Local Analysis of a New Multipliers Method

Luís N. Vicente [*]

Abstract

In this paper we introduce a penalty function and a corresponding multipliers method for the solution of a class of nonlinear programming problems where the equality constraints have a particular structure. The class models optimal control and engineering design problems with bounds on the state and control variables and has wide applicability.

The multipliers method updates multipliers corresponding to inequality constraints (maintaining their nonnegativity) instead of dealing with multipliers associated with equality constraints. The basic local convergence properties of the method are proved and a dual framework is introduced. We also analyze the properties of the penalized problem related with the penalty function.

Keywords. Nonlinear programming, optimal control, state constraints, penalty function, multipliers method, augmented Lagrangian.

AMS subject classifications. 49M37, 90C06, 90C30

1 Introduction

Consider the following nonlinear optimization problem:

$$\begin{aligned} \text{minimize} \quad & f(y, u) \\ \text{subject to} \quad & c(y, u) = 0 \,, \\ & (y, u) \geq 0 \,, \end{aligned} \qquad (1)$$

in the variables

$$x = \left(\begin{array}{c} y \\ u \end{array} \right) ,$$

where the structure of the equality constraints $c(y, u) = 0$ arises from optimal control and design engineering problems. So, we assume that the state variables y are in $\mathbb{R}^{n_y}$ and the control or design variables u lie in $\mathbb{R}^{n_u}$, where n_y and n_u are positive integers. The functions f and c are considered smooth and defined as $f : \Omega \longrightarrow \mathbb{R}$ and $c : \Omega \longrightarrow \mathbb{R}^{n_y}$, where Ω is an open set of $\mathbb{R}^{n_y + n_u}$. The nonlinear system $c(y, u) = 0$ with n_y equalities is the state equation. Due to the partition of x in y and u, we can write the Jacobian matrix of c as:

$$\nabla c(y, u)^\top = \left(\begin{array}{cc} c_y(y, u) & c_u(y, u) \end{array} \right) ,$$

where the partial Jacobian $c_y(y, u)$ is a square matrix of order n_y. We assume that $c_y(y, u)$ is nonsingular in Ω. In the Appendix, the optimality conditions for this problem are described, assuming that the functions f and c are twice continuously differentiable in Ω.

This problem setting formulates a broad class of optimal control and design engineering problems. For instance, any problem of the form

$$\begin{aligned} \text{minimize} \quad & f(w, u) \\ \text{subject to} \quad & d(w, u) = 0 \,, \\ & g(w, u) \geq 0 \,, \\ & (w, u) \geq 0 \,, \end{aligned}$$

where the partial Jacobian $d_w(w, u)$ is nonsingular, can be reformulated as problem (1) by setting

$$y \;=\; \begin{pmatrix} w \\ s \end{pmatrix} \qquad \text{and} \qquad c(y, u) \;=\; c(w, s, u) \;=\; \begin{pmatrix} d(w, u) \\ g(w, u) - s \end{pmatrix} .$$

In this situation, as it can be easily checked, $c_y(y, u)$ would still be nonsingular.

Since we are assuming that $c_y(y, u)$ is nonsingular, the implicit function theorem guarantees the local existence of a smooth vector function $y(u)$ from $\mathbb{R}^{n_u}$ to $\mathbb{R}^{n_y}$ defined by $c(y(u), u) = 0$. This allows us to reduce the minimization problem (1) to the space of the control variables u:

$$\begin{aligned} &\text{minimize} \quad f(y(u), u) \\ &\text{subject to} \quad y(u) \geq 0, \quad u \geq 0 \,. \end{aligned} \qquad\qquad (2)$$

The formulation (1) is called "all-at-once" whereas the formulation (2) is referred to as "black-box". Note that bounds on states in the all-at-once formulation correspond to nonlinear inequality constraints on the controls in the "black-box" formulation. Further, if the optimal control problem is given in the form (2) it can be easily reformulated in the form (1):

$$\begin{aligned} &\text{minimize} \quad f(y(u), u) \\ &\text{subject to} \quad y(u) - s = 0, \\ &\qquad\qquad\quad (s, u) \geq 0 \,. \end{aligned}$$

The multipliers method was proposed by Hestenes [18] and Powell [26] for nonlinear optimization problems with equality constraints. The book by Bertsekas [1] gives a comprehensive treatment of this topic with connections to other approaches. Extensions of the multipliers method for inequality constraints have been proposed using slack and square slack variables and nondifferentiable penalty functions (see the books by Bertsekas [1, Sections 3.1-3.2] and Fletcher [10, Section 12.2], the papers by Rockafellar [27, 28], and the references therein). A common feature in these approaches is that the nonnegativity of the multipliers corresponding to inequality constraints has to be explicitly imposed. Another approach was introduced later by Conn, Gould, and Toint [5]. Their algorithm, implemented in the optimization solver LANCELOT [6], first converts the nonlinear programming problem into a problem with equality constraints and simple bounds. The objective function and the equality constraints then define the augmented Lagrangian function that does not take into account the simple bounds. Their method then consists of minimizing a sequence of augmented Lagrangian functions within the bounds.

A (smooth) penalty function P for the solution of problem (1) is introduced in Section 2 and its derivatives computed in Section 3. The penalty function is based on Fletcher's augmented Lagrangian penalty function [9]. In fact, the penalty function is derived by eliminating from the first-order necessary conditions for (1) the multipliers λ corresponding to the state equation $c(y, u) = 0$. This derivation also means that the penalty function P depends on the values of the multipliers z_y corresponding to the bounds $y \geq 0$. Because of the dependence on the multipliers z_y, this penalty function is not exact in the traditional sense [24, 25]. However it does share some exactness properties as it is shown in Section 4.

The multipliers method we propose in this paper is based on the penalty function P. Its novelty comes from the fact that the multipliers that are explicitly updated are the multipliers z_y corresponding to the bounds on the state variables y. The update formula maintains these multipliers z_y nonnegative without any artificial use of the $\max(0, \cdot)$ operator. The multipliers λ (that depend in turn on z_y) corresponding to the state equation $c(y, u) = 0$ are then implicitly updated within the formula for P. The traditional multipliers method for problems with equality constraints requires at each major iteration the solution of an unconstrained minimization problem; the method proposed here (like the one in [5]) requires the solution of a minimization problem with simple bounds.

The description of the multipliers method and the proof of its basic local convergence properties are given in Section 5. In Section 6 we introduce the dual interpretation of this multipliers method. In Section 7 we state some conclusions and point out directions for future work.

2   The penalty function

The penalty function we propose in this paper is based on the Lagrangian of f with respect to the constraints $c(y, u) = 0$:

$$\ell(y, u, \lambda) \;=\; f(y, u) + c(y, u)^\top \lambda$$

and on the corresponding augmented Lagrangian function:

$$L(y, u, \lambda; \mu) \;=\; f(y, u) + c(y, u)^\top \lambda + \frac{1}{2\mu} c(y, u)^\top c(y, u) \,,$$

where $\lambda \in \mathbb{R}^{n_y}$ are multipliers corresponding to the state equation and μ is a positive penalty parameter. We would like to come up with an expression for λ in terms of (y, u), hoping that that would lead us to a penalty function with some interesting exactness properties. To do so, we take a close look to the structure of our problem. A point (y, u) satisfies the first-order necessary conditions for problem (1) (see the Appendix) if there exist $\lambda \in \mathbb{R}^{n_y}$ and $(z_y, z_u) \in \mathbb{R}^{n_y + n_u}$ such that

$$\nabla_y f(y, u) + c_y(y, u)^\top \lambda - z_y = 0 \,,$$

$$\nabla_u f(y, u) + c_u(y, u)^\top \lambda - z_u = 0 \,,$$

$$c(y, u) = 0 \,, \;\; (y, u) \geq 0 \,,$$

$$y^\top z_y \;=\; u^\top z_u \;=\; 0 \,, \;\; (z_y, z_u) \geq 0 \,.$$

Given y, u, and z_y, the first-order necessary conditions indicate a formula for λ:

$$\lambda(y, u, z_y) \;=\; -c_y(y, u)^{-\top} \left(\nabla_y f(y, u) - z_y \right) . \tag{3}$$

Given the parameters $z_y \geq 0$ and $\mu > 0$, we introduce the penalty function

$$P(y, u; z_y, \mu) \;=\; f(y, u) + c(y, u)^\top \lambda(y, u, z_y) + \frac{1}{2\mu} c(y, u)^\top c(y, u)$$

and the corresponding penalized problem:

$$\begin{aligned} \text{minimize} \quad & P(y, u; z_y, \mu) \\ \text{subject to} \quad & (y, u) \geq 0 \,, \end{aligned} \tag{4}$$

in the optimization variables y and u. The parameter z_y is updated explicitly rather than incorporated in the optimization variables. The penalty function P shares some exactness properties as it is shown in Section 4.

Throughout this paper we will make the following assumptions.

Assumptions 2.1 *The functions f and c are three times continuously differentiable in an open set $\Omega \subset \mathbb{R}^{n_y + n_u}$. The partial Jacobian $c_y(y, u)$ is nonsingular in Ω.*

To alleviate the notation we will omit the arguments $x = (y, u)$ and $x^* = (y^*, u^*)$ when it is clear from the context where the functions are evaluated. For instance, $c_u = c_u(y, u)$ and $\nabla_y f^* = \nabla_y f(y^*, u^*)$. The symbol e represents a vector of ones with appropriate size. Also, for any vector v, V is the diagonal matrix for which the diagonal elements are the elements of v. We use the notation $\|v\| = \mathcal{O}(\delta, \epsilon)$ or $v = \mathcal{O}(\delta, \epsilon)$ (and say that v is of size δ and ϵ) to denote $\|v\| \leq \kappa(\delta + \epsilon)$, where κ is a positive constant.

3   Derivatives of the penalty function

We calculate now the gradient and the Hessian of $P(y, u; z_y, \mu)$ with respect to y and u.

