UPDATING MULTIPLIERS CORRESPONDING TO INEQUALITY CONSTRAINTS IN AN AUGMENTED LAGRANGIAN MULTIPLIERS METHOD

CATARINA P. AVELINO * AND LUíS N. VICENTE †

Abstract. This paper adds to the development of the field of augmented Lagrangian multipliers methods for general nonlinear programming by introducing a new update for multipliers corresponding to inequality constraints. The update naturally maintains the nonnegativity of the multipliers without the need for a positive-orthant projection, as result of the verification of the first-order necessary conditions for the minimization of a modified augmented Lagrangian penalty function.

In the new multipliers method the roles of the multipliers are interchanged: the multipliers corresponding to inequality constraints are updated explicitly whereas the multipliers corresponding to equality constraints are implicitly approximated. It is shown that the basic properties of local convergence of the traditional multipliers method are also valid for the proposed method.

Key words. Nonlinear Programming, Multipliers Methods, Augmented Lagrangian.

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

1. Introduction. We consider the general nonlinear programming problem, in the format

$$\min\ f(x) \quad \text{s.t.} \quad h(x)=0,\ x\geq 0, \tag{1.1}$$

where $x \in \mathbb{R}^n$, the functions f and h are considered smooth and defined as $f:\Omega \longrightarrow \mathbb{R}$ and $h:\Omega \longrightarrow \mathbb{R}^m$, n and m are positive integers satisfying $n > m$, and Ω is an open set of $\mathbb{R}^n$.

The multipliers method [1, 2] is based on the augmented Lagrangian penalty function

$$L(x;\lambda,\mu) \ = \ f(x)+h(x)^\top\lambda+\frac{1}{2\mu}h(x)^\top h(x) \ = \ \ell(x,\lambda)+\frac{1}{2\mu}h(x)^\top h(x)\,,$$

where $\mu > 0$ is a penalty parameter and

$$\ell(x,\lambda)=f(x)+h(x)^\top\lambda$$

is the Lagrangian of f with respect to the equality constraints $h(x)=0$, with corresponding multipliers $\lambda \in \mathbb{R}^m$. Note that the Lagrangian term of the augmented Lagrangian penalty function involves only the equality constraints $h(x)=0$. In each outer iteration of this method, the primal iterate x_k is computed by solving the problem

$$\min\ L(x;\lambda_k,\mu_k) \quad \text{s.t.} \quad x\geq 0 \tag{1.2}$$

* Ph.D. student at Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal (cavelino@mat.uc.pt). Support for this author was provided by Departamento de Matemática da UTAD, Centro de Matemática da Universidade de Coimbra, by FCT under grant POCTI/35059/MAT/2000, and by the PRODEP 5.3 program.

† Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal (lvicente@mat.uc.pt). Support for this author was provided by Centro de Matemática da Universidade de Coimbra, by FCT under grant POCTI/35059/MAT/2000, by the European Union under grant IST-2000-26063, and by Fundação Calouste Gulbenkian.

for some λ_k and $\mu_k > 0$. The multipliers method [1] updates the multipliers λ for the next iteration by using the formula

$$\lambda_{k+1} \;=\; \lambda_k + \frac{1}{\mu_k} h(x_k)\,.$$

Thus, the multipliers λ corresponding to the equality constraints $h(x) = 0$ are updated explicitly. The nonnegative multipliers $w \in \mathbb{R}^n$, corresponding to the inequality constraints $x \geq 0$ in problem (1.1), can be implicitly approximated from the multipliers associated with $x \geq 0$ in problem (1.2), see [2].

The question addressed in this paper is the interchange of the implicit *vs* explicit roles of the multipliers in the multipliers method. It turns out that it is possible to derive a multipliers method where the multipliers w corresponding to the inequality constraints are updated explicitly — and kept nonnegative — whereas the multipliers corresponding to the equality constraints are implicitly approximated.

For this purpose let us consider the duality part of the first-order necessary optimality conditions for problem (1.1),

$$\nabla f(x) + \nabla h(x) \lambda - w = 0\,,$$

and the corresponding least-squares Lagrange multipliers estimate (when $\nabla h(x)$ has full rank),

$$\lambda(x, w) \;=\; -\left(\nabla h(x)^\top \nabla h(x)\right)^{-1} \nabla h(x)^\top (\nabla f(x) - w).$$

It is therefore possible to consider an augmented Lagrangian penalty function in the variables x, parameterized by the penalty parameter $\mu > 0$ and by the multipliers $w \geq 0$,

$$P(x; w, \mu) = f(x) + h(x)^\top \lambda(x, w) + \frac{1}{2\mu} h(x)^\top h(x),$$

and to pose the corresponding penalized problem

$$\text{(1.3)} \qquad \min \; P(x; w, \mu) \quad \text{s.t.} \quad x \geq 0.$$

Each outer iteration of the new multipliers method involves the computation of the primal variables x_k by solving the problem

$$\text{(1.4)} \qquad \min \; P(x; w_k, \mu_k) \quad \text{s.t.} \quad x \geq 0,$$

for some $w_k \geq 0$ and $\mu_k > 0$. The outer iteration provides then a formula to update w for the next iteration:

$$\text{(1.5)} \qquad w_{k+1} \;=\; \nabla P(x_k; w_k, \mu_k).$$

This formula results naturally from the first-order necessary conditions for problem (1.4) and guarantees the nonnegativity of the new multipliers estimate w_{k+1}.

In this paper we establish the local convergence properties of the new multipliers method based on (1.4) and (1.5) for general programming problems of the form (1.1). Although the analysis presented here has a lot in common with the proof of local convergence for the original multipliers method [1], several difficulties inherent to the nature of the new update had to be overcome. In particular, it is shown that the neighborhood of local convergence is smaller than in the original multipliers method, see (3.4). The new multipliers method was originally developed in [6] for nonlinear optimization problems of the form

$$\min\ f(y,u) \quad \text{s.t.} \quad c(y,u)=0,\ (y,u)\geq 0, \tag{1.6}$$

where it was assumed that the partial Jacobian of c with respect to y is square and invertible.

The paper is structured as follows. In section 2 we describe the new multipliers method for (1.1) in more detail. The local convergence properties are then presented in section 3. In section 4 we state some conclusions and comments. The proof of the main result of the local convergence analysis, stated in theorem 3.2, is given in the appendix of the paper.

