# ON INTERIOR–POINT NEWTON ALGORITHMS FOR DISCRETIZED OPTIMAL CONTROL PROBLEMS WITH STATE CONSTRAINTS

LUíS N. VICENTE *

**Abstract.** In this paper we consider a class of nonlinear programming problems that arise from the discretization of optimal control problems with bounds on both the state and the control variables. For this class of problems, we analyze constraint qualifications and optimality conditions in detail. We derive an affine–scaling and two primal–dual interior–point Newton algorithms by applying, in an interior–point way, Newton's method to equivalent forms of the first–order optimality conditions. Under appropriate assumptions, the interior–point Newton algorithms are shown to be locally well–defined with a q–quadratic rate of local convergence. By using the structure of the problem, the linear algebra of these algorithms can be reduced to the null space of the Jacobian of the equality constraints. The similarities between the three algorithms are pointed out, and their corresponding versions for the general nonlinear programming problem are discussed.

**Keywords.** Nonlinear programming, Newton's method, interior–point algorithms, affine scaling, primal dual, optimal control problems, state constraints.

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

**1. Introduction.** We are interested in the solution of the following nonlinear programming problem:

$$\begin{aligned} &\text{minimize} && f(y,u) \\ &\text{subject to} && C(y,u) = 0\,, \\ & && y \geq 0\,,\ u \geq 0\,, \end{aligned} \tag{1.1}$$

where $y \in \mathbb{R}^m$, $u \in \mathbb{R}^{n-m}$, and $n$ and $m$ are positive integers satisfying $n > m$. The functions $f$ and $C$ are considered smooth and defined as $f : \Omega \longrightarrow \mathbb{R}$ and $C : \Omega \longrightarrow \mathbb{R}^m$, where $\Omega$ is an open set of $\mathbb{R}^n$ containing $\{y : y \geq 0\} \times \{u : u \geq 0\}$. The results in this paper are easily extended for bounds of the form $a_y \leq y \leq b_y$ and $a_u \leq u \leq b_u$.

The nonlinear programming problem (1.1) often arises from the discretization of optimal control problems (see references [2], [5], [10], [15], [16], [17], [18]). In this case $y$ is the vector of state variables, $u$ is the vector of control variables, and $C(y,u) = 0$ is the discretized state equation. Problem (1.1) also appears frequently in engineering design, inversion, and parameter identification.

The purpose of this paper is to present an affine–scaling and two primal–dual interior–point Newton algorithms to solve problems of the form (1.1). The affine–scaling framework extends the approach followed by [10], [17], [28] for problems of the type (1.1) but with no bounds on the state variables $y$. First, it is shown that the Karush–Kuhn–Tucker constraint qualifications hold for most feasible points of (1.1). We also derive optimality conditions for (1.1) using affine–scaling matrices. Departing from different but equivalent forms of the first–order optimality conditions, we apply Newton's method in an interior–point way and derive an affine–scaling and two primal–dual interior–point Newton algorithms. Table 1.1 summarizes the correspondence between the algorithms and the respective forms of the first–order optimality conditions. By using the structure

* Departamento de Matemática, Universidade de Coimbra, 3000 Coimbra, Portugal, E-Mail: lvicente@mat.uc.pt.
Support for this work has been provided by Instituto de Telecomunicações, JNICT, and Praxis XXI 2/2.1/MAT/346/94.

of the problem, we show how to reduce the computation of a step of the interior–point Newton algorithms to the null space of the Jacobian of $C(y,u)=0$. It is shown that in this reduced version these algorithms exhibit coefficient matrices with the same structure.

A few instances of the affine–scaling interior–point Newton algorithms have already been implemented and tested with problems from [17]. The preliminary numerical results are encouraging. The preconditioners proposed in [1] will certainly play an important role in the implementation of efficient and robust algorithms for this class of problems.