To obtain the first-order partial derivatives of $P(y, u; z_y, \mu)$, we first calculate

$$\nabla_y \lambda(y, u, z_y) \;=\; -c_y(y, u)^{-\top} \nabla_{yy}^2 \ell(y, u, \lambda(y, u, z_y)) \,,$$

$$\nabla_u \lambda(y, u, z_y) \;=\; -c_y(y, u)^{-\top} \nabla_{yu}^2 \ell(y, u, \lambda(y, u, z_y)) \,,$$

by differentiating $c_y(y,u)^\top \lambda(y,u,z_y) = -\nabla_y f(y,u) + z_y$ with respect to y and u, respectively. Thus the gradient of $P(y,u;z_y,\mu)$ is given by

$$\nabla P(y,u;z_y,\mu) \;=\; G_1(y,u;z_y,\mu) + G_2(y,u;z_y,\mu) + G_3(y,u;z_y,\mu)\,,$$

where

$$G_1(y,u;z_y,\mu) \;=\; \begin{pmatrix} z_y \\ \nabla_u f(y,u) - c_u(y,u)^\top c_y(y,u)^{-\top}(\nabla_y f(y,u) - z_y) \end{pmatrix}\,,$$

$$G_2(y,u;z_y,\mu) \;=\; \begin{pmatrix} -\nabla_{yy}^2 \ell(y,u,\lambda(y,u,z_y))c_y(y,u)^{-1}c(y,u) \\ -\nabla_{uy}^2 \ell(y,u,\lambda(y,u,z_y))c_y(y,u)^{-1}c(y,u) \end{pmatrix}\,,$$

and

$$G_3(y,u;z_y,\mu) \;=\; \begin{pmatrix} \frac{1}{\mu}c_y(y,u)^\top c(y,u) \\ \frac{1}{\mu}c_u(y,u)^\top c(y,u) \end{pmatrix}\,.$$

If $c(y,u)=0$, the gradient $\nabla P(y,u;z_y,\mu)$ is just:

$$\nabla P(y,u;z_y,\mu) \;=\; G_1(y,u;z_y,\mu) \;=\; \begin{pmatrix} z_y \\ w(y,u)^\top \begin{pmatrix} \nabla_y f(y,u) - z_y \\ \nabla_u f(y,u) \end{pmatrix} \end{pmatrix}\,,$$

where

$$w(y,u) \;=\; \begin{pmatrix} -c_y(y,u)^{-1}c_u(y,u) \\ I \end{pmatrix}\,. \tag{5}$$

To compute the gradient one has to solve one linearized state equation $c_y(y,u)^{-1}c(y,u)$ and one adjoint equation $c_y(y,u)^{-\top}(\nabla_y f(y,u) - z_y)$.

We calculate now the Hessian of $P(y,u;z_y,\mu)$. The procedures are similar as before. For instance, we need to obtain the partial derivatives of $c_y(y,u)^{-1}c(y,u)$ with respect to y and u and we accomplish this task by noticing that $s(y,u) = c_y(y,u)^{-1}c(y,u)$ and by differentiating $c_y(y,u)s(y,u) = c(y,u)$:

$$\nabla_y s(y,u) \;=\; I - c_y(y,u)^{-1} \sum_{i=1}^{n_y} \left(c_y(y,u)^{-1}c(y,u)\right)_i \nabla_{yy}^2 c_i(y,u)\,,$$

$$\nabla_u s(y,u) \;=\; c_y(y,u)^{-1}c_u(y,u) - c_y(y,u)^{-1}\sum_{i=1}^{n_y}\left(c_y(y,u)^{-1}c(y,u)\right)_i \nabla_{yu}^2 c_i(y,u)\,.$$

Moreover, we need to differentiate the u component of $G_1(y,u;z_y,\mu)$ with respect to y and u. This component is nothing else than $\nabla_u \ell(y,u,z_y) = \nabla_u f(y,u) + c_u(y,u)^\top \lambda(y,u,z_y)$, where $\ell(y,u,z_y) = \ell(y,u,\lambda(y,u,z_y))$. So, its derivatives are given by

$$\begin{aligned}
\nabla_y\left(\nabla_u \ell(y,u,z_y)\right) &=& \nabla_{uy}^2 f(y,u) + \sum_{i=1}^{n_y}\lambda(y,u,z_y)_i\nabla_{uy}^2 c_i(y,u) \\
&& -c_u(y,u)^\top c_y(y,u)^{-\top}\nabla_{yy}^2\ell(y,u,\lambda(y,u,z_y))\,, \\
\nabla_u\left(\nabla_u \ell(y,u,z_y)\right) &=& \nabla_{uu}^2 f(y,u) + \sum_{i=1}^{n_y}\lambda(y,u,z_y)_i\nabla_{uu}^2 c_i(y,u) \\
&& -c_u(y,u)^\top c_y(y,u)^{-\top}\nabla_{yu}^2\ell(y,u,\lambda(y,u,z_y))\,.
\end{aligned}$$

Hence, the Hessian of $P(y,u;z_y,\mu)$ is expressed as

$$\nabla^2 P(y,u;z_y,\mu) \;=\; H_1(y,u;z_y,\mu) + H_2(y,u;z_y,\mu) + H_3(y,u;z_y,\mu)\,,$$

where

$$H_1(y,u;z_y,\mu) \;=\; \begin{pmatrix} 0 & 0 \\ -c_u^\top c_y^{-\top}\nabla_{yy}^2\ell + \nabla_{uy}^2\ell & -c_u^\top c_y^{-\top}\nabla_{yu}^2\ell + \nabla_{uu}^2\ell \end{pmatrix}\,,$$

$$H_2(y, u; z_y, \mu) = \begin{pmatrix} -\nabla_{yy}^2 \ell & -\nabla_{yy}^2 \ell \, c_y^{-1} c_u \\ -\nabla_{uy}^2 \ell & -\nabla_{uy}^2 \ell \, c_y^{-1} c_u \end{pmatrix} + \begin{pmatrix} \nabla_{yy}^2 \ell (I - \nabla_y s) & \nabla_{yy}^2 \ell (c_y^{-1} c_u - \nabla_u s) \\ \nabla_{uy}^2 \ell (I - \nabla_y s) & \nabla_{uy}^2 \ell (c_y^{-1} c_u - \nabla_u s) \end{pmatrix}$$

$$+ \begin{pmatrix} -\sum_{j=1}^{n_y} (c_y^{-1} c)_j \nabla_y \left(\nabla_{yy}^2 \ell \right)_j & -\sum_{j=1}^{n_y} (c_y^{-1} c)_j \nabla_u \left(\nabla_{yy}^2 \ell \right)_j \\ -\sum_{j=1}^{n_y} (c_y^{-1} c)_j \nabla_y \left(\nabla_{uy}^2 \ell \right)_j & -\sum_{j=1}^{n_y} (c_y^{-1} c)_j \nabla_u \left(\nabla_{uy}^2 \ell \right)_j \end{pmatrix},$$

and

$$H_3(y, u; z_y, \mu) = \begin{pmatrix} \frac{1}{\mu} c_y^\top c_y + \frac{1}{\mu} \sum_{i=1}^{n_y} c_i \nabla_{yy}^2 c_i & \frac{1}{\mu} c_y^\top c_u + \frac{1}{\mu} \sum_{i=1}^{n_y} c_i \nabla_{yu}^2 c_i \\ \frac{1}{\mu} c_u^\top c_y + \frac{1}{\mu} \sum_{i=1}^{n_y} c_i \nabla_{uy}^2 c_i & \frac{1}{\mu} c_u^\top c_u + \frac{1}{\mu} \sum_{i=1}^{n_y} c_i \nabla_{uu}^2 c_i \end{pmatrix}.$$

If $c(y, u) = 0$, the Hessian of $P(y, u; z_y, \mu)$ reduces to

$$\begin{pmatrix} -\nabla_{yy}^2 \ell & -\nabla_{yy}^2 \ell \, c_y^{-1} c_u \\ -c_u^\top c_y^{-\top} \nabla_{yy}^2 \ell & \begin{matrix} -c_u^\top c_y^{-\top} \nabla_{yu}^2 \ell + \nabla_{uu}^2 \ell \\ -\nabla_{uy}^2 \ell \, c_y^{-1} c_u \end{matrix} \end{pmatrix} + \begin{pmatrix} \frac{1}{\mu} c_y^\top c_y & \frac{1}{\mu} c_y^\top c_u \\ \frac{1}{\mu} c_u^\top c_y & \frac{1}{\mu} c_u^\top c_u \end{pmatrix}.$$

4 Properties of the penalty function

In this section we analyze the relationships between problem (1) and the penalized problem (4). We start with a result that is analogous to [1, Proposition 2.3]. We appeal to the first-order necessary conditions for problem (4):

$$z_y - \nabla_{yy}^2 \ell \, c_y^{-1} c + \frac{1}{\mu} c_y^\top c - \bar{z}_y = 0, \tag{6}$$

$$\nabla_u f - c_u^\top c_y^{-\top} \left(\nabla_y f - z_y \right) - \nabla_{uy}^2 \ell \, c_y^{-1} c + \frac{1}{\mu} c_u^\top c - \bar{z}_u = 0, \tag{7}$$

$$(y, u) \geq 0, \tag{8}$$

$$y^\top \bar{z}_y = u^\top \bar{z}_u = 0, \ (\bar{z}_y, \bar{z}_u) \geq 0. \tag{9}$$

Theorem 4.1 *Let Assumptions 2.1 hold. If $(x_k, (\bar{z}_y)_k, (\bar{z}_u)_k)$ satisfies the first-order necessary conditions for (4) with $\mu_k > 0$ and $(z_y)_k \geq 0$, $\{(z_y)_k\}$ is bounded, and $\lim_{k \to +\infty} \mu_k = 0$, then every limit point of $\{(x_k, (\bar{z}_y)_k, (\bar{z}_u)_k)\}$ satisfies the first-order necessary conditions for the original problem (1).*

Proof: Let $(x, \bar{z}_y, \bar{z}_u)$ be a limit point of $\{(x_k, (\bar{z}_y)_k, (\bar{z}_u)_k)\}$. Since $\{(z_y)_k\}$ is bounded, there exists a subsequence $\{(x_k, (\bar{z}_y)_k, (\bar{z}_u)_k)\}_\mathcal{K}$ such that $\lim_{k \in \mathcal{K}} (x_k, (\bar{z}_y)_k, (\bar{z}_u)_k) = (x, \bar{z}_y, \bar{z}_u)$ and $\lim_{k \in \mathcal{K}} (z_y)_k = z_y$. Now we set

$$(\tilde{z}_y)_k = \nabla_y P(x_k; (z_y)_k, \mu_k).$$

It follows from the first-order necessary conditions for (4) that

$$\lim_{k \in \mathcal{K}} (\tilde{z}_y)_k = \lim_{k \in \mathcal{K}} \nabla_y P(x_k; (z_y)_k, \mu_k) = \bar{z}_y \geq 0.$$

Thus, $\lim_{k \in \mathcal{K}} \frac{1}{\mu_k} c_y(x_k)^\top c(x_k)$ exists. Since $\lim_{k \to +\infty} \mu_k = 0$, we have that

$$\lim_{k \in \mathcal{K}} c(x_k) = 0$$

and $c(x) = 0$. Then $\tilde{z}_y = \lim_{k \in \mathcal{K}} (z_y)_k + \frac{1}{\mu_k} c_y(x_k)^\top c(x_k)$ and

$$
\begin{aligned}
0 &= \lim_{k \in \mathcal{K}} \Big(\nabla_u f(x_k) - c_u(x_k)^\top c_y(x_k)^{-\top} (\nabla_y f(x_k) - (z_y)_k) \\
&\qquad - \nabla_{uy}^2 \ell(x_k, (z_y)_k) c_y(x_k)^{-1} c(x_k) + \frac{1}{\mu} c_u(x_k)^\top c(x_k) - (\bar{z}_u)_k \Big) \\
&= \lim_{k \in \mathcal{K}} \Big(\nabla_u f(x_k) \\
&\qquad - c_u(x_k)^\top c_y(x_k)^{-\top} (\nabla_y f(x_k) - (z_y)_k - \tfrac{1}{\mu_k} c_y(x_k)^\top c(x_k)) - (\bar{z}_u)_k \Big) \\
&= \nabla_u f(x) - c_u(x)^\top c_y(x)^{-\top} (\nabla_y f(x) - \tilde{z}_y) - \bar{z}_u \,.
\end{aligned}
$$

Since $y^\top \tilde{z}_y = u^\top \bar{z}_u = 0$ the proof is completed. $\square$

The next two theorems relate problems (1) and (4).