2. The new multipliers method. A point x satisfies the first-order necessary optimality conditions for problem (1.1) if there exist $\lambda \in \mathbb{R}^m$ and $w \in \mathbb{R}^n$ such that

$$\nabla_x \ell(x,\lambda) - w = 0, \tag{2.1a}$$

$$h(x)=0,\ x\geq 0, \tag{2.1b}$$

$$x^\top w = 0,\ w\geq 0. \tag{2.1c}$$

Conditions (2.1a)-(2.1c) are know as the first-order Karush-Kuhn-Tucker conditions and can be written in the equivalent form

$$\begin{aligned} Z(x)^\top(\nabla f(x)-w) &= 0, \\ h(x)=0,\ x &\geq 0, \\ x^\top w = 0,\ w &\geq 0, \end{aligned}$$

where $Z(x)$ is a matrix whose columns form an orthonormal basis for the null space of $\nabla h(x)^\top$, i.e., where $Z(x)$ satisfies

$$Z(x)^\top Z(x) = I \quad \text{and} \quad \nabla h(x)^\top Z(x) = 0.$$

The matrix $Z(x)$ can be obtained from the QR factorization of $\nabla h(x)$.

Note that the matrix

$$Z(x)Z(x)^\top = I - \nabla h(x)\left(\nabla h(x)^\top \nabla h(x)\right)^{-1}\nabla h(x)^\top$$

is an orthogonal projector onto the null space $\mathcal{N}\left(\nabla h(x)^\top\right)$ of the matrix $\nabla h(x)^\top$. Similarly, $I - Z(x)Z(x)^\top$ is an orthogonal projector onto $\mathcal{R}\left(\nabla h(x)\right)$, the range space of $\nabla h(x)$.

First and second order derivatives of the penalty function P require second and third order derivatives of f and h, respectively. To establish local convergence properties we will therefore need the following assumptions that will be assumed throughout this paper.

A.1 The functions f and h are three times continuously differentiable in Ω, where Ω is an open set of $\mathbb{R}^n$. The Jacobian matrix $\nabla h(x)^\top$ of $h(x)$ has full rank in Ω.

We point out that an implementation of the multipliers method (with or without a globalization scheme) could require only first or second order derivatives.

To derive the first-order necessary conditions for problem (1.3), we need first to calculate the gradient of $P(x; w, \mu)$ with respect to x. First, we note that

$$\nabla_x \lambda(x, w) = -\left(\nabla^2_{xx} \ell(x, \lambda(x, w)) \nabla h(x) + R(x, w)^\top\right) \left(\nabla h(x)^\top \nabla h(x)\right)^{-1},$$

where the i-th row of $R(x, w)$ is given by

$$R(x, w)_i = \left(\nabla_x \ell(x, \lambda(x, w)) - w\right)^\top \nabla^2 h_i(x) \ , \quad i = 1, ..., m.$$

Thus, the gradient of $P(x; w, \mu)$ is given by

$$\nabla P(x; w, \mu) = G_1(x; w, \mu) + G_2(x; w, \mu) + G_3(x; w, \mu), \tag{2.2}$$

where

$$G_1(x; w, \mu) = Z(x) Z(x)^\top \nabla f(x) + \left(I - Z(x) Z(x)^\top\right) w,$$

$$G_2(x; w, \mu) = -\left(\nabla^2_{xx} \ell(x, \lambda(x, w)) \nabla h(x) + R(x, w)^\top\right) \left(\nabla h(x)^\top \nabla h(x)\right)^{-1} h(x),$$

$$G_3(x; w, \mu) = \frac{1}{\mu} \nabla h(x) h(x).$$

To alleviate the notation, we will omit the arguments x and x^* when it is clear from the context where the functions are evaluated. For instance, $\nabla h = \nabla h(x)$ and $\nabla f^* = \nabla f(x^*)$.

A point x satisfies the first-order necessary conditions for problem (1.3) if there exists $\bar{w} \in \mathbb{R}^n$ such that

$$ZZ^\top \nabla f + \left(I - ZZ^\top\right) w - \left(\nabla^2_{xx} \ell \nabla h + R^\top\right) \left(\nabla h^\top \nabla h\right)^{-1} h + \frac{1}{\mu} \nabla h \ h - \bar{w} = 0, \tag{2.3a}$$

$$x \geq 0, \tag{2.3b}$$

$$x^\top \bar{w} = 0, \ \bar{w} \geq 0. \tag{2.3c}$$

Equation (2.3a) provides an update formula for the multipliers corresponding to the constraints $x \geq 0$, that is the basis of the multipliers method considered in this paper.

The penalty function P, together with the penalized problem (1.3) and the equation (2.3a), suggest a new multipliers method to solve the nonlinear programming problem (1.1), which is presented below without any globalization strategy.

ALGORITHM 2.1.

Step 0. Choose initial values: μ_0 for the penalty parameter and w_0 for the approximation of the multipliers.

Step 1. For $k = 0, 1, 2, ...$ do

- **1.1** Solve problem (1.4).
- **1.2** Update the multipliers approximation:

$$w_{k+1} = ZZ^\top \nabla f + (I - ZZ^\top) w_k - \left(\nabla^2_{xx} \ell \nabla h + R^\top\right) \left(\nabla h^\top \nabla h\right)^{-1} h + \frac{1}{\mu_k} \nabla h h,$$

  where the functions ∇f, h, ∇h, and Z are evaluated at the solution $\tilde{x}(w_k, \mu_k)$ obtained in step 1.1, and the functions $\nabla^2_{xx} \ell$ and R are evaluated at $(\tilde{x}(w_k, \mu_k), \lambda(\tilde{x}(w_k, \mu_k), w_k))$.
- **1.3** Update the penalty parameter μ_{k+1}.

The local convergence analysis of the multipliers method, based on algorithm 2.1, is presented in section 3 and corresponds to the analysis given in Bertsekas [1] for the traditional augmented Lagrangian multipliers method.

3. Local convergence analysis. The study of the rate of local convergence of the multipliers method (as described in algorithm 2.1) requires second derivatives of the penalty function $P(x; w, \mu)$. One can easily show that the Hessian matrix of $P(x; w, \mu)$ is given by

$$\nabla^2 P(x; w, \mu) = H_1(x; w, \mu) + H_2(x; w, \mu) + H_3(x; w, \mu), \tag{3.1}$$

where

$$H_1(x; w, \mu) = \nabla^2_{xx} \ell Z Z^\top - R^T \left(\nabla h^\top \nabla h\right)^{-1} \nabla h^T,$$

$$H_2(x; w, \mu) = -\left(I - Z Z^\top\right) \nabla^2_{xx} \ell - \nabla h \left(\nabla h^\top \nabla h\right)^{-1} R + \sum_{i=1}^{m} h_i \nabla^2_{xx} \lambda_i,$$

$$H_3(x; w, \mu) = \frac{1}{\mu} \nabla h \nabla h^\top + \frac{1}{\mu} \sum_{i=1}^{m} h_i \nabla^2 h_i,$$

correspond to the derivatives of $G_1(x; w, \mu)$, $G_2(x; w, \mu)$, and $G_3(x; w, \mu)$ in (2.2), respectively.