TABLE 1.1

*Interior–point Newton algorithms corresponding to the application of Newton's method to equivalent forms of the first–order optimality conditions defined on the set, or subset, of primal and dual variables. $x$ are primal variables, $\lambda$ are dual variables associated with the equality constraints, and $z$ are dual variables associated with the inequality (or bound) constraints.*

| Interior–Point Newton Algorithm | Variables | | |
|---|---|---|---|
| | Primal $x$ | Dual $\lambda$ | Dual $z$ |
| Affine Scaling | $\surd$ | $\surd$ | |
| Primal Dual | $\surd$ | $\surd$ | $\surd$ |
| Reduced Primal Dual | $\surd$ | | $\surd$ |

Coleman and Li introduced affine–scaling interior–point methods for minimization problems with simple bounds (see [3], [6], [7], [11]). Interior–point methods for nonlinear programming have been proposed and analyzed in [12], [30], [31] (primal dual), [4], [24] (primal using trust regions and approximation to the multipliers corresponding to inequality constraints), [8], [19], [20] (affine scaling). For discretized optimal control problems see also [18], [29].

Some of the ideas presented in this paper were discovered independently by Das [9] for the general nonlinear programming problem, namely the formulation of the optimality conditions and the application of Newton's method using affine–scaling matrices. The full–space version of the primal–dual interior–point Newton algorithm (second in Table 1.1) is the algorithm introduced in [12] for the general nonlinear programming problem.

The paper is structured as follows. In Section 2 we introduce notation, describe the optimality conditions, and analyze what conditions assure that the constraint qualifications hold. In Section 3 we derive a form for the optimality conditions using affine–scaling matrices. Sections 4 and 5 introduce respectively the affine–scaling and the primal–dual interior–point Newton algorithms and establish their well–posedness. Finally, we consider the general nonlinear programming problem in Section 6 and summarize our conclusions in Section 7.

**2. Constraint qualifications and optimality conditions.** In our notation, we have

$$x=\begin{pmatrix} y \\ u \end{pmatrix}.$$

Also, $(z)_y$ and $(z)_u$ represent the subvectors of $z\in\mathbb{R}^n$ corresponding to the $y$ and $u$ components, respectively, and $I_p$ represents the identity matrix of order $p$.

We denote the Lagrangian of $f(x)$ with respect to the equality constraints $C(x)=0$ by $\ell(x,\lambda)=f(x)+\lambda^\top C(x)$. The Jacobian matrix of $C(x)$ is denoted by $J(x)$. Due to the partition of $x$ in $y$

and $u$, we have

$$J(x) \;=\; \begin{pmatrix} C_y(x) & C_u(x) \end{pmatrix} ,$$

where the partial Jacobian $C_y(x)$ is a square matrix of order $m$.

In this paper we assume that $f$ and $C$ satisfy the following assumptions, where $\Omega$ is an open set of $\mathbb{R}^n$ containing $\{x : x \geq 0\}$.

ASSUMPTIONS 2.1.

1 The functions $f$ and $C$ are twice continuously differentiable with Lipschitz second derivatives in $\Omega$.

2 The partial Jacobian $C_y(x)$ is nonsingular in $\Omega$.

Although the analysis that motivates our framework requires $f$ and $C$ to be twice continuously differentiable, the resulting algorithms might use only first derivatives.

Let $\mathcal{U}$ be an open set containing $\{u : u \geq 0\}$ such that for all $u \in \mathcal{U}$ there exists a solution $y$ of $C(y,u) = 0$ and such that the matrix $C_y(x)$ is invertible for all $x = (y^\top \; u^\top)^\top$ with $u \in \mathcal{U}$ and $C(y,u) = 0$. Then the implicit function theorem guarantees the existence of a twice continuously differentiable function

$$y : \mathcal{U} \to \mathbb{R}^m \tag{2.1}$$

defined by

$$C(y(u), u) = 0 \, .$$

This allows us to reduce the minimization problem (1.1) to the space of the control variables $u$. The reduced problem is given by

$$\begin{aligned} &\text{minimize} \quad f(y(u), u) \\ &\text{subject to} \quad y(u) \geq 0 \, , u \geq 0 \, . \end{aligned} \tag{2.2}$$

It is important to note that the bound constraints in the state variables $y$ that are imposed in formulation (1.1), are transformed, in formulation (2.2), into nonlinear inequality constraints in the control variables $u$. The vector–valued function $y(u)$ will be used later.