Theorem 4.2 *Let Assumptions 2.1 hold. Also, let $\mu > 0$ and x be such that $c(x) = 0$. The point x satisfies the first-order necessary conditions for the original problem (1) with multipliers z_y corresponding to $y \geq 0$ if and only if x satisfies the first-order necessary conditions for the penalized problem (4) for some $z_y \geq 0$.*

Proof: If $c(y, u) = 0$, the first-order necessary conditions for the penalized problem (4) imply

$$
\begin{aligned}
z_y - \bar{z}_y &= 0 \,, \\
\nabla_u f(y, u) - c_u(y, u)^\top c_y(y, u)^{-\top} (\nabla_y f(y, u) - z_y) - \bar{z}_u &= 0 \,, \\
(y, u) &\geq 0 \,, \\
y^\top \bar{z}_y = u^\top \bar{z}_u = 0 \,, \quad (\bar{z}_y, \bar{z}_u) &\geq 0 \,.
\end{aligned}
$$

Since $z_y = \bar{z}_y$, these conditions can be rewritten as

$$
\begin{aligned}
\nabla_u f(y, u) - c_u(y, u)^\top c_y(y, u)^{-\top} (\nabla_y f(y, u) - z_y) - \bar{z}_u &= 0 \,, \\
(y, u) &\geq 0 \,, \\
y^\top z_y = u^\top \bar{z}_u = 0 \,, \quad (z_y, \bar{z}_u) &\geq 0 \,,
\end{aligned}
$$

which are the first-order necessary conditions for the original problem (1) with $z_u = \bar{z}_u$. $\square$

Also, if x satisfies the first-order necessary conditions for the penalized problem (4) for some $z_y \geq 0$ and $z_y = \bar{z}_y$, then we get

$$
-\nabla_{yy}^2 \ell(x, \lambda(x, z_y)) c_y(x)^{-1} c(x) + \frac{1}{\mu} c_y(x)^\top c(x) = 0 \,.
$$

So,

$$
\big(c_y(x)^{-1} c(x) \big)^\top \left(-\nabla_{yy}^2 \ell(x, \lambda(x, z_y)) + \frac{1}{\mu} c_y(x)^\top c_y(x) \right) \big(c_y(x)^{-1} c(x) \big) = 0
$$

and there exists $\mu^* > 0$ such that for all $0 < \mu \leq \mu^*$, $c_y(x)^{-1} c(x) = 0$, i.e., $c(x) = 0$.

Theorem 4.3 *Let Assumptions 2.1 hold. Also, let $\mu > 0$ and (y, u) be such that $c(y, u) = 0$. If (y, u) satisfies the second-order necessary (sufficient) conditions for the penalized problem (4) for some $z_y \geq 0$, then (y, u) satisfies the second-order necessary (sufficient) conditions for the original problem (1).*

Proof: The first-order part of this result follows from the previous theorem. To establish the second-order assertions we need to calculate

$$
\left(\begin{array}{c} \Delta y \\ \Delta u \end{array} \right)^\top \nabla^2 P(y, u; z_y, \mu) \left(\begin{array}{c} \Delta y \\ \Delta u \end{array} \right) \tag{10}
$$

in the situation where $(\Delta y, \Delta u)$ satisfies:

$$(\Delta y)_i = 0 \quad \text{if} \quad y_i = 0 \quad \text{and} \quad (\bar{z}_y)_i > 0\,, \tag{11}$$

$$(\Delta y)_i \geq 0 \quad \text{if} \quad y_i = 0 \quad \text{and} \quad (\bar{z}_y)_i = 0\,, \tag{12}$$

$$(\Delta u)_i = 0 \quad \text{if} \quad u_i = 0 \quad \text{and} \quad (\bar{z}_u)_i > 0\,, \tag{13}$$

$$(\Delta u)_i \geq 0 \quad \text{if} \quad u_i = 0 \quad \text{and} \quad (\bar{z}_u)_i = 0\,. \tag{14}$$

We will also assume that $(\Delta y, \Delta u)$ lies in the null-space of the Jacobian of $c(y, u)$, i.e., that $\Delta y = -c_y(y, u)^{-1} c_u(y, u)\Delta u$. Since $c(y, u) = 0$, the value of (10) is given by:

$$-\Delta y^\top \nabla^2_{yy}\ell\Delta y - \Delta y^\top \nabla^2_{yy}\ell\, c_y^{-1} c_u \Delta u - \Delta u^\top c_u^\top c_y^{-\top} \nabla^2_{yy}\ell\Delta y$$

$$+\Delta u^\top \nabla^2_{uu}\ell\Delta u - \Delta u^\top c_u^\top c_y^{-\top} \nabla^2_{yu}\ell\Delta u - \Delta u^\top \nabla^2_{uy}\ell\, c_y^{-1} c_u \Delta u$$

$$+\tfrac{1}{\mu}\Delta y^\top c_y^\top c_y \Delta y + \tfrac{1}{\mu}\Delta y^\top c_y^\top c_u \Delta u + \tfrac{1}{\mu}\Delta u^\top c_u^\top c_y \Delta y + \tfrac{1}{\mu}\Delta u^\top c_u^\top c_u \Delta u\,.$$

Thus, from $\Delta y = -c_y^{-1} c_u \Delta u$, a number of terms cancel out and we obtain:

$$\left(\begin{array}{c} \Delta y \\ \Delta u \end{array}\right)^\top \nabla^2_{xx}\ell(y, u, z_y) \left(\begin{array}{c} \Delta y \\ \Delta u \end{array}\right) \;=\; \left(\begin{array}{c} \Delta y \\ \Delta u \end{array}\right)^\top \nabla^2 P(y, u; z_y, \mu) \left(\begin{array}{c} \Delta y \\ \Delta u \end{array}\right)\,.$$

Since $(\Delta y, \Delta u)$ satisfies (11)-(14) and $c_y(y, u)\Delta y + c_u(y, u)\Delta u = 0$ the proof is completed. $\qquad\square$

Finally, we establish the exactness property of the penalty function P (a result that can be seen as the reciprocal of Theorem 4.3).

Theorem 4.4 *Let Assumptions 2.1 hold. Also, let x be such that $c(x) = 0$. If (x, z_y) satisfies the second-order necessary (sufficient) conditions for the original problem (1) with multipliers z_y corresponding to $y \geq 0$, then there exists a $\mu^* > 0$ such that (y, u) satisfies the second-order necessary (sufficient) conditions for the penalized problem (4) for this z_y and any $0 < \mu \leq \mu^*$.*

Proof: We prove the result only for the sufficient conditions. Let Δy and Δu satisfy (11)-(14). The proof is based on the fact that

$$\left(\begin{array}{c} \Delta y \\ \Delta u \end{array}\right)^\top \nabla^2 P(y, u; z_y, \mu) \left(\begin{array}{c} \Delta y \\ \Delta u \end{array}\right)$$

$$= \left(\begin{array}{c} -c_y(y, u)^{-1} c_u(y, u)\Delta u \\ \Delta u \end{array}\right)^\top \nabla^2_{xx}\ell(y, u, z_y) \left(\begin{array}{c} -c_y(y, u)^{-1} c_u(y, u)\Delta u \\ \Delta u \end{array}\right)$$

$$+ \left(\begin{array}{c} \Delta e \\ 0 \end{array}\right)^\top \left(\begin{array}{cc} -\nabla^2_{yy}\ell(y, u, z_y) + \tfrac{1}{\mu}c_y(y, u)^\top c_y(y, u) & 0 \\ 0 & 0 \end{array}\right) \left(\begin{array}{c} \Delta e \\ 0 \end{array}\right)$$

$$= \left(\begin{array}{c} \Delta y \\ \Delta u \end{array}\right)^\top \left(\begin{array}{cc} I & 0 \\ c_u^\top c_y^{-\top} & I \end{array}\right) \left(\begin{array}{cc} -\nabla^2_{yy}\ell + \tfrac{1}{\mu}c_y^\top c_y & 0 \\ 0 & w^\top \nabla^2_{xx}\ell w \end{array}\right) \left(\begin{array}{cc} I & c_y^{-1} c_u \\ 0 & I \end{array}\right) \left(\begin{array}{c} \Delta y \\ \Delta u \end{array}\right)$$

where $\Delta e = \Delta y - (-c_y(y, u)^{-1} c_u(y, u)\Delta u)$. Since (13)-(14) hold, the second-order sufficient conditions for problem (1) imply that

$$\Delta u^\top w(x)^\top \nabla^2_{xx}\ell(x, z_y)w(x)\Delta u \;>\; 0\,.$$

The proof is completed by setting:

$$\mu^* = \left\{ \begin{array}{ll} \text{any positive real} & \text{if } \nabla^2_{yy}\ell(x, z_y) \text{ is negative semi-definite,} \\[2mm] \frac{\alpha(x)}{\beta(x, z_y)} & \text{otherwise,} \end{array} \right.$$

where $c_y(x)^\top c_y(x)$ has eigenvalue decomposition $c_y(x)^\top c_y(x) = Q(x)A(x)Q(x)^\top$ and smallest eigenvalue $\alpha(x)$, and $\beta(x, z_y)$ is the largest eigenvalue of $Q(x)^\top \nabla^2_{yy}\ell(x, z_y)Q(x)$. $\qquad\square$

5   The multipliers method

The penalty function P together with the penalized problem (4) and the equation (6) motivate a new multipliers method.