We start by showing that the penalty function $P(x; w, \mu)$ exhibits some exactness properties. The result stated in the next theorem will be helpful later in the analysis of local convergence, in particular the fact that the Hessian of $P(x; w, \mu)$ is positive definite for μ in $(0, \mu^*]$, where $\mu^* > 0$ is specified later, provided that x satisfies the second-order sufficient conditions for the original problem (1.1) with multipliers $\lambda(x, w)$ and w.

THEOREM 3.1. *Let assumptions A.1 hold. If $(x, \lambda(x, w))$ satisfies the second-order necessary (resp. sufficient) conditions for the original problem (1.1), with multipliers w corresponding to $x \geq 0$, then there exists $\mu^* > 0$ such that x satisfies the second-order necessary (resp. sufficient) conditions for the penalized problem (1.3), for this w and for any $\mu \in (0, \mu^*]$.*

Proof. We start by pointing out that because the matrix

$$\begin{pmatrix} Z(x)^\top \\ \nabla h(x)^\top \end{pmatrix}$$

is nonsingular, the equation (2.3a), when $h(x) = 0$, is equivalent to

$$Z(x)^\top (\nabla f(x) - \bar{w}) = 0 \quad \text{and} \quad \nabla h(x)^\top (w - \bar{w}) = 0.$$

Thus, from the fact that x satisfies the first-order necessary conditions for the original problem (1.1) with multipliers $\lambda(x, w)$ and w, we conclude that x also satisfies the first-order necessary conditions (2.3) for the penalized problem (1.3) with multipliers $\bar{w} = w$.

Now, let us prove the result concerning the second-order sufficient conditions. For this purpose, let Δx satisfy

$$(\Delta x)_i = 0 \;\text{ if }\; x_i = 0 \;\text{ and }\; \bar{w}_i > 0, \tag{3.2a}$$

$$(\Delta x)_i \geq 0 \;\text{ if }\; x_i = 0 \;\text{ and }\; \bar{w}_i = 0. \tag{3.2b}$$

Since $h(x) = 0$ and $R(x, w) = 0$, we have

$$\Delta x^\top \nabla^2 P(x; w, \mu) \Delta x = \Delta x^\top Z Z^\top \nabla^2_{xx} \ell Z Z^\top \Delta x - \Delta x^\top (I - Z Z^\top) \nabla^2_{xx} \ell (I - Z Z^\top) \Delta x + \tfrac{1}{\mu} \Delta x^\top \nabla h \nabla h^\top \Delta x.$$

On the other hand, the second-order sufficient conditions for the original problem (1.1) say that $\nabla^2_{xx}\ell(x,\lambda(x,w))$ has to be positive definite for all vectors Δx satisfying (3.2) and $\nabla h(x)^\top \Delta x = 0$, i.e., $\Delta x = ZZ^\top \Delta x$. Thus,

$$\Delta x^\top ZZ^\top \nabla^2_{xx}\ell ZZ^\top \Delta x > 0$$

for all vectors Δx satisfying (3.2).

So, since $I - ZZ^\top = \nabla h(\nabla h^\top \nabla h)^{-1}\nabla h^\top$,

$$\begin{aligned}&\Delta x^\top \nabla^2 P(x;w,\mu)\Delta x\\ &> \Delta x^\top \nabla h \left(-(\nabla h^\top \nabla h)^{-1}\nabla h^\top \nabla^2_{xx}\ell \nabla h(\nabla h^\top \nabla h)^{-1} + \frac{1}{\mu}I\right)\nabla h^\top \Delta x,\end{aligned}$$

and the proof is completed by setting:

$$\mu^* = \begin{cases} \text{any positive real} & \text{if } (\nabla h^\top \nabla h)^{-1}\nabla h^\top \nabla^2_{xx}\ell \nabla h(\nabla h^\top \nabla h)^{-1} \\ & \text{is negative semi-definite,} \\ \frac{1}{\alpha(x,w)} & \text{otherwise,} \end{cases}$$

where $\alpha(x,w)$ is the largest eigenvalue of $(\nabla h^\top \nabla h)^{-1}\nabla h^\top \nabla^2_{xx}\ell \nabla h(\nabla h^\top \nabla h)^{-1}$. □

The local convergence properties of the multipliers method are established under assumptions A.1 and A.2, where A.2 is given below.

A.2 The point $x^* \in \Omega$ is a nondegenerate point (i.e., the gradients of the functions defining the active constraints are linearly independent) satisfying the second-order sufficient conditions for problem (1.1) with corresponding multipliers $\lambda(x^*,\bar{w}^*)$ and $\bar{w}^*$. The pair $(x^*,\bar{w}^*)$ satisfies strict complementarity.

The main result is presented in theorem 3.2 and bounds the distance between a local minimizer of (1.3) and $(x^*,\bar{w}^*)$ by the penalty parameter μ times the distance between the approximation w and the corresponding multipliers $\bar{w}^*$. The proof of this theorem is quite long and technical and is postponed to the appendix of this paper.

THEOREM 3.2. *Let x^*, with corresponding multipliers $\bar{w}^*$, satisfy assumptions A.1-A.2. There exist positive scalars $\bar{\mu}$, δ, ϵ, κ_1, and κ_2 such that*

$$Z^*Z^{*\top}\nabla^2_{xx}\ell^* Z^*Z^{*\top} - (I - Z^*Z^{*\top})\nabla^2_{xx}\ell^*(I - Z^*Z^{*\top}) + \frac{1}{\bar{\mu}}\nabla h^* \nabla h^{*\top}$$

is positive definite, the problem

$$\min\ P(x;w,\mu) \quad \text{s.t.} \quad x \geq 0,\ x \in B(x^*;\epsilon), \tag{3.3}$$

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

$$D^* = \left\{(w,\mu) : \|w - \bar{w}^*\| < \min\left\{\delta, \frac{\delta}{\mu}\right\}; 0 < \mu \leq \bar{\mu}\right\}, \tag{3.4}$$

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

$$\|\tilde{x}(w,\mu) - x^*\| \leq \kappa_1 \mu \|w - \bar{w}^*\|, \tag{3.5a}$$

$$\|\tilde{w}(w,\mu) - \bar{w}^*\| \leq \kappa_2 \mu \|w - \bar{w}^*\|, \tag{3.5b}$$

where $\tilde{w} = \tilde{w}(w, \mu)$ *are the multipliers corresponding to* $\tilde{x} = \tilde{x}(w, \mu)$*, and*

$$\begin{aligned} \tilde{w} \;&=\; Z(\tilde{x})Z(\tilde{x})^\top \nabla f(\tilde{x}) + (I - Z(\tilde{x})Z(\tilde{x})^\top)\, w \\ &\quad - \left(\nabla^2_{xx}\ell(\tilde{x}, w)\nabla h(\tilde{x}) + R(\tilde{x}, w)^\top\right) \left(\nabla h(\tilde{x})^\top \nabla h(\tilde{x})\right)^{-1} h(\tilde{x}) + \tfrac{1}{\mu}\nabla h(\tilde{x}) h(\tilde{x}). \end{aligned} \tag{3.6}$$

Theorem 3.2 can be used to state the basic local convergence properties of the multipliers method given in algorithm 2.1, which we summarize in the next corollary.