The optimality conditions and constraint qualifications we work with in this paper are those originally proposed by Karush, Kuhn, and Tucker (see, for instance, the books [21] and [22]).

For further analysis, we introduce the active sets

$$\begin{aligned} \mathcal{A}(x) &= \Big\{ i \in \{1, \ldots, n\} : x_i = 0 \Big\} , \\ \mathcal{A}(y) &= \Big\{ i \in \{1, \ldots, m\} : y_i = 0 \Big\} , \quad \text{and} \\ \mathcal{A}(u) &= \Big\{ i \in \{1, \ldots, n-m\} : u_i = 0 \Big\} , \end{aligned}$$

and their corresponding complementary sets:

$$\begin{aligned} \mathcal{N}(x) &= \{1, \ldots, n\} \setminus \mathcal{A}(x) \, , \\ \mathcal{N}(y) &= \{1, \ldots, m\} \setminus \mathcal{A}(y) \, , \quad \text{and} \\ \mathcal{N}(u) &= \{1, \ldots, n-m\} \setminus \mathcal{A}(u) \, . \end{aligned} \tag{2.3}$$

The Karush–Kuhn–Tucker (KKT) optimality conditions are necessary conditions for $x_*$ to be a local solution of (1.1) if certain constraint qualifications are satisfied at this point. An often used, but rather strong constraint qualification is given in the following definition.

DEFINITION 2.1. *A point $x_*$ satisfying $C(x_*) = 0$ and $x_* \geq 0$ is called feasible. A feasible point $x_*$ is said to be regular if the matrix*

$$\begin{pmatrix} C_y(x_*) & C_u(x_*) \\ I_{\mathcal{A}(y_*)} & 0 \\ 0 & I_{\mathcal{A}(u_*)} \end{pmatrix}$$

*has full row rank, where $I_{\mathcal{A}(y_*)}$ and $I_{\mathcal{A}(u_*)}$ are submatrices of the identity matrix formed by rows corresponding to indices in $\mathcal{A}(y_*)$ and $\mathcal{A}(u_*)$.*

For discretized optimal control problems of the form (1.1) with state constraints, regularity is not satisfied if the cardinal of $\mathcal{A}(y_*) \cup \mathcal{A}(u_*)$ is larger than $n - m$. For problems with bounds only on the control variables regularity is always given independently of the cardinal of $\mathcal{A}(u_*)$, provided $C_y(x_*)$ is nonsingular. For more details see [1] and [10].

However, the KKT optimality conditions are indeed necessary conditions for $x_*$ to be a local solution of (1.1) if the constraint qualifications are satisfied. We recall the definition of KKT constraint qualifications (see [21] and [22]).

DEFINITION 2.2. *A feasible point $x_*$ satisfies the first–order KKT constraint qualification if for all vectors $p \neq 0$ verifying*

$$\begin{pmatrix} C_y(x_*) & C_u(x_*) \end{pmatrix} \begin{pmatrix} p_y \\ p_u \end{pmatrix} = 0\,, \tag{2.4}$$

$$(p_y)_{\mathcal{A}(y_*)} \geq 0\,, \quad \text{and} \quad (p_u)_{\mathcal{A}(u_*)} \geq 0\,, \tag{2.5}$$