Algorithm 5.1 (Multipliers method for (1))

1. Choose initial values: μ_0 for the penalty parameter and z_y^0 for the approximation of the multipliers.

2. For $k = 0, 1, 2, \ldots$ do

   2.1 Solve problem (4) with $z_y = z_y^k$ and $\mu = \mu_k$.

   2.2 Update the multipliers approximation:

   $$
   \begin{aligned}
   z_y^{k+1} \;=\; & z_y^k - \nabla_{yy}^2 \ell(\tilde{x}(z_y^k, \mu_k), z_y^k) c_y(\tilde{x}(z_y^k, \mu_k))^{-1} c(\tilde{x}(z_y^k, \mu_k)) \\
   & + \tfrac{1}{\mu_k} c_y(\tilde{x}(z_y^k, \mu_k))^\top c(\tilde{x}(z_y^k, \mu_k)) \,,
   \end{aligned}
   \tag{15}
   $$

   where $\tilde{x}(z_y^k, \mu_k)$ is the solution obtained in Step 2.1.

   2.3 Update the penalty parameter μ_{k+1}.

If $\nabla_{yy}^2 \ell(\tilde{x}(z_y^k, \mu_k), z_y^k) c_y(\tilde{x}(z_y^k, \mu_k))^{-1} c(\tilde{x}(z_y^k, \mu_k)) \simeq 0$, then the update (15) is just

$$
z_y^{k+1} \;\simeq\; z_y^k + \frac{1}{\mu_k} c_y(\tilde{x}(z_y^k, \mu_k))^\top c(\tilde{x}(z_y^k, \mu_k)) \,,
$$

which differs from the classical update of the multipliers methods for equality constrained optimization only because $c_y(\tilde{x}(z_y^k, \mu_k))^\top$ is multiplying $c(\tilde{x}(z_y^k, \mu_k))$ on the left.

The basic local convergence properties of the multipliers method 5.1 are proved under the following set of assumptions.

Assumptions 5.1 *The point $x^* = (y^*, u^*) \in \Omega$ with corresponding multipliers $\bar{z}^* = (\bar{z}_y^*, \bar{z}_u^*)$ is a nondegenerate point satisfying the second-order sufficient conditions and strict complementarity for problem (1).*

The main result is proved in Theorem 5.1 and bounds the distance between a local minimizer of (4) and $(x^*, \bar{z}^*)$ by the penalty parameter μ times the distance between the parameter z_y and the corresponding multipliers $\bar{z}_y^*$. Although the structure of the proof follows [1, Proposition 2.4], we have additional difficulties here due to the presence of the bound constraints on the variables. Another difficulty arises when dealing with the cross term in the multipliers update. This term is not multiplied by $1/\mu_k$ but involves z_y^k. A consequence of having to handle this extra term is that the region D^* in (17) becomes smaller than the one in [1, Proposition 2.4] where instead of $\min\{\delta, \delta/\mu\}$ we only have δ/μ.

Theorem 5.1 *Let $x^* = (y^*, u^*)$ with corresponding multipliers $\bar{z}^* = (\bar{z}_y^*, \bar{z}_u^*)$ satisfy Assumptions 2.1-5.1.*
There exist positive scalars $\bar{\mu}$, δ, ϵ, κ_1, κ_2, κ_3, and κ_4 such that

$$
-\nabla_{yy}^2 \ell(y^*, u^*, \bar{z}_y^*) + \frac{1}{\bar{\mu}} c_y(y^*, u^*)^\top c_y(y^*, u^*)
$$

is positive definite, the problem

$$
\begin{aligned}
minimize \quad & P(y, u; z_y, \mu) \\
subject\ to \quad & (y, u) \geq 0 \,, \\
& (y, u) \in B(y^*, u^*; \epsilon) \,,
\end{aligned}
\tag{16}
$$

has a unique solution $\tilde{x}(z_y, \mu)$ for all (z_y, μ) in

$$
D^* \;=\; \left\{ (z_y, \mu) : \; \|z_y - \bar{z}_y^*\| < \min\left\{\delta, \frac{\delta}{\mu}\right\} ; \quad 0 < \mu \leq \bar{\mu} \right\} \,,
\tag{17}
$$

the function $\tilde{x}(z_y, \mu)$ is continuously differentiable in D^*, and for all $(z_y, \mu) \in D^*$ we have

$$\|\tilde{y}(z_y, \mu) - y^*\| \quad \leq \quad \kappa_1 \mu \|z_y - \bar{z}_y^*\|, \tag{18}$$

$$\|\tilde{u}(z_y, \mu) - u^*\| \quad \leq \quad \kappa_2 \mu \|z_y - \bar{z}_y^*\|, \tag{19}$$

$$\|\tilde{z}_y(z_y, \mu) - \bar{z}_y^*\| \quad \leq \quad \kappa_3 \mu \|z_y - \bar{z}_y^*\|, \tag{20}$$

$$\|\tilde{z}_u(z_y, \mu) - \bar{z}_u^*\| \quad \leq \quad \kappa_4 \mu \|z_y - \bar{z}_y^*\|, \tag{21}$$

where $\tilde{z}_y(z_y, \mu)$ and $\tilde{z}_u(z_y, \mu)$ are the multipliers corresponding to $\tilde{y}(z_y, \mu)$ and $\tilde{u}(z_y, \mu)$, and $\tilde{z}_y(z_y, \mu)$ is such that:

$$\begin{aligned} \tilde{z}_y(z_y, \mu) \quad = \quad & z_y - \nabla_{yy}^2 \ell(\tilde{x}(z_y, \mu), z_y) c_y(\tilde{x}(z_y, \mu))^{-1} c(\tilde{x}(z_y, \mu)) \\ & + \tfrac{1}{\mu} c_y(\tilde{x}(z_y, \mu))^\top c(\tilde{x}(z_y, \mu)) \,. \end{aligned} \tag{22}$$

Proof: Consider, for $\mu > 0$, the system of nonlinear equations that results from the first-order necessary conditions (6)-(9) for problem (4). We replace z_y in the second equation (7) from its expression in the first equation (6)

$$z_y - \nabla_{yy}^2 \ell(x, z_y) c_y^{-1} c + \frac{1}{\mu} c_y^\top c - \bar{z}_y = 0 \,, \tag{23}$$

$$\nabla_u f - c_u^\top c_y^{-\top} \left(\nabla_y f - \nabla_{yy}^2 \ell(x, z_y) c_y^{-1} c - \bar{z}_y \right) - \nabla_{uy}^2 \ell(x, z_y) c_y^{-1} c - \bar{z}_u = 0 \,, \tag{24}$$

$$y^\top \bar{z}_y = u^\top \bar{z}_u = 0 \,, \tag{25}$$

canceling the term $\frac{1}{\mu} c_u^\top c$. Now we multiply equation (23) by μ and do the changes of variables

$$r \quad = \quad \mu(z_y - \bar{z}_y^*) \,, \tag{26}$$

$$s \quad = \quad z_y - \bar{z}_y^* \,, \tag{27}$$

to obtain

$$r - \mu \nabla_{yy}^2 \ell(x, \bar{z}_y^* + s) c_y^{-1} c + c_y^\top c + \mu \bar{z}_y^* - \mu \bar{z}_y \quad = \quad 0 \,, \tag{28}$$

$$\nabla_u f - c_u^\top c_y^{-\top} \left(\nabla_y f - \nabla_{yy}^2 \ell(x, \bar{z}_y^* + s) c_y^{-1} c - \bar{z}_y \right)$$

$$-\nabla_{uy}^2 \ell(x, \bar{z}_y^* + s) c_y^{-1} c - \bar{z}_u \quad = \quad 0 \,, \tag{29}$$

$$Y \bar{Z}_y e \quad = U \bar{Z}_u e \quad = \quad 0 \,. \tag{30}$$

We analyze this system for $\mu \in [0, \mu^*]$, where μ^* is such that $-\mu \nabla_{yy}^2 \ell^* + c_y^{*\top} c_y^*$ is positive definite for all $\mu \in [0, \mu^*]$. For $r = s = 0$ and $\mu \in [0, \mu^*]$, it is easy to check that the system (28)-(30) has the solution $(y^*, u^*, \bar{z}_y^*, \bar{z}_u^*)$.

For $r = s = 0$, the Jacobian of (28)-(30) with respect to $(y, u, \bar{z}_y, \bar{z}_u)$ at $(y^*, u^*, \bar{z}_y^*, \bar{z}_u^*)$ is given by

$$J^*(0, 0, \mu) \quad = \quad \begin{pmatrix} -\mu \nabla_{yy}^2 \ell^* + c_y^{*\top} c_y^* & -\mu \nabla_{yy}^2 \ell^* c_y^{*-1} c_u^* + c_y^{*\top} c_u^* & -\mu I & 0 \\ & \nabla_{uu}^2 \ell^* - c_u^{*\top} c_y^{*-\top} \nabla_{yu}^2 \ell^* & & \\ 0 & +c_u^{*\top} c_y^{*-\top} \nabla_{yy}^2 \ell^* c_y^{*-1} c_u^* & c_u^{*\top} c_y^{*-\top} & -I \\ & -\nabla_{uy}^2 \ell^* c_y^{*-1} c_u^* & & \\ \bar{Z}_y^* & 0 & Y^* & 0 \\ 0 & \bar{Z}_u^* & 0 & U^* \end{pmatrix}$$

because the cross derivative with respect to y and u vanishes:

$$\nabla_{uy}^2 \ell^* - c_u^{*\top} c_y^{*-\top} \nabla_{yy}^2 \ell^* + c_u^{*\top} c_y^{*-\top} \nabla_{yy}^2 \ell^* - \nabla_{uy}^2 \ell^* \quad = \quad 0 \,.$$

We mean $J^*(0,0,\mu) \equiv J(x^*, \bar{z}^*; 0, 0, \mu)$. When $\mu = 0$, $J^*(0,0,\mu)$ is just

$$J^*(0,0,0) \;=\; \begin{pmatrix} c_y^{*\top} c_y^* & c_y^{*\top} c_u^* & 0 \\ 0 & w^{*\top} \nabla_{xx}^2 \ell^* w^* & -w^{*\top} \\ \bar{Z}_y^* & 0 & X^* \\ 0 & \bar{Z}_u^* & \end{pmatrix}.$$

The nonsingularity of this matrix is a direct consequence of the nonsingularity of the matrix

$$\begin{pmatrix} \nabla c^{*\top} & 0 & 0 \\ \nabla_{xx}^2 \ell^* & \nabla c^* & -I \\ \bar{Z}^* & 0 & X^* \end{pmatrix},$$

which in turn results from the assumptions on $(y^*, u^*, \bar{z}_y^*, \bar{z}_u^*)$.