COROLLARY 3.3. *Let* x^**, with corresponding multipliers* $\bar{w}^*$*, satisfy assumptions A.1-A.2. 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* $\|w_0 - \bar{w}^*\| < \min\left\{\delta_0, \frac{\delta_0}{\mu_0}\right\}$*, then the sequence* $\{w_k\}$*, generated by* $w_{k+1} = \nabla P(\tilde{x}(w_k, \mu_k); w_k, \mu_k)$*, is well defined (in the sense that* $(w_k, \mu_k) \in D^*$ *for all k) and satisfies*

$$\limsup_{k \to +\infty} \frac{\|w_{k+1} - \bar{w}^*\|}{\|w_k - \bar{w}^*\|} \leq \kappa, \tag{3.7}$$

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

$$\lim_{k \to +\infty} \frac{\|w_{k+1} - \bar{w}^*\|}{\|w_k - \bar{w}^*\|} = 0, \tag{3.8}$$

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

$$\lim_{k \to +\infty} \tilde{x}(w_k, \mu_k) = x^*, \tag{3.9}$$

$$\lim_{k \to +\infty} w_k = \bar{w}^*. \tag{3.10}$$

Proof. The limits (3.7), (3.8) and (3.10) follow from inequality (3.5b). The limit (3.9) is a consequence of (3.5a). □

It is also worthwhile to note that the multipliers update (3.6) can be seen as an approximation to the steepest ascent iteration applied to the dual function associated with problem (3.3); see [6] for details on how this was carried out in the context of problem (1.6).

4. Conclusions and future research. The augmented Lagrangian multipliers method proposed in this paper is based on the solution of a sequence of bound-constrained minimization problems. Each outer iteration of the method involves the minimization, within the bounds, of the augmented Lagrangian penalty function $P(x; w, \mu)$ for specific values of the penalty parameter μ and of the multipliers w. The evaluation of $P(x; w, \mu)$ and of its gradient requires the solution of systems of linear equations with $\nabla h(x)^\top \nabla h(x)$. The gradient of $P(x; w, \mu)$ involves a cross term where second-order derivatives of the problem functions f and h appear. Thus, each inner or minor iteration, i.e., each iteration of the iterative process applied to minimize $P(x; w, \mu)$ within the bounds, is relatively costly.

This augmented Lagrangian multipliers method was proposed originally in [6] for a class of nonlinear programming problems with a structure arising from optimal control or design, see (1.6). There, the role of the matrix $\nabla h(x)^\top \nabla h(x)$ is played by the matrix $c_y(y,u)$, the partial Jacobian of $c(y,u)$ with respect to the state variables y. The computation of the gradient of the penalty function involves there the solution of linear systems with $c_y(y,u)$ (linearized state equations) and with $c_y(y,u)^\top$ (adjoint equations), for which solvers are available in many applications, see [4].

One major open question is weather a globalization scheme, similar to what was developed in [2] for the original multipliers method, would be applicable to the new multipliers method of this paper, yielding the same type of global convergence. In contrast to what happens in [2], we do not have here the equality $\nabla P(x;w,\mu) = \nabla_x \ell(x,\lambda(x,\bar{w}))$ that seems to us to be crucial to the derivation of global convergence. What we get instead is the following:

$$\begin{aligned} &\nabla P(x;w,\mu) - \nabla_x \ell(x,\lambda(x,\bar{w})) = \\ &-Z(x)Z(x)^\top (\nabla^2_{xx}\ell(x,\lambda(x,w))\nabla h(x) + R(x,w)^\top)\,(\nabla h(x)^\top \nabla h(x))^{-1}\, h(x). \end{aligned} \tag{4.1}$$

When $h(x) = 0$ we do have, of course, $\nabla P(x;w,\mu) = \nabla_x \ell(x,\lambda(x,\bar{w}))$. The fact that there is a term depending on the size of the feasibility function $h(x)$ in (4.1) makes the global analysis considerably more difficult.

Numerical results obtained for small-scale dimension problems have shown that the method is competitive with Lancelot [3], sharing some of the advantages and disadvantages of the class of augmented Lagrangian multipliers methods.

Appendix. We prove here the main result of local convergence established in theorem 3.2. We will use the following notation. The symbol e represents a vector of ones with appropriate size and e_i denotes a vector whose i-th component is unity and the others zero. Also, for any vector v, V is the diagonal matrix for which the diagonal elements are the elements of v.

Although the structure of the proof follows the one in [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 $\frac{1}{\mu_k}$ but involves w_k. A consequence of having to handle this extra term is that the region D^* in (3.4) becomes smaller than the one in [1, proposition 2.4], where instead of $\min\{\delta, \delta/\mu\}$ we only have δ/μ.

We need first to organize some of the calculations that will appear later. The derivative of $s(x) = (\nabla h(x)^\top \nabla h(x))^{-1} h(x)$ is given by

$$\begin{aligned} \nabla s(x)^\top &= (\nabla h^\top \nabla h)^{-1} \nabla h^\top - (\nabla h^\top \nabla h)^{-1} \sum_{i=1}^{m} \nabla (\nabla h^\top \nabla h)_i \left[(\nabla h^\top \nabla h)^{-1} h \right]_i \\ &\stackrel{\text{def}}{=} (\nabla h^\top \nabla h)^{-1} \nabla h^\top - F(h), \end{aligned}$$

where we have omitted the argument x in the right hand side. The size of $F(h(x))$ varies continuously with $h(x)$.

Further, we note that from $\nabla h(x)^\top Z(x) = 0$ one obtains

$$\nabla h(x)^\top \nabla Z(x)_j^\top = - \begin{pmatrix} Z(x)_j^\top \nabla^2 h(x)_1 \\ \vdots \\ Z(x)_j^\top \nabla^2 h(x)_m \end{pmatrix}, \tag{4.2}$$

for $j = 1, \ldots, n-m$, where $Z(x)_j$ denotes the j-th column of $Z(x)$. By using (4.2), we can write

$$\begin{pmatrix} (\nabla f - w)^\top (I - ZZ^\top)\nabla Z_1^\top \\ \vdots \\ (\nabla f - w)^\top (I - ZZ^\top)\nabla Z_{n-m}^\top \end{pmatrix} = -Z^\top \sum_{i=1}^{m} \left[(\nabla h^\top \nabla h)^{-1} \nabla h^\top (\nabla f - w)\right]_i \nabla^2 h_i.$$

We have assumed that $Z(x)$ is differentiable. Goodman [5] has shown how to extend locally an orthonormal basis $Z(x)$ given by the QR factorization of $\nabla h(x)$ so that $Z(x)$ exhibits the same smoothness of $h(x)$.