*there exists a continuously differentiable arc $x(t)$, $t \in [0, \bar{t})$, with $\bar{t} > 0$, such that $x(t)$ is feasible for all $t \in [0, \bar{t})$, $x(0) = x_*$, and $x'(0) = p$.*

*A feasible point $x_*$ satisfies the second–order KKT (or strong) constraint qualification if for all nonzero vectors $p$ verifying (2.4),*

$$(p_y)_{\mathcal{A}(y_*)} = 0\,, \quad \text{and} \quad (p_u)_{\mathcal{A}(u_*)} = 0\,, \tag{2.6}$$

*there exists a twice continuously differentiable arc $x(t)$, $t \in [0, \bar{t})$, with $\bar{t} > 0$, such that $x(t)$ is feasible for all $t \in [0, \bar{t})$, $x(0) = x_*$, $x'(0) = p$, and*

$$y(t)_{\mathcal{A}(y_*)} = 0\,, \quad u(t)_{\mathcal{A}(u_*)} = 0 \quad \text{for all} \quad t \in [0, \bar{t})\,. \tag{2.7}$$

In the next proposition we show that, if $-\left(C_y(x_*)^{-1} C_u(x_*) p_u\right)_i \neq 0$ for all $i \in \mathcal{A}(y_*)$ and all nonzero $p_u$ such that $(p_u)_{\mathcal{A}(u_*)} \geq 0$, then the KKT constraint qualifications hold at the point $x_*$.

PROPOSITION 2.1. *The first–order and second–order KKT constraint qualifications hold at the feasible point $x_*$ if*

$$\text{for all nonzero} \quad p_u \in \mathbb{R}^{n-m} \quad \text{with} \quad (p_u)_{\mathcal{A}(u_*)} \geq 0\,, \quad \left(-C_y(x_*)^{-1} C_u(x_*) p_u\right)_i \neq 0 \quad \text{for all} \quad i \in \mathcal{A}(y_*)\,. \tag{2.8}$$

*Proof.* Let $x_* = (y_*^\top \; u_*^\top)^\top$ be a feasible point for (1.1) and consider a vector $p = (p_y^\top \; p_u^\top)^\top$ satisfying (2.4).

First, let us prove that the first–order constraint qualification holds. So, let also $p$ satisfy (2.5). Now let us define the following arc:

$$x(t) \;=\; \begin{pmatrix} y(t) \\ u(t) \end{pmatrix} \;=\; \begin{pmatrix} y\,(u_* + tp_u) \\ \\ u_* + tp_u \end{pmatrix}, \tag{2.9}$$

where $y(u)$ is the vector–valued function defined in (2.1). The arc (2.9) is continuously differentiable. It is easy to see that $x(0) = x_*$. Also, from (2.4) and the definition of the arc (2.9),

$$\begin{aligned} y'(0) \;&=\; -C_y(x_*)^{-1} C_u(x_*) p_u \;=\; p_y\,, \\ u'(0) \;&=\; p_u\,, \end{aligned} \tag{2.10}$$

and this shows that $x'(0) = p$. It remains to show that the arc (2.9) is feasible. If the arc (2.9) is not feasible then there exists an index $i$ in $\mathcal{A}(y_*)$ and a sequence $\{t_k\}$ of positive numbers converging to zero for which $y(u_*)_i = 0$ and $y(u_* + t_k p_u)_i < 0$ for all $k$. If we define $h(t) = y(u_* + tp_u)_i$, we conclude that $h'(0) \leq 0$. From (2.5), (2.10), and $(p_y)_i = h'(0) \geq 0$, we have $h'(0) = 0$. But this contradicts assumption (2.8). Thus, the arc (2.9) is feasible.

The second–order constraint qualification also holds because there is no vector $p$ satisfying (2.4), (2.6), and (2.8) simultaneously. In fact (2.6) says that $(p_y)_{\mathcal{A}(y_*)} = 0$, and (2.4) and (2.8) together imply that $(p_y)_i \neq 0$ for all $i \in \mathcal{A}(y_*)$. □

We depict the situation where the condition (2.8) is false in Figure 2.1 for a problem that satisfies Assumptions 2.1.