The Jacobian $J^*(0,0,\mu)$ is also nonsingular for $\mu \in (0, \mu^*]$. In fact, let $(\Delta y, \Delta u, \Delta z_y, \Delta z_u)$ be a solution of the homogeneous linear system with the matrix $J^*(0,0,\mu)$:

$$\left(-\mu \nabla_{yy}^2 \ell^* + c_y^{*\top} c_y^*\right) \Delta y + \left(-\mu \nabla_{yy}^2 \ell^* + c_y^{*\top} c_y^*\right) c_y^{*-1} c_u^* \Delta u - \mu \Delta z_y \;=\; 0, \tag{31}$$

$$w^{*\top} \nabla_{xx}^2 \ell^* w^* \Delta u + c_u^{*\top} c_y^{*-\top} \Delta z_y - \Delta z_u \;=\; 0, \tag{32}$$

$$\bar{Z}_y^* \Delta y + Y^* \Delta z_y \;=\; 0, \tag{33}$$

$$\bar{Z}_u^* \Delta u + U^* \Delta z_u \;=\; 0. \tag{34}$$

Equations (33)-(34) and strict complementarity between x^* and $\bar{z}^*$ imply $\Delta y^\top \Delta z_y = \Delta u^\top \Delta z_u = 0$. Multiplying (32) on the left by $\Delta u^\top$ yields

$$\Delta u^\top w^{*\top} \nabla_{xx}^2 \ell^* w^* \Delta u + \Delta u^\top c_u^{*\top} c_y^{*-\top} \Delta z_y \;=\; 0.$$

Using (31) to eliminate $\Delta u^\top c_u^{*\top} c_y^{*-\top}$, this last equation is equivalent to

$$\Delta u^\top w^{*\top} \nabla_{xx}^2 \ell^* w^* \Delta u - \Delta y^\top \Delta z_y + \mu \Delta z_y^\top \left(-\mu \nabla_{yy}^2 \ell^* + c_y^{*\top} c_y^*\right)^{-1} \Delta z_y \;=\; 0.$$

From (33)-(34) and the assumptions on $(x^*, \bar{z}^*)$, we know that $w^{*\top} \nabla_{xx}^2 \ell^* w^*$ is positive definite for all Δu such that $(\Delta u)_i = 0$ if $u_i^* = 0$. From the choice of μ_*, $-\mu \nabla_{yy}^2 \ell^* + c_y^{*\top} c_y^*$ is also positive definite. Thus, we conclude that $\Delta u = 0$ and $\Delta z_y = 0$. From (31) and the choice of μ_*, $\Delta y = 0$. Finally, from (32), $\Delta z_u = 0$. The conclusion is that the Jacobian $J^*(0,0,\mu)$ is nonsingular for $\mu \in (0, \mu^*]$.

We now apply the implicit function theorem [1, pp. 12] to the system (28)-(30). We identify the set $K = \{0\} \times \{0\} \times [0, \mu^*]$ with the compact set $\bar{X}$ of that theorem. The consequence is that there exist positive scalars ϵ and δ and unique continuously differentiable functions $\hat{y}(r,s,\mu)$, $\hat{u}(r,s,\mu)$, $\hat{z}_y(r,s,\mu)$, and $\hat{z}_u(r,s,\mu)$ defined on a neighborhood $B(K,\delta) = \{(r,s,\mu) : \|(r,s,\mu) - (0,0,\mu')\| < \delta \text{ for some } (0,0,\mu') \in K\}$ of K satisfying

$$r - \mu \nabla_{yy}^2 \ell(\hat{x}(r,s,\mu), \bar{z}_y^* + s) c_y(\hat{x}(r,s,\mu))^{-1} c(\hat{x}(r,s,\mu))$$
$$+ c_y(\hat{x}(r,s,\mu))^\top c(\hat{x}(r,s,\mu)) + \mu \bar{z}_y^* - \mu \hat{z}_y(r,s,\mu) \;=\; 0, \tag{35}$$

$$\nabla_u f(\hat{x}(r,s,\mu)) - c_u(\hat{x}(r,s,\mu))^\top c_y^{-\top}(\hat{x}(r,s,\mu)) \left(\nabla_y f(\hat{x}(r,s,\mu))\right.$$
$$\left. - \nabla_{yy}^2 \ell(\hat{x}(r,s,\mu), \bar{z}_y^* + s) c_y(\hat{x}(r,s,\mu))^{-1} c(\hat{x}(r,s,\mu)) - \hat{z}_y(r,s,\mu)\right)$$
$$- \nabla_{uy}^2 \ell(\hat{x}(r,s,\mu), \bar{z}_y^* + s) c_y(\hat{x}(r,s,\mu))^{-1} c(\hat{x}(r,s,\mu)) - \hat{z}_u(r,s,\mu) \;=\; 0, \tag{36}$$

$$\hat{X}(r,s,\mu) \hat{Z}(r,s,\mu) e \;=\; 0, \tag{37}$$

and such that

$$\left\| \begin{pmatrix} \hat{x}(r,s,\mu) - x^* \\ \hat{z}_y(r,s,\mu) - \bar{z}_y^* \\ \hat{z}_u(r,s,\mu) - \bar{z}_u^* \end{pmatrix} \right\| \leq \epsilon$$

for all $(r,s,\mu) \in B(K,\delta)$. Making use of (37) and strict complementarity of the pair $(x^*, \bar{z}^*)$, and reducing ϵ and δ if necessary, one can easily show that:

(i) $\hat{x}(r,s,\mu) \geq 0$, $\hat{z}(r,s,\mu) \geq 0$,

(ii) the pair $(\hat{x}(r,s,\mu), \hat{z}(r,s,\mu))$ also satisfies strict complementarity,

(iii) the gradients of the active constraints are linearly independent at $\hat{x}(r,s,\mu)$,

for all $(r,s,\mu) \in B(K,\delta)$.

To derive the bounds (18)-(21), we differentiate $\hat{x}(r,s,\mu)$ and $\hat{z}(r,s,\mu)$ with respect to (r,s,μ), and write

$$\begin{pmatrix} \nabla_r \hat{x}(r,s,\mu)^\top & \nabla_s \hat{x}(r,s,\mu)^\top & \nabla_\mu \hat{x}(r,s,\mu)^\top \\ \nabla_r \hat{z}_y(r,s,\mu)^\top & \nabla_s \hat{z}_y(r,s,\mu)^\top & \nabla_\mu \hat{z}_y(r,s,\mu)^\top \\ \nabla_r \hat{z}_u(r,s,\mu)^\top & \nabla_s \hat{z}_u(r,s,\mu)^\top & \nabla_\mu \hat{z}_u(r,s,\mu)^\top \end{pmatrix} = -J(r,s,\mu)^{-1} B(r,s,\mu),$$

where the Jacobian of the vector function given in (28)-(30) at the pair $(\hat{x}(r,s,\mu), \hat{z}(r,s,\mu))$, and in the situation where the parameters r, s, and μ are not necessarily zero, is given by

$$J(r,s,\mu) \equiv J(\hat{x}(r,s,\mu), \hat{z}(r,s,\mu); r,s,\mu)$$

$$= \begin{pmatrix} -\mu\nabla_{yy}^2\ell + c_y^\top c_y & -\mu\nabla_{yy}^2\ell c_y^{-1}c_u + c_y^\top c_u & -\mu I & 0 \\ 0 & \begin{matrix} \nabla_{uu}^2\ell - c_u^\top c_y^{-\top}\nabla_{yu}^2\ell \\ +c_u^\top c_y^{-\top}\nabla_{yy}^2\ell c_y^{-1}c_u \\ -\nabla_{uy}^2\ell c_y^{-1}c_u \end{matrix} & c_u^\top c_y^{-\top} & -I \\ Z_y & 0 & Y & 0 \\ 0 & Z_u & 0 & U \end{pmatrix}$$

$$+ \begin{pmatrix} \begin{matrix} -\mu\sum_{j=1}^{n_y}(c_y^{-1}c)_j\nabla_y\left(\nabla_{yy}^2\ell\right)_j \\ +\sum_{j=1}^{n_y}c_j\nabla_{yy}^2 c_j + \mu\nabla_{yy}^2\ell(I - \nabla_y s) \end{matrix} & \begin{matrix} -\mu\sum_{j=1}^{n_y}(c_y^{-1}c)_j\nabla_u\left(\nabla_{yy}^2\ell\right)_j \\ +\sum_{j=1}^{n_y}c_j\nabla_{yu}^2 c_j + \mu\nabla_{yy}^2\ell(c_y^{-1}c_u - \nabla_u s) \end{matrix} & 0 & 0 \\ \begin{matrix} \sum_{j=1}^{n_y}\left[(c_y^{-1}c)_j \right. \\ \left. \nabla_y\left(c_u^\top c_y^{-\top}\nabla_{yy}^2\ell\right)_j\right] \\ -\sum_{j=1}^{n_y}(c_y^{-1}c)_j\nabla_y\left(\nabla_{uy}^2\ell\right)_j \\ -c_u^\top c_y^{-\top}\nabla_{yy}^2\ell(I - \nabla_y s) \\ +\nabla_{uy}^2\ell(I - \nabla_y s) \end{matrix} & \begin{matrix} \sum_{j=1}^{n_y}\left[(c_y^{-1}c)_j \right. \\ \left. \nabla_u\left(c_u^\top c_y^{-\top}\nabla_{yy}^2\ell\right)_j\right] \\ -\sum_{j=1}^{n_y}(c_y^{-1}c)_j\nabla_u\left(\nabla_{uy}^2\ell\right)_j \\ -c_u^\top c_y^{-\top}\nabla_{yy}^2\ell(c_y^{-1}c_u - \nabla_u s) \\ +\nabla_{uy}^2\ell(c_y^{-1}c_u - \nabla_u s) \end{matrix} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The matrix B is defined by

$$B(r,s,\mu) = \begin{pmatrix} I & B_{12}(r,s,\mu) & B_{13}(r,s,\mu) \\ 0 & B_{22}(r,s,\mu) & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

with

$$
\left(B_{12}(r, s, \mu) \right)_k = -\mu \left[L_{12}(r, s, \mu) \right]_{s_k} c_y(\hat{x}(r, s, \mu))^{-1} c(\hat{x}(r, s, \mu)) ,
$$

$$
B_{13}(r, s, \mu) = -\nabla_{yy}^2 \ell(\hat{x}(r, s, \mu)) c_y(\hat{x}(r, s, \mu))^{-1} c(\hat{x}(r, s, \mu)) + \bar{z}_y^* - \hat{z}_y(r, s, \mu) ,
$$

$$
\left(B_{22}(r, s, \mu) \right)_k = -c_u(\hat{x}(r, s, \mu))^\top c_y(\hat{x}(r, s, \mu))^{-\top} \left[L_{12}(r, s, \mu) \right]_{s_k} c_y(\hat{x}(r, s, \mu))^{-1} c(\hat{x}(r, s, \mu))
$$
$$
- \left[L_{22}(r, s, \mu) \right]_{s_k} c_y(\hat{x}(r, s, \mu))^{-1} c(\hat{x}(r, s, \mu)) .
$$