We finally get an expression that will be used later on:

$$(4.3)\quad \begin{pmatrix} (\nabla f - w)^\top \nabla Z_1^\top \\ \vdots \\ (\nabla f - w)^\top \nabla Z_{n-m}^\top \end{pmatrix} = \begin{pmatrix} (\nabla f - w)^\top ZZ^\top \nabla Z_1^\top \\ \vdots \\ (\nabla f - w)^\top ZZ^\top \nabla Z_{n-m}^\top \end{pmatrix} - Z^\top \nabla^2 f + Z^\top \nabla^2_{xx} \ell.$$

We are ready now to prove theorem 3.2. The proof is divided in six major steps.

Proof. **A. Preparing the system of nonlinear equations.** Consider, for $\mu > 0$, the system of nonlinear equations that results from the first-order necessary conditions (2.3a)-(2.3c) for problem (1.3). If we multiply equation (2.3a) by $\nabla h^\top$ and $Z^\top$, we obtain the equivalent system

$$(4.4a)\quad \begin{aligned} &\nabla h^\top w - \nabla h^\top (\nabla^2_{xx}\ell(x,w)\nabla h + R(x,w)^\top)\left(\nabla h^\top \nabla h\right)^{-1} h + \tfrac{1}{\mu}\nabla h^\top \nabla h\; h \\ &-\nabla h^\top \bar{w} = 0, \end{aligned}$$

$$(4.4b)\quad Z^\top \nabla f - Z^\top (\nabla^2_{xx}\ell(x,w)\nabla h + R(x,w)^\top)\left(\nabla h^\top \nabla h\right)^{-1} h - Z^\top \bar{w} = 0,$$

$$(4.4c)\quad X\bar{W}e = 0.$$

Now we multiply equation (4.4a) by μ and perform the changes of variables

$$(4.5a)\quad r = \mu(w - \bar{w}^*),$$

$$(4.5b)\quad s = w - \bar{w}^*,$$

to obtain the system of nonlinear equations

$$(4.6a)\quad \begin{aligned} &\nabla h^\top r - \mu \nabla h^\top (\nabla^2_{xx}\ell(x,\bar{w}^* + s)\nabla h + R(x,\bar{w}^* + s)^\top)\left(\nabla h^\top \nabla h\right)^{-1} h \\ &+\nabla h^\top \nabla h h + \mu \nabla h^\top \bar{w}^* - \mu \nabla h^\top \bar{w} = 0, \end{aligned}$$

$$(4.6b)\quad \begin{aligned} &Z^\top \nabla f - Z^\top (\nabla^2_{xx}\ell(x,\bar{w}^* + s)\nabla h + R(x,\bar{w}^* + s)^\top)\left(\nabla h^\top \nabla h\right)^{-1} h \\ &-Z^\top \bar{w} = 0, \end{aligned}$$

$$(4.6c)\quad X\bar{W}e = 0,$$

that we write as

$$J(x, w, \mu) = 0.$$

We analyze this system for $\mu \in [0, \mu^*]$, where μ^* is such that

$$(4.7)\quad \mu Z^* Z^{*\top} \nabla^2_{xx}\ell^* Z^* Z^{*\top} - \mu(I - Z^* Z^{*\top})\nabla^2_{xx}\ell^* (I - Z^* Z^{*\top}) + \nabla h^* \nabla h^{*\top}$$

is positive definite for all $\mu \in (0, \mu^*]$. The existence of such $\mu^* > 0$ is guaranteed by theorem 3.1.

B. Nonsingularity at the solution when the penalty parameter is zero. When $r = s = 0$ and $\mu \in [0, \mu^*]$, it is easy to check that the system (4.6a)-(4.6c) has the solution $(x^*, \bar{w}^*)$. For $r = s = 0$, the Jacobian of (4.6a)-(4.6c) with respect to $(x, \bar{w})$, at the point $(x^*, \bar{w}^*)$, is given by

$$J^*(0,0,\mu) = \begin{pmatrix} -\mu \nabla h^{*\top} \nabla^2_{xx} \ell^* \left(I - Z^* Z^{*\top}\right) + \nabla h^{*\top} \nabla h^* \nabla h^{*\top} & -\mu \nabla h^{*\top} \\ Z^{*\top} \nabla^2_{xx} \ell^* Z^* Z^{*\top} & -Z^{*\top} \\ \bar{W}^* & X^* \end{pmatrix}.$$

When $\mu = 0$, $J^*(0,0,\mu)$ reduces to

$$J^*(0,0,0) = \begin{pmatrix} \nabla h^{*\top} \nabla h^* \nabla h^{*\top} & 0 \\ Z^{*\top} \nabla^2_{xx} \ell^* Z^* Z^{*\top} & -Z^{*\top} \\ \bar{W}^* & X^* \end{pmatrix}. \tag{4.8}$$

One can see that $J^*(0,0,0)$ is nonsingular. In fact, the assumptions on $(x^*, \bar{w}^*)$ imply that the following matrix is nonsingular:

$$\begin{pmatrix} \nabla h^{*\top} & 0 & 0 \\ Z^* Z^{*\top} \nabla^2_{xx} \ell^* Z^* Z^{*\top} & \nabla h^* & -I \\ \bar{W}^* & 0 & X^* \end{pmatrix}. \tag{4.9}$$

The nonsingularity of (4.9) implies the nonsingularity of (4.8).

C. Nonsingularity at the solution for positive values of the penalty parameter. Let $(\Delta x, \Delta w)$ be a solution of the homogeneous linear system with the matrix $J^*(0,0,\mu)$:

$$\left(-\mu \nabla h^{*\top} \nabla^2_{xx} \ell^* \left(I - Z^* Z^{*\top}\right) + \nabla h^{*\top} \nabla h^* \nabla h^{*\top}\right) \Delta x - \mu \nabla h^{*\top} \Delta w = 0, \tag{4.10a}$$

$$Z^{*\top} \nabla^2_{xx} \ell^* Z^* Z^{*\top} \Delta x - Z^{*\top} \Delta w = 0, \tag{4.10b}$$

$$\bar{W}^* \Delta x + X^* \Delta w = 0. \tag{4.10c}$$

The equation (4.10c) and strict complementarity between x^* and $\bar{w}^*$ imply $\Delta x^\top \Delta w = 0$. By multiplying (4.10a) and (4.10b) on the left by $\nabla h^* \left(\nabla h^{*\top} \nabla h^*\right)^{-1}$ and μZ^*, respectively, we obtain

$$\left(-\mu \left(I - Z^* Z^{*\top}\right) \nabla^2_{xx} \ell^* \left(I - Z^* Z^{*\top}\right) + \nabla h^* \nabla h^{*\top}\right) \Delta x - \mu \left(I - Z^* Z^{*\top}\right) \Delta w = 0,$$

$$\mu Z^* Z^{*\top} \nabla^2_{xx} \ell^* Z^* Z^{*\top} \Delta x - \mu Z^* Z^{*\top} \Delta w = 0.$$