Proposition 2.1 is important because it allows us to state necessary (first and second order) optimality conditions in the absence of regularity.

A point $x_*$ satisfies the first–order optimality conditions for problem (1.1) if there exist $\lambda_* \in \mathbb{R}^m$ and $z_* \in \mathbb{R}^n$ such that

$$\begin{aligned} &C(x_*) = 0,\; x_* \geq 0\,, \\ &\nabla_x \ell(x_*, \lambda_*) - z_* = 0\,, \\ &(x_*)_i\,(z_*)_i = 0\,, \quad i = 1, \ldots, n\,, \\ &z_* \geq 0\,. \end{aligned}$$

Consider a point $(x_*, \lambda_*, z_*)$ satisfying the first–order optimality conditions. We say that such point $(x_*, \lambda_*, z_*)$ satisfies the second–order necessary optimality conditions if $s^\top \nabla^2_{xx} \ell(x_*, \lambda_*) s \geq 0$ for all $s$ verifying $J(x_*)s = 0$ and $s_i = 0$ if $i \in \mathcal{A}(x_*)$. On the other hand, the second–order sufficient optimality conditions require $s^\top \nabla^2_{xx} \ell(x_*, \lambda_*) s > 0$ for all $s \neq 0$ verifying $J(x_*)s = 0$ and $s_i = 0$ if $i \in \mathcal{A}(x_*)$ and $(z_*)_i > 0$.

**3. Optimality conditions using affine–scaling matrices.** Now we define the matrix $D(x, \lambda)$ to be the diagonal matrix with diagonal elements given by:

$$\Big(D(x,\lambda)\Big)_{ii} = \begin{cases} (x_i)^{\frac{1}{2}} & \text{if} \quad (\nabla_x \ell(x,\lambda))_i \geq 0\,, \\ \\ 1 & \text{if} \quad (\nabla_x \ell(x,\lambda))_i < 0\,, \end{cases}$$

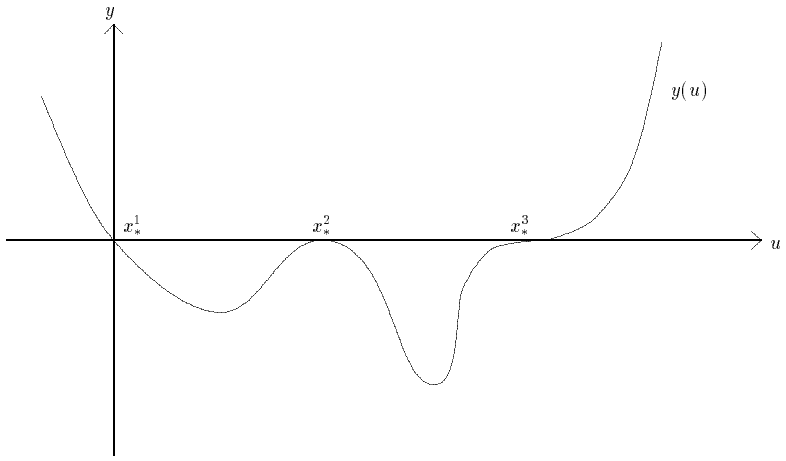


FIG. 2.1. *The points $x_*^1$, $x_*^2$, and $x_*^3$ are not regular. The points $x_*^2$ and $x_*^3$ do not satisfy the first and second order KKT constraint qualifications since the derivative $y'(u) = -\frac{C_u(x)}{C_y(x)}$ is zero at these points. The point $x_*^1$ satisfies the constraint qualifications.*

for $i = 1, \ldots, n$.

In the following proposition we give the form of the first–order and second–order necessary optimality conditions that results of using this matrix instead of the multipliers $z$. The proof is straightforward, and we omit it. A similar result is stated in [6], [7].