In the expressions for $B_{12}(r, s, \mu)$ and $B_{22}(r, s, \mu)$, the index k goes through the n_y columns of these matrices, the subscript s_k denotes derivative with respect to s_k, and

$$
L_{12}(r, s, \mu) = \sum_{j=1}^{n_y} \nabla_{yy}^2 c_j(\hat{x}(r, s, \mu)) \left(c_y(\hat{x}(r, s, \mu))^{-\top} s \right)_j ,
$$
$$
L_{22}(r, s, \mu) = \sum_{j=1}^{n_y} \nabla_{uy}^2 c_j(\hat{x}(r, s, \mu)) \left(c_y(\hat{x}(r, s, \mu))^{-\top} s \right)_j .
$$

Hence,

$$
\begin{pmatrix} \hat{x}(r, s, \mu) - x^* \\ \hat{z}_y(r, s, \mu) - \bar{z}_y^* \\ \hat{z}_u(r, s, \mu) - \bar{z}_u^* \end{pmatrix} = \begin{pmatrix} \hat{x}(r, s, \mu) - \hat{x}(0, 0, 0) \\ \hat{z}_y(r, s, \mu) - \hat{z}_y(0, 0, 0) \\ \hat{z}_u(r, s, \mu) - \hat{z}_u(0, 0, 0) \end{pmatrix}
$$
$$
= -\int_0^1 J(\tau r, \tau s, \tau \mu)^{-1} B(\tau r, \tau s, \tau \mu) \begin{pmatrix} r \\ s \\ \mu \end{pmatrix} d\tau . \tag{38}
$$

Since $J^*(0, 0, \mu)$ is nonsingular for all $\mu \in [0, \mu^*]$, we can show that for ϵ and δ sufficiently small $J(r, s, \mu)^{-1}$ is bounded on

$$
\{(r, s, \mu) : \|(r, s)\| < \delta , \ \mu \in [0, \mu^*]\} \subset B(K, \delta) .
$$

In fact, it is quite clear from the continuity assumptions that the first matrix term of $J(r, s, \mu)$ is a perturbation of size δ and ϵ of $J^*(0, 0, 0)$. Furthermore, from (35) we write

$$
c(\hat{x}(r, s, \mu)) = c_y(\hat{x}(r, s, \mu)) \left(-\mu \nabla_{yy}^2 \ell(\hat{x}(r, s, \mu) + \right.
$$
$$
\left. c_y(\hat{x}(r, s, \mu))^\top c_y(\hat{x}(r, s, \mu)) \right)^{-1} \left(-r + \mu \hat{z}_y(r, s, \mu) - \mu \bar{z}_y^* \right) ,
$$

which enables us to say that the second matrix term of $J(r, s, \mu)$ is $\mathcal{O}(\delta, \epsilon)$.

Now we can finally show (18)-(21). Since the integral in (38) is bounded above by

$$
\max_{\tau \in [0,1]} \| J(\tau r, \tau s, \tau \mu)^{-1} \| \int_0^1 B(\tau r, \tau s, \tau \mu) \begin{pmatrix} r \\ s \\ \mu \end{pmatrix} d\tau ,
$$

there exist positive constants κ_5-κ_9 such that

$$
\| \hat{x}(r, s, \mu) - x^* \| + \| \hat{z}(r, s, \mu) - \bar{z}^* \| \leq \kappa_5 \|r\| + \kappa_6 \mu \|s\| d(r, s, \mu) + \kappa_7 \mu d(r, s, \mu)
$$
$$
+ \kappa_8 \mu \max_{\tau \in [0,1]} \| \hat{z}_y(\tau r, \tau s, \tau \mu) - \bar{z}_y^* \|
$$
$$
+ \kappa_9 \|s\| d(r, s, \mu) ,
$$

where

$$
d(r, s, \mu) = \max_{\tau \in [0,1]} \| c_y(\hat{x}(\tau r, \tau s, \tau \mu))^{-1} c(\hat{x}(\tau r, \tau s, \tau \mu)) \| .
$$

We developed this bound applying the continuity assumptions to the four terms that appeared in $B_{12}(r, s, \mu)$, $B_{13}(r, s, \mu)$, and $B_{22}(r, s, \mu)$. However, from equation (35), the choice of μ^*, and the continuity assumptions, we can derive

$$
d(r, s, \mu) \leq \kappa_{10} \|r\| + \kappa_{11} \mu \max_{\tau \in [0,1]} \| \hat{z}_y(\tau r, \tau s, \tau \mu) - \bar{z}_y^* \| ,
$$

for some positive constants κ_{10} and κ_{11}. Thus, since $\mu \leq \mu^*$ and $\|s\| < \delta$, there exist constants $\kappa_{12}, \kappa_{13} > 0$ such that

$$\|\hat{x}(r,s,\mu) - x^*\| + \|\hat{z}(r,s,\mu) - \bar{z}^*\| \ \leq \ \kappa_{12}\|r\| + \kappa_{13}\mu \max_{\tau \in [0,1]} \|\hat{z}_y(\tau r, \tau s, \tau\mu) - \bar{z}_y^*\|,$$

from which we get for (r,s,μ) replaced by $(\tau r, \tau s, \tau\mu)$

$$\|\hat{z}_y(\tau r, \tau s, \tau\mu) - \bar{z}_y^*\| \ \leq \ \kappa_{12}\|r\| + \kappa_{13}\mu \max_{\beta \in [0,1]} \|\hat{z}_y(\beta\tau r, \beta\tau s, \beta\tau\mu) - \bar{z}_y^*\|.$$

Thus

$$\max_{\tau \in [0,1]} \|\hat{z}_y(\tau r, \tau s, \tau\mu) - \bar{z}_y^*\| \ \leq \ \frac{\kappa_{12}}{1 - \kappa_{13}\mu}\|r\|,$$

for $\mu \in [0, \bar{\mu}]$ with $\bar{\mu} < \min\left\{\mu^*, \frac{1}{\kappa_{13}}\right\}$. Therefore

$$
\begin{aligned}
\|\hat{x}(r,s,\mu) - x^*\| + \|\hat{z}(r,s,\mu) - \bar{z}^*\| &\leq \left(\kappa_{12} + \frac{\kappa_{12}\kappa_{13}\mu}{1 - \kappa_{13}\mu}\right)\|r\| \\
&\leq \left(\kappa_{12} + \frac{\kappa_{12}\kappa_{13}\bar{\mu}}{1 - \kappa_{13}\bar{\mu}}\right)\mu\|z_y - \bar{z}_y^*\|.
\end{aligned}
\tag{39}
$$

For $\mu \in (0, \mu^*]$ and $\|z_y - \bar{z}_y^*\| < \min\left\{\delta, \frac{\delta}{\mu}\right\}$ let us define

$$
\begin{aligned}
\tilde{y}(z_y, \mu) &= \hat{y}(r, s, \mu) &= \hat{y}(\mu(z_y - \bar{z}_y^*), z_y - \bar{z}_y^*, \mu), \\
\tilde{u}(z_y, \mu) &= \hat{u}(r, s, \mu) &= \hat{u}(\mu(z_y - \bar{z}_y^*), z_y - \bar{z}_y^*, \mu), \\
\tilde{z}_y(z_y, \mu) &= \hat{z}_y(r, s, \mu) &= \hat{z}_y(\mu(z_y - \bar{z}_y^*), z_y - \bar{z}_y^*, \mu), \\
\tilde{z}_u(z_y, \mu) &= \hat{z}_u(r, s, \mu) &= \hat{z}_u(\mu(z_y - \bar{z}_y^*), z_y - \bar{z}_y^*, \mu).
\end{aligned}
\tag{40}
$$

Hence, the bounds (18)-(21) follow immediately from (39).

We end the proof by showing that $\tilde{x}(z_y, \mu)$ is the solution of problem (16). First we point out that $(\tilde{x}(z_y, \mu), \tilde{z}(z_y, \mu))$ satisfies the first-order necessary conditions for (16) as it can be seen by rewriting system (35)-(37) using the changes of variables (26)-(27) and (40). The first equation of the first-order necessary conditions is

$$
\begin{aligned}
z_y \ - \ & \nabla_{yy}^2 \ell(\tilde{x}(z_y, \mu), z_y)c_y(\tilde{x}(z_y, \mu))^{-1}c(\tilde{x}(z_y, \mu)) \\
+ \ & \tfrac{1}{\mu}c_y(\tilde{x}(z_y, \mu))^\top c(\tilde{x}(z_y, \mu)) \ - \ \tilde{z}_y(z_y, \mu) \ = \ 0
\end{aligned}
\tag{41}
$$

and (22) is clearly true. Then, we show that the Hessian of $P(x; z_y, \mu)$ is positive definite at $\tilde{x}(z_y, \mu)$ for all vectors

$$(\Delta y)_i = 0 \quad \text{if} \quad \tilde{y}(z_y, \mu)_i = 0 \quad \text{and} \quad (\tilde{z}_y(z_y, \mu))_i > 0, \tag{42}$$

$$(\Delta u)_i = 0 \quad \text{if} \quad \tilde{u}(z_y, \mu)_i = 0 \quad \text{and} \quad (\tilde{z}_u(z_y, \mu))_i > 0. \tag{43}$$

The cases $(\Delta y)_i \geq 0$ and $(\Delta u)_i \geq 0$ are eliminated because the pair $(\tilde{x}(z_y, \mu), \tilde{z}(z_y, \mu))$ is strictly complementary (see ii above). The scalar ϵ can be chosen sufficiently small so that we can consider

$$(\Delta y)_i = 0 \quad \text{if} \quad y_i^* = 0 \quad \text{and} \quad (\bar{z}_y^*)_i > 0,$$

$$(\Delta u)_i = 0 \quad \text{if} \quad u_i^* = 0 \quad \text{and} \quad (\bar{z}_u^*)_i > 0.$$