Thus,

$$\mu Z^* Z^{*\top} \nabla^2_{xx} \ell^* Z^* Z^{*\top} \Delta x - \mu \left(I - Z^* Z^{*\top}\right) \nabla^2_{xx} \ell^* \left(I - Z^* Z^{*\top}\right) \Delta x + \nabla h^* \nabla h^{*\top} \Delta x$$

$$-\mu \Delta w = 0.$$

By multiplying this equation on the left by $\Delta x^\top$, we derive

$$\Delta x^\top \left(\mu Z^* Z^{*\top} \nabla^2_{xx} \ell^* Z^* Z^{*\top} - \mu \left(I - Z^* Z^{*\top}\right) \nabla^2_{xx} \ell^* \left(I - Z^* Z^{*\top}\right) + \nabla h^* \nabla h^{*\top}\right) \Delta x = 0.$$

Since (4.7) is positive definite for $\mu \in (0, \mu^*]$, we conclude that $\Delta x = 0$. Now, using $\Delta x = 0$, we get $\nabla h^{*\top} \Delta w = 0$ and $Z^{*\top} \Delta w = 0$, implying that $\Delta w = 0$. We have therefore proved that $J^*(0, 0, \mu)$ is nonsingular for $\mu \in (0, \mu^*]$.

D. The use of the implicit function theorem. We now apply the implicit function theorem [1, page 12] to the system (4.6a)-(4.6c). By identifying the set $K = \{0\} \times \{0\} \times [0, \mu^*]$ as the compact set $\bar{X}$ of that theorem, we guarantee the existing of positive scalars ϵ and δ and unique continuously differentiable functions $\hat{x} = \hat{x}(r, s, \mu)$ and $\hat{w} = \hat{w}(r, s, \mu)$, defined on a neighborhood of K,

$$B(K, \delta) = \{(r, s, \mu) : \|(r, s, \mu) - (0, 0, \mu')\| < \delta \text{ for some } (0, 0, \mu') \in K\},$$

satisfying (4.6a)-(4.6c) with $x = \hat{x} = \hat{x}(r, s, \mu)$ and $\bar{w} = \hat{w} = \hat{w}(r, s, \mu)$, and such that

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

for all $(r, s, \mu) \in B(K, \delta)$. Using (4.6c) and strict complementarity of the pair $(x^*, \bar{w}^*)$, and reducing ϵ and δ if necessary, one can easily show for all $(r, s, \mu) \in B(K, \delta)$ that: $\hat{x}(r, s, \mu) \geq 0$; $\hat{w}(r, s, \mu) \geq 0$; the pair $(\hat{x}(r, s, \mu), \hat{w}(r, s, \mu))$ also verifies strict complementarity; the gradients of the active constraints are linearly independent at $\hat{x}(r, s, \mu)$.

E. The bounds (3.5a)-(3.5b). We differentiate (4.6a)-(4.6c) with respect to (r, s, μ), and write

$$(4.11)\; J(r, s, \mu) \begin{pmatrix} \nabla_r \hat{x}(r, s, \mu)^\top & \nabla_s \hat{x}(r, s, \mu)^\top & \nabla_\mu \hat{x}(r, s, \mu)^\top \\ \nabla_r \hat{w}(r, s, \mu)^\top & \nabla_s \hat{w}(r, s, \mu)^\top & \nabla_\mu \hat{w}(r, s, \mu)^\top \end{pmatrix} = -B(r, s, \mu).$$

Here $J(r, s, \mu)$ is the Jacobian of the vector function of the left-hand side of (4.6) with respect to x and $\bar{w}$, given by

$$\begin{pmatrix} -\mu \nabla h^\top \nabla^2_{xx} \ell \left(I - Z Z^\top\right) + \nabla h^\top \nabla h \nabla h^\top & -\mu \nabla h^\top \\ Z^\top \nabla^2_{xx} \ell Z Z^\top & -Z^\top \\ \hat{W} & \hat{X} \end{pmatrix} +$$

$$\begin{pmatrix} \begin{matrix} A_{11} - \mu \nabla h^\top R^\top \left(\nabla h^\top \nabla h\right)^{-1} \nabla h^\top + \mu \nabla h^\top \left(\nabla^2_{xx} \ell \nabla h + R^\top\right) F(h) \\ -\mu \sum_{i=1}^{m} \nabla_x (\nabla h^\top (\nabla^2_{xx} \ell \nabla h + R^\top))_i [(\nabla h^\top \nabla h)^{-1} h]_i + \sum_{i=1}^{m} h_i \nabla (\nabla h^\top \nabla h)_i \end{matrix} & 0 \\ \begin{matrix} A_{21} - \sum_{i=1}^{m} \left[(\nabla h^\top \nabla h)^{-1} h\right]_i \nabla_x (Z^\top \left(\nabla^2_{xx} \ell \nabla h + R^\top\right))_i \\ + Z^\top \left(\nabla^2_{xx} \ell \nabla h + R^\top\right) F(h) - Z^\top R^\top \left(\nabla h^\top \nabla h\right)^{-1} \nabla h^\top \end{matrix} & 0 \\ 0 & 0 \end{pmatrix},$$

where the functions h, Z, ∇h, $\nabla^2 h_i$, $i = 1, ..., m$, are evaluated at $\hat{x}(r,s,\mu)$ and the functions $\nabla^2_{xx}\ell$ and R are evaluated at $(\hat{x}(r,s,\mu), \hat{w}(r,s,\mu))$, and where the rows of A_{11} are given by $(A_{11})_i = (r + \mu(\bar{w}^* - \bar{w}))^\top \nabla^2 h_i$, $i = 1, ..., m$. The term A_{21} is given by

$$A_{21} = \begin{pmatrix} (\nabla f - \bar{w})^\top \nabla Z_1^\top \\ \vdots \\ (\nabla f - \bar{w})^\top \nabla Z_{n-m}^\top \end{pmatrix} + Z^\top \nabla^2 f - Z^\top \nabla^2_{xx}\ell$$

$$= \begin{pmatrix} (\nabla f - \bar{w})^\top Z Z^\top \nabla Z_1^\top \\ \vdots \\ (\nabla f - \bar{w})^\top Z Z^\top \nabla Z_{n-m}^\top \end{pmatrix},$$

where the last equality is justified by the derivation (4.3).