This means that we can check the positive definiteness of $\nabla^2 P(x; z_y, \mu)$ in the same subspace that we consider for $\nabla^2 P(x^*; \bar{z}_y^*, \mu)$. Moreover, we proved in Theorem 4.4 that $\nabla^2 P(x^*; \bar{z}_y^*, \mu)$ is positive definite for $\mu \in (0, \mu^*]$ in the above mentioned subspace. To achieve our goal we show that the Hessian of P at $(\tilde{x}(z_y, \mu); z_y, \mu)$ is

a perturbation of size ϵ and δ of the Hessian $\nabla^2 P(x^*; \bar{z}_y^*, \mu)$. In fact, $\nabla^2 P(\tilde{x}(z_y, \mu); z_y, \mu)$ is given by:

$$
\begin{pmatrix}
-\nabla_{yy}^2 \ell + \frac{1}{\mu} c_y^\top c_y & -\nabla_{yy}^2 \ell \, c_y^{-1} c_u + \frac{1}{\mu} c_y^\top c_u \\
-c_u^\top c_y^{-\top} \nabla_{yy}^2 \ell + \frac{1}{\mu} c_u^\top c_y & -c_u^\top c_y^{-\top} \nabla_{yu}^2 \ell + \nabla_{uu}^2 \ell - \nabla_{uy}^2 \ell \, c_y^{-1} c_u + \frac{1}{\mu} c_u^\top c_u
\end{pmatrix} -
$$

$$
\begin{pmatrix}
\sum_{j=1}^{n_y} (c_y^{-1} c)_j \nabla_y \left(\nabla_{yy}^2 \ell \right)_j - \frac{1}{\mu} c_j \nabla_{yy}^2 c_j & \sum_{j=1}^{n_y} (c_y^{-1} c)_j \nabla_u \left(\nabla_{yy}^2 \ell \right)_j - \frac{1}{\mu} c_j \nabla_{yu}^2 c_j \\
\sum_{j=1}^{n_y} (c_y^{-1} c)_j \nabla_y \left(\nabla_{uy}^2 \ell \right)_j - \frac{1}{\mu} c_j \nabla_{uy}^2 c_j & \sum_{j=1}^{n_y} (c_y^{-1} c)_j \nabla_u \left(\nabla_{uy}^2 \ell \right)_j - \frac{1}{\mu} c_j \nabla_{uu}^2 c_j
\end{pmatrix} +
$$

$$
\begin{pmatrix}
\nabla_{yy}^2 \ell (I - \nabla_y s) & \nabla_{yy}^2 \ell (c_y^{-1} c_u - \nabla_u s) \\
\nabla_{uy}^2 \ell (I - \nabla_y s) & \nabla_{uy}^2 \ell (c_y^{-1} c_u - \nabla_u s)
\end{pmatrix}
$$

with the Lagrangian evaluated at $(\tilde{x}(z_y, \mu), z_y)$ and the remaining functions at $\tilde{x}(z_y, \mu)$. It is quite clear that the first matrix term is a perturbation of size ϵ and δ of the Hessian $\nabla^2 P(x^*; \bar{z}_y^*, \mu)$. To bound the second matrix term we rewrite (41) as

$$
\frac{1}{\mu} c(\tilde{x}(z_y, \mu)) = c_y(\tilde{x}(z_y, \mu)) \left(-\mu \nabla_{yy}^2 \ell(\tilde{x}(z_y, \mu), z_y) + \\
c_y(\tilde{x}(z_y, \mu))^\top c_y(\tilde{x}(z_y, \mu)) \right)^{-1} (\tilde{z}_y(z_y, \mu) - z_y).
$$

Thus, using the continuity assumptions and adding and subtracting $\bar{z}_y^*$, we obtain

$$
\begin{aligned}
\left\| \frac{1}{\mu} c(\tilde{x}(z_y, \mu)) \right\| &\leq \kappa_{14} \left(\|\tilde{z}_y(z_y, \mu) - \bar{z}_y^*\| + \|z_y - \bar{z}_y^*\| \right) \\
&\leq \kappa_{14}(\epsilon + \delta)
\end{aligned}
$$

and

$$
\|c(\tilde{x}(z_y, \mu))\| \leq \bar{\mu} \kappa_{14}(\epsilon + \delta).
$$

The conclusion is that the second and third matrix terms are also $\mathcal{O}(\delta, \epsilon)$ and the proof that the Hessian of $P(y, u; z_y, \mu)$ is positive definite for all vectors Δy and Δu satisfying (42)-(43) is terminated. $\quad\square$

Using Theorem 5.1 we can state the basic properties of local convergence of the multipliers method 5.1.

Corollary 5.1 *Let $x^* = (y^*, u^*)$ with corresponding multipliers $\bar{z}^* = (\bar{z}_y^*, \bar{z}_u^*)$ satisfy Assumptions 2.1-5.1.*

There exist scalars $\delta_0 \in (0, \delta]$, $\kappa \in (0, 1)$, and $\mu_0 \in (0, \bar{\mu}]$ such that if the sequence $\{\mu_k\}$ is monotone decreasing and $\|z_y^0 - \bar{z}_y^*\| \leq \min\{\delta_0, \delta_0/\mu_0\}$, then the sequence $\{z_y^k\}$ generated by (15) is well defined (in the sense that $(z_y^k, \mu_k) \in D^*$ for all k) and satisfies

$$
\limsup_{k \to +\infty} \frac{\|z_y^{k+1} - \bar{z}_y^*\|}{\|z_y^k - \bar{z}_y^*\|} \leq \kappa \tag{44}
$$

when $\liminf_{k \to +\infty} \mu_k > 0$, and

$$
\lim_{k \to +\infty} \frac{\|z_y^{k+1} - \bar{z}_y^*\|}{\|z_y^k - \bar{z}_y^*\|} = 0 \tag{45}
$$

when $\lim_{k \to +\infty} \mu_k = 0$. In both cases, we have

$$
\lim_{k \to +\infty} \tilde{x}(z_y^k, \mu_k) = x^*, \qquad \lim_{k \to +\infty} \tilde{z}_u(z_y^k, \mu_k) = \bar{z}_u^*, \tag{46}
$$

$$
\lim_{k \to +\infty} z_y^k = \bar{z}_y^*. \tag{47}
$$

Proof: The limits (44), (45), and (47) follow from inequality (20). The limits (46) are a consequence of (18), (19), and (21). $\quad\square$

We point out that the constant κ in (44) depends on the condition number of c_y^*, as we have seen in the proof of Theorem 5.1.

6   Dual interpretation of the multipliers method

In the context of Theorem 5.1, we introduce the dual function

$$d_\mu(z_y) \;=\; \min_{x \in B(x^*, \epsilon)} \; P(x; z_y, \mu)\,, \tag{48}$$

for $(z_y, \mu) \in D^*$, where z_y are the dual variables and μ is a positive parameter. From Theorem 5.1, we know that

$$d_\mu(z_y) \;=\; P(\tilde{x}(z_y, \mu); z_y, \mu)\,, \tag{49}$$

where $\tilde{x}(z_y, \mu)$ and the corresponding multipliers $\tilde{z}(z_y, \mu)$ are continuously differentiable functions satisfying

$$\nabla_x P(\tilde{x}(z_y, \mu); z_y, \mu) - \tilde{z}(z_y, \mu) \;=\; 0\,, \tag{50}$$

$$(\tilde{x}(z_y, \mu))_i (\tilde{z}(z_y, \mu))_i \;=\; 0\,, \; i = 1, \ldots, n_y + n_u\,. \tag{51}$$

We now need to introduce the following notation

$$\mathcal{B} \;=\; \{i : x_i^* > 0\}\,, \qquad \mathcal{N} \;=\; \{i : x_i^* = 0\}\,.$$

Differentiating (49) and (51) with respect to z_y and using (50)

$$
\begin{aligned}
\nabla_{z_y} d_\mu(z_y) \;&=\; \nabla_{z_y} \tilde{x}(z_y, \mu) \nabla_x P(\tilde{x}(z_y, \mu); z_y, \mu) + c_y(\tilde{x}(z_y, \mu))^{-1} c(\tilde{x}(z_y, \mu)) \\
&=\; \nabla_{z_y} \tilde{x}(z_y, \mu) \tilde{z}(z_y, \mu) + c_y(\tilde{x}(z_y, \mu))^{-1} c(\tilde{x}(z_y, \mu)) \\
&=\; c_y(\tilde{x}(z_y, \mu))^{-1} c(\tilde{x}(z_y, \mu))\,.
\end{aligned} \tag{52}
$$

Differentiating (52) with respect to z_y and (50) with respect to x and appealing to strict complementarity

$$
\begin{aligned}
\nabla_{z_y z_y}^2 d_\mu(z_y) \;&=\; \nabla_{z_y} \tilde{x}(z_y, \mu)_\mathcal{B} C(\tilde{x}(z_y, \mu))_\mathcal{B} \\
&=\; -C(\tilde{x}(z_y, \mu))_\mathcal{B}^\top \nabla_{xx}^2 P(\tilde{x}(z_y, \mu); z_y, \mu)_\mathcal{B}^{-1} C(\tilde{x}(z_y, \mu))_\mathcal{B}\,,
\end{aligned} \tag{53}
$$

where

$$C(x) \;=\; \begin{pmatrix} I \\ c_u(x)^\top c_y(x)^{-\top} \end{pmatrix} - \begin{pmatrix} \sum_{i=1}^{n_y} (c_y(x)^{-1} c(x))_i \nabla_{yy}^2 c_i(x)\, c_y(x)^{-\top} \\ \sum_{i=1}^{n_y} (c_y(x)^{-1} c(x))_i \nabla_{uy}^2 c_i(x)\, c_y(x)^{-\top} \end{pmatrix}\,.$$

If we neglect the terms in (53) involving $c(\tilde{x}(z_y, \mu))$, the Hessian $\nabla_{z_y z_y}^2 d_\mu(z_y)$ reduces to

$$-c_y^{-1} \nabla c_\mathcal{B}^\top \nabla_{xx}^2 P_\mathcal{B}^{-1} \nabla c_\mathcal{B} c_y^{-\top} \;=\; -(c_y^\top c_y)^{-1} \left[c_y^\top \nabla c_\mathcal{B}^\top \nabla_{xx}^2 P_\mathcal{B}^{-1} \nabla c_\mathcal{B} c_y^{-\top} \right]\,.$$

Locally (around x^*) and for sufficiently large values of μ, solving the dual problem

$$\max_{z_y} \quad d_\mu(z_y)$$

reduces to solve problem (1). The following corollary of Theorem 5.1 states this relationship formally. The proof follows from the form (52) for the gradient of the dual function and from Theorems 4.2 and 4.3.

Corollary 6.1 *Let $x^* = (y^*, u^*)$ with corresponding multipliers $\bar{z}^* = (\bar{z}_y^*, \bar{z}_u^*)$ satisfy Assumptions 2.1-5.1.*
 If for $(z_y^d, \mu_d) \in D^*$, z_y^d is a stationary point for $d_{\mu_d}(\cdot)$, i.e., if $\nabla_{z_y} d_{\mu_d}(z_y^d) = 0$, then the pair $(\tilde{x}(z_y^d, \mu_d),$ $\tilde{z}(z_y^d, \mu_d))$ satisfies the second-order sufficient conditions for problem (1).

The gradient of the dual function provides an interpretation of the multipliers method 5.1. In fact, the steepest ascent method to maximize $d_\mu(\cdot)$ is of the form

$$z_y^{k+1} \;=\; z_y^k + \rho_k (c_y^{-1})_k c_k$$

for some search parameter ρ_k, while the multipliers method update (15) can be rewritten as

$$z_y^{k+1} \;=\; z_y^k + \frac{1}{\mu_k} \left(-\mu_k \nabla_{yy}^2 \ell_k + (c_y)_k^\top (c_y)_k \right) (c_y^{-1})_k c_k \,.$$

As in [1, Section 2.3.2], a q-quadratically convergent multipliers method can be derived by applying Newton's method to the dual function $d_\mu(\cdot)$:

$$z_y^{k+1} \;=\; z_y^k - \nabla_{z_y z_y}^2 d_{\mu_k}(z_y^k)^{-1} \nabla_{z_y} d_{\mu_k}(z_y^k)$$

with $(z_y^k, \mu_k) \in D^*$.

7   Conclusions and future work

The nonlinear programming problems (1) and (2) often arise from the discretization of optimal control problems (see references [2, 3, 4, 8, 11, 12, 13, 14, 15, 19, 21, 20, 22, 23]).

The presence of bounds on the states variables makes the problem particularly difficult but also with wider applicability. The contribution of this paper was to set a framework to solve (1) and (2) based on a new multipliers method motivated in turn by a penalty function with some exactness properties. The method explores the structure of the problem, requiring the solution of linearized state and adjoint linear systems, and can be implemented using either adjoints or sensitivities [17].

The application of the multipliers method to solve the general nonlinear programming problem is the subject of a forthcoming paper that will address the local analysis, a globalization scheme, and numerical results. The use of the least squares multipliers (and the use of the orthogonal null space basis in the analysis) in general nonlinear programming raises a number of questions about the efficiency of our approach in such general setting. Our new multipliers method is perhaps better tailored to problems of the form (1) and (2) where the null space basis (5) and the adjoint multipliers appear naturally in the problem structure.

Other topics of future research are: number of inner and outer iterations [7] and the identification of active constraints; local analysis under weaker assumptions like degeneracy [29] and lack of strict complementarity [16].

8   Appendix

The point $x = (y, u)$ is a regular or nondegenerate point for problem (1) if the gradients of the active constraints are linearly independent, i.e., if the matrix

$$\begin{pmatrix} c_y(y, u) & c_u(y, u) \\ I_\mathcal{Y} & 0 \\ 0 & I_\mathcal{U} \end{pmatrix}$$

has full row rank, where $I_\mathcal{Y}$ (respectively $I_\mathcal{U}$) is the submatrix of the identity corresponding to indices i such that $y_i = 0$ (respectively $u_i = 0$).

A point (y, u) satisfies the first-order necessary conditions for problem (1) if there exist $\lambda \in \mathbb{R}^{n_y}$ and $(z_y, z_u) \in \mathbb{R}^{n_y + n_u}$ such that

$$\nabla_y \ell(y, u, \lambda) - z_y = 0 \,, \tag{54}$$

$$\nabla_u \ell(y, u, \lambda) - z_u = 0 \,, \tag{55}$$

$$c(y, u) = 0 \,, \ (y, u) \geq 0 \,, \tag{56}$$

$$y^\top z_y \;=\; u^\top z_u \;=\; 0 \,, \ (z_y, z_u) \geq 0 \,, \tag{57}$$

where $\ell(y, u, \lambda) = f(y, u) + c(y, u)^\top \lambda$. A pair (x, z) formed by a point $x = (y, u)$ and corresponding multipliers $z = (z_y, z_u)$ satisfies strict complementarity if

$$y_i = 0 \implies (z_y)_i > 0, \ i = 1, \ldots, n_y,$$
$$u_i = 0 \implies (z_u)_i > 0, \ i = 1, \ldots, n_u.$$

The second-order necessary conditions for problem (1) are composed by (54)-(57) and the condition

$$\left(\begin{array}{c} \Delta y \\ \Delta u \end{array} \right)^\top \nabla_{xx}^2 \ell(y, u, \lambda) \left(\begin{array}{c} \Delta y \\ \Delta u \end{array} \right) \geq 0 \tag{58}$$

for all $(\Delta y, \Delta u)$ satisfying

$$\nabla c(y, u)^\top \left(\begin{array}{c} \Delta y \\ \Delta u \end{array} \right) = c_y(y, u)\Delta y + c_u(y, u)\Delta u = 0, \tag{59}$$

$(\Delta y)_i = 0$ if $y_i = 0$, and $(\Delta u)_i = 0$ if $u_i = 0$. It is easy to check that the validation of (58) on the subspace defined by (59) reduces to

$$\Delta u^\top w(y, u)^\top \nabla_{xx}^2 \ell(y, u, \lambda) w(y, u)\Delta u \geq 0.$$

The second-order sufficient conditions are the conjunction of (54)-(57) with

$$\Delta u^\top w(y, u)^\top \nabla_{xx}^2 \ell(y, u, \lambda) w(y, u)\Delta u > 0 \tag{60}$$

for all Δu such that $(\Delta u)_i = 0$ if $u_i = 0$ and $(z_u)_i > 0$, and $(\Delta u)_i \geq 0$ if $u_i = 0$ and $(z_u)_i = 0$. A point that satisfies the second-order sufficient conditions is a strict local minimizer. If the pair (x, z) is strictly complementary then (60) has to hold only for Δu such that $(\Delta u)_i = 0$ if $u_i = 0$.

Acknowledgments

References

[1] D. P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Academic Press, New York, 1982.

[2] J. BORGGAARD AND J. BURNS, *A PDE sensitivity equation method for optimal aerodynamic design*, Journal of Computational Physics, (1997), pp. 366–384.

[3] Z. CHEN AND K.-H. HOFFMANN, *Numerical solutions of the optimal control problem governed by a phase field model*, in Optimal Control of Partial Differential Equations, W. Desch, F. Kappel, and K. Kunisch, eds., Basel, Boston, Berlin, 1991, Birkhäuser Verlag.

[4] E. M. CLIFF, M. HEINKENSCHLOSS, AND A. SHENOY, *An optimal control problem for flows with discontinuities*, J. Optim. Theory Appl., 94 (1997), pp. 273–309.

[5] A. R. CONN, N. I. M. GOULD, AND P. L. TOINT, *A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds*, SIAM J. Numer. Anal., 28 (1991), pp. 545–572.

[6] ——, *LANCELOT: A Fortran Package for Large–Scale Nonlinear Optimization (Release A)*, Springer–Verlag, Berlin, Heidelberg, New York, Tokyo, 1992.

[7] ——, *On the number of inner iterations per outer iteration of a globally convergent algorithm for optimization with general equality constraints and simple bounds*, in Proceedings of the 14th Biennal Numerical Analysis Conference Dundee 1991, G. A. Watson, ed., Longmans, 1992.

[8] J. E. Dennis, M. Heinkenschloss, and L. N. Vicente, *Trust–region interior–point SQP algorithms for a class of nonlinear programming problems*, SIAM J. Control Optim., 36 (1998), pp. 1750–1794.

[9] R. Fletcher, *A class of methods for nonlinear programming with termination and convergence properties*, in Integer and nonlinear programming, J. Abadie, ed., North–Holland, Amesterdam, 1970.

[10] ——, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, second ed., 1987.

[11] A. Friedman and B. Hu, *Optimal control of a chemical vapor deposition reactor*, J. Optim. Theory Appl., 97 (1998), pp. 623–644.

[12] M. D. Gunzburger, L. S. Hou, and T. P. Svobotny, *Optimal control and optimization of viscous, incompressible flows*, in Incompressible Computational Fluid Dynamics, M. D. Gunzburger and R. A. Nicolaides, eds., Cambridge, New York, 1993, Cambridge University Press, pp. 109–150.

[13] N. Handagama and S. Lenhart, *Optimal control of a PDE/ODE system modeling a gas-phase bioreactor*, in Mathematical Models in Medical and Health Sciences, M. A. Horn, G. Simonett, and G. Webb, eds., Nashville, TN, 1998, Vanderbilt University Press.

[14] M. Heinkenschloss, *Projected sequential quadratic programming methods*, SIAM J. Optim., 6 (1996), pp. 373–417.

[15] ——, *Formulation and analysis of a sequential quadratic programming method for the optimal dirichlet boundary control of navier-stokes flow*, in Optimal Control: Theory, Algorithms, and Applications, W. W. Hager and P. M. Pardalos, eds., Kluwer Academic Publishers B.V., 1998, pp. 178–203.

[16] M. Heinkenschloss, M. Ulbrich, and S. Ulbrich, *Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumption*, Math. Programming, 86 (1999), pp. 615–635.

[17] M. Heinkenschloss and L. N. Vicente, *An interface between optimization and application for the numerical solution of optimal control problems*, ACM Trans. Math. Software, 25 (1999), pp. 157–190.

[18] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.

[19] K. Ito and K. Kunisch, *Augmented Lagrangian-SQP-methods in Hilbert spaces and application to control in the coefficients problems*, SIAM J. Optim., 6 (1996), pp. 96–125.

[20] F.-S. Kupfer and E. W. Sachs, *Numerical solution of a nonlinear parabolic control problem by a reduced SQP method*, Comput. Optim. and Appl., 1 (1992), pp. 113–135.

[21] F. Leibfritz and E. W. Sachs, *Numerical solution of parabolic state constrained control problems using SQP- and interior-point-methods*, in Large Scale Optimization: State of the Art, W. W. Hager, D. Hearn, and P. Pardalos, eds., Kluwer, 1994, pp. 251–264.

[22] J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer Verlag, Berlin, Heidelberg, New York, 1971.

[23] P. Neittaanmäki and D. Tiba, *Optimal Control of Nonlinear Parabolic Systems. Theory, Algorithms, and Applications*, Marcel Dekker, New York, Basel, Hong Kong, 1994.

[24] G. D. Pillo, *Exact penalty methods*, in Algorithms for continuous optimization, E. Spedicato, ed., Kluwer Academic Publishers, Dordrecht, 1994, pp. 1–45.

[25] G. D. Pillo and L. Grippo, *Exact penalty functions in constrained optimization*, SIAM J. Control Optim., 27 (1989), pp. 1333–1360.

[26] M. J. D. POWELL, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher, ed., Academic Press, New York, 1969, pp. 283–298.

[27] R. T. ROCKAFELLAR, *The multiplier method of Hestenes and Powell applied to convex programming*, J. Optim. Theory Appl., 12 (1973), pp. 555–562.

[28] ——, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM Journal on Control, 12 (1974), pp. 268–285.

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