In (4.11), B is the Jacobian of the vector function of the left-hand side of (4.6) with respect to r, s and μ, defined by

$$B(r,s,\mu) = \begin{pmatrix} B_{11}(r,s,\mu) & B_{12}(r,s,\mu) & B_{13}(r,s,\mu) \\ 0 & B_{22}(r,s,\mu) & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

with

$$B_{11}(r,s,\mu) = \nabla h^\top,$$

$$B_{12}(r,s,\mu)\ e_j = -\mu \nabla h^\top \sum_{i=1}^{m} \left[(\nabla h^\top \nabla h)^{-1} \nabla h^\top e_j \right]_i \nabla^2 h_i\ \nabla h (\nabla h^\top \nabla h)^{-1} h$$

$$+\mu \nabla h^\top \begin{pmatrix} (Z Z^\top e_j)^\top \nabla^2 h_1 \\ \vdots \\ (Z Z^\top e_j)^\top \nabla^2 h_m \end{pmatrix} (\nabla h^\top \nabla h)^{-1} h,$$

$$B_{13}(r,s,\mu) = -\nabla h^\top \left(\left(\nabla^2_{xx}\ell\ \nabla h + R^\top \right) \left(\nabla h^\top \nabla h \right)^{-1} h + \bar{w}^* - \hat{w}(r,s,\mu) \right),$$

$$B_{22}(r,s,\mu)\ e_j = -Z^\top \sum_{i=1}^{m} \left[(\nabla h^\top \nabla h)^{-1} \nabla h^\top e_j \right]_i \nabla^2 h_i\ \nabla h (\nabla h^\top \nabla h)^{-1} h$$

$$+Z^\top \begin{pmatrix} (Z Z^\top e_j)^\top \nabla^2 h_1 \\ \vdots \\ (Z Z^\top e_j)^\top \nabla^2 h_m \end{pmatrix} (\nabla h^\top \nabla h)^{-1} h,$$

where $j = 1, ..., n$.

Hence, for all $(r,s,\mu) \in B(K,\delta)$, we have

$$\begin{pmatrix} \hat{x}(r,s,\mu) - x^* \\ \hat{w}(r,s,\mu) - \bar{w}^* \end{pmatrix} = \begin{pmatrix} \hat{x}(r,s,\mu) - \hat{x}(0,0,0) \\ \hat{w}(r,s,\mu) - \hat{w}(0,0,0) \end{pmatrix}$$

$$= -\int_0^1 J(\tau r, \tau s, \tau \mu)^{-1} B(\tau r, \tau s, \tau \mu) \begin{pmatrix} r \\ s \\ \mu \end{pmatrix} d\tau.$$

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 $J^*(0,0,0)$. If we look carefully at the second term of $J(r,s,\mu)$, we come to the conclusion that all the expressions involved depend continuously on either r, μ, h, R or $Z^\top(\nabla f - \bar{w})$, quantities that are of size δ and ϵ.

Now we can finally show (3.5a)-(3.5b). By appealing to

$$\left\| \begin{pmatrix} \hat{x}(r,s,\mu) - x^* \\ \hat{w}(r,s,\mu) - \bar{w}^* \end{pmatrix} \right\| \le \max_{\tau\in[0,1]} \|J(\tau r,\tau s,\tau\mu)^{-1}\| \int_0^1 \left\| B(\tau r,\tau s,\tau\mu) \begin{pmatrix} r \\ s \\ \mu \end{pmatrix} \right\| d\tau,$$

and by applying the continuity assumptions to the terms that appear in $B_{11}(r,s,\mu)$, $B_{12}(r,s,\mu)$, $B_{13}(r,s,\mu)$, $B_{22}(r,s,\mu)$, we can assume the existence of positive constants κ_3-κ_7 such that

$$\begin{aligned} \|\hat{x}(r,s,\mu) - x^*\| + \|\hat{w}(r,s,\mu) - \bar{w}^*\| \le \kappa_3\|r\| + \kappa_4\mu\|s\|d(r,s,\mu) + \kappa_5\mu d(r,s,\mu) + \\ \kappa_6\mu \max_{\tau\in[0,1]} \|\hat{w}(\tau r,\tau s,\tau\mu) - \bar{w}^*\| \\ +\kappa_7\|s\|d(r,s,\mu), \end{aligned}$$

where

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

Furthermore, from (4.6a) we write, with $\hat{x} = \hat{x}(r,s,\mu)$ and $\hat{w} = \hat{w}(r,s,\mu)$,

$$\begin{aligned} h(\hat{x}) \;=\; & \nabla h(\hat{x})^\top \nabla h(\hat{x}) \left\{-\mu \nabla h(\hat{x})^\top \left(\nabla^2_{xx}\ell(\hat{x},\hat{w}) \nabla h(\hat{x}) + R(\hat{x},\hat{w})^\top\right)\right. \\ & \left. +\nabla h(\hat{x})^\top \nabla h(\hat{x}) \nabla h(\hat{x})^\top \nabla h(\hat{x})\right\}^{-1} \nabla h(\hat{x})^\top (-r + \mu\hat{w} - \mu\bar{w}^*). \end{aligned} \tag{4.12}$$

Thus, the choice of μ^* and the continuity assumptions, together with the expression (4.12) for $h(\hat{x})$, imply that

$$d(r,s,\mu) \le \kappa_8\|r\| + \kappa_9\mu \max_{\tau\in[0,1]} \|\hat{w}(\tau r,\tau s,\tau\mu) - \bar{w}^*\|,$$

for some positive constants κ_8 and κ_9. Since $\mu \le \mu^*$ and $\|s\| < \delta$, there exist positive constants κ_{10} and κ_{11} such that

$$\|\hat{x}(r,s,\mu) - x^*\| + \|\hat{w}(r,s,\mu) - \bar{w}^*\| \le \kappa_{10}\|r\| + \kappa_{11}\mu \max_{\tau\in[0,1]} \|\hat{w}(\tau r,\tau s,\tau\mu) - \bar{w}^*\|,$$

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

$$\max_{\tau\in[0,1]} \|\hat{w}(\tau r,\tau s,\tau\mu) - \bar{w}^*\| \le \frac{\kappa_{10}}{1-\kappa_{11}\mu}\|r\|,$$

for $\mu \in [0,\bar{\mu}]$, with $\bar{\mu} < \min\left\{\mu^*, \frac{1}{\kappa_{11}}\right\}$. Therefore,

$$\begin{aligned} \|\hat{x}(r,s,\mu) - x^*\| + \|\hat{w}(r,s,\mu) - \bar{w}^*\| \;&\le\; \left(\kappa_{10} + \tfrac{\kappa_{10}\kappa_{11}\mu}{1-\kappa_{11}\mu}\right)\|r\| \\ &\le\; \tfrac{\kappa_{10}}{1-\kappa_{11}\bar{\mu}}\mu\|w - \bar{w}^*\|. \end{aligned} \tag{4.13}$$

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

$$\tilde{x}(w,\mu) = \hat{x}(r,s,\mu) = \hat{x}\left(\mu\left(w - \bar{w}^*\right), w - \bar{w}^*, \mu\right), \tag{4.14}$$
$$\tilde{w}(w,\mu) = \hat{w}(r,s,\mu) = \hat{w}\left(\mu\left(w - \bar{w}^*\right), w - \bar{w}^*, \mu\right).$$

Hence, the bounds (3.5a)-(3.5b) follow immediately from (4.13).

F. Optimality of $\tilde{x}(w, \mu)$. We finish the proof by showing that $\tilde{x}(w,\mu)$ is the solution of problem (3.3). First we point out that $(\tilde{x}(w,\mu), \tilde{w}(w,\mu))$ satisfies the first-order necessary conditions for (3.3) as it can be seen by rewriting system (4.6a)-(4.6c) using the changes of variables (4.5a)-(4.5b) and (4.14). The first equation of the first-order necessary conditions is, with $\tilde{x} = \tilde{x}(w,\mu)$ and $\tilde{w} = \tilde{w}(w,\mu)$,

$$Z(\tilde{x})Z(\tilde{x})^\top \nabla f(\tilde{x}) + (I - Z(\tilde{x})Z(\tilde{x})^\top)\, w - \left(\nabla^2_{xx}\ell(\tilde{x},w)\nabla h(\tilde{x}) + R(\tilde{x},w)^\top\right) \tag{4.15}$$
$$\left(\nabla h(\tilde{x})^\top \nabla h(\tilde{x})\right)^{-1} h(\tilde{x}) + \tfrac{1}{\mu}\nabla h(\tilde{x}) h(\tilde{x}) - \tilde{w} = 0,$$

and (3.6) is clearly true. We show now that the Hessian of $P(x; w, \mu)$ is positive definite at $\tilde{x}(w,\mu)$ for all vectors

$$(\Delta x)_i = 0 \;\text{ if }\; \left(\tilde{x}(w,\mu)\right)_i = 0 \;\text{ and }\; \left(\tilde{w}(w,\mu)\right)_i > 0. \tag{4.16}$$

The case $(\Delta x)_i \geq 0$ is eliminated, because the pair $(\tilde{x}(w,\mu), \tilde{w}(w,\mu))$ is strictly complementary. The scalar ϵ can be chosen sufficiently small so that we can consider

$$(\Delta x)_i = 0 \;\text{ if }\; x^*_i = 0 \;\text{ and }\; \left(\bar{w}^*\right)_i > 0.$$

This means that we can check the positive definiteness of the Hessian of $P(x; w, \mu)$ in the same subspace that we consider for $P(x^*; \bar{w}^*, \mu)$. Moreover, we proved in theorem 3.1 that the Hessian of $P(x^*; \bar{w}^*, \mu)$ is positive definite for $\mu \in (0, \mu^*]$ in the above mentioned subspace. To achieve our goal, we show that the Hessian of $P\left(\tilde{x}(w,\mu); w, \mu\right)$ is a perturbation of size ϵ and δ of the Hessian of $P(x^*; \bar{w}^*, \mu)$. In fact, the Hessian of $P\left(\tilde{x}(w,\mu); w, \mu\right)$ is given by

$$\nabla^2_{xx}\ell Z Z^\top - R^T \left(\nabla h^\top \nabla h\right)^{-1} \nabla h^T - \left(I - Z Z^\top\right) \nabla^2_{xx}\ell - \nabla h \left(\nabla h^\top \nabla h\right)^{-1} R,$$
$$+ \sum_{i=1}^{m} h_i \nabla^2_{xx}\lambda_i + \frac{1}{\mu}\nabla h \nabla h^\top + \frac{1}{\mu}\sum_{i=1}^{m} h_i \nabla^2 h_i,$$

see (3.1), with the Lagrangian and the residual R evaluated at $(\tilde{x}(w,\mu), w)$ and the remaining functions at $\tilde{x}(w,\mu)$. The term

$$\nabla^2_{xx}\ell Z Z^\top - R^T \left(\nabla h^\top \nabla h\right)^{-1} \nabla h^T - \left(I - Z Z^\top\right) \nabla^2_{xx}\ell - \nabla h \left(\nabla h^\top \nabla h\right)^{-1} R + \frac{1}{\mu}\nabla h \nabla h^\top$$

is a perturbation of size ϵ and δ of the Hessian of $P(x^*; \bar{w}^*, \mu)$. To bound the remaining terms, we can rewrite (4.15), using $\tilde{x} = \tilde{x}(w,\mu)$ and $\tilde{w} = \tilde{w}(w,\mu)$, as

$$\frac{1}{\mu} h(\tilde{x}) = \nabla h(\tilde{x})^\top \nabla h(\tilde{x}) \{-\mu \nabla h(\tilde{x})^\top \left(\nabla^2_{xx}\ell(\tilde{x},w)\nabla h(\tilde{x}) + R(\tilde{x},w)^\top\right)$$
$$+ (I - Z(\tilde{x})Z(\tilde{x})^\top)\}^{-1} \nabla h(\tilde{x})^\top \left(\tilde{w} - w\right).$$

Thus, using the continuity assumptions and adding and subtracting $\bar{w}^*$, we obtain, for some positive constant κ_{12},

$$\begin{aligned}\left|\left|\frac{1}{\mu}h(\tilde{x}(w,\mu))\right|\right| &\leq \kappa_{12}\left(||\tilde{w}(w,\mu)-\bar{w}^*||+||w-\bar{w}^*||\right)\\ &\leq \kappa_{12}(\epsilon+\delta)\end{aligned}$$

and

$$||h(\tilde{x}(w,\mu))|| \leq \bar{\mu}\kappa_{12}(\epsilon+\delta).$$

The conclusion is that $\sum_{i=1}^m h_i \nabla^2_{xx}\lambda_i + \frac{1}{\mu}\sum_{i=1}^m h_i\nabla^2 h_i$ is also of size δ and ϵ, and the proof that the Hessian of $P(x;w,\mu)$ is positive definite for all vectors Δx satisfying (4.16) is terminated. □

The proof also shows that κ_1 and κ_2 in the bounds (3.5) grow with the condition number of $\nabla h^\top \nabla h$.

REFERENCES

[1] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Computer Science and Applied Mathematics, Academic Press, New York, 1982.

[2] 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.

[3] ———, *LANCELOT: A Fortran package for large-scale nonlinear optimization (Release A)*, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1992.

[4] 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.

[5] J. Goodman, *Newton's method for constrained optimization*, Math. Programming, 33 (1985), pp. 162–171.

[6] L. N. Vicente, *Local analysis of a new multipliers method*, European Journal of Operational Research, 143 (2002), pp. 432–451.