PROPOSITION 3.1. *A point $x_*$ satisfies the first–order optimality conditions if and only if there exists $\lambda_* \in \mathbb{R}^m$ such that*

$$D(x_*, \lambda_*)^2 \nabla_x \ell(x_*, \lambda_*) = 0\,,$$
$$C(x_*) = 0, \quad x_* \geq 0\,.$$

*Such a point $(x_*, \lambda_*)$ satisfies the second–order necessary optimality conditions if and only if it satisfies the first–order optimality conditions and*

$$D(x_*, \lambda_*) \nabla^2_{xx} \ell(x_*, \lambda_*) D(x_*, \lambda_*)$$

*is positive semi–definite on the null space of $J(x_*)D(x_*, \lambda_*)$.*

*If the constraint qualification (2.8) holds, these conditions are necessary for $x_*$ to be a local minimizer of (1.1).*

The form of the sufficient optimality conditions that is stated using the affine–scaling matrix $D(x, \lambda)$ requires the definition of nondegeneracy or strict complementarity.

DEFINITION 3.1. *A feasible point $x_*$, with corresponding multipliers $\lambda_*$, is said to be nondegenerate if $(\nabla_x \ell(x_*, \lambda_*))_i = 0$ implies $(x_*)_i > 0$ for all $i \in \{1, \ldots, n\}$.*

We now define a diagonal matrix $E(x, \lambda)$ of order $n$ with diagonal elements given by

$$\Big(E(x,\lambda)\Big)_{ii} \;=\; \begin{cases} (\nabla_x \ell(x,\lambda))_i & \text{if} \quad (\nabla_x \ell(x,\lambda))_i > 0\,, \\ \\ 0 & \text{otherwise}\,, \end{cases}$$

for $i = 1, \ldots, n$. The significance of this matrix will become clear in the next sections. From the definitions of $D(x, \lambda)$ and $E(x, \lambda)$ we have the following property for which the proof is simple and also omitted.

PROPOSITION 3.2. *A nondegenerate point $x_*$, with corresponding multipliers $\lambda_*$, satisfies the second–order sufficient optimality conditions if and only if it satisfies the first–order conditions and*

$$D(x_*,\lambda_*)\nabla^2_{xx}\ell(x_*,\lambda_*)D(x_*,\lambda_*) + E(x_*,\lambda_*) \tag{3.1}$$

*is positive definite on the null space of $J(x_*)D(x_*, \lambda_*)$.*

The following proposition is important in the next section to justify the well–posedness of the affine–scaling and primal–dual interior–point Newton algorithms.

PROPOSITION 3.3.

*If $x_*$ is a regular point, then the matrix $D(x_*, \lambda_*)J(x_*)^\top$ has full column rank.*

*If the regular point $x_*$, with corresponding Lagrange multipliers $\lambda_*$, is such that the matrix (3.1) is positive definite, then the matrix*

$$\begin{pmatrix} D(x_*,\lambda_*)\nabla^2_{xx}\ell(x_*,\lambda_*)D(x_*,\lambda_*) + E(x_*,\lambda_*) & D(x_*,\lambda_*)J(x_*)^\top \\ J(x_*)D(x_*,\lambda_*) & 0 \end{pmatrix} \tag{3.2}$$

*is nonsingular.*

*Proof.* Let us consider a vector $s_y \in \mathbb{R}^m$ such that

$$D(x_*,\lambda_*)\begin{pmatrix} C_y(x_*)^\top \\ C_u(x_*)^\top \end{pmatrix} s_y \;=\; 0\,.$$

We need to prove that $s_y = 0$. It follows from the definition of $D(x_*, \lambda_*)$ that

$$\begin{aligned} \Big(C_y(x_*)^\top s_y\Big)_i = 0 \quad &\text{for all} \quad i \in \mathcal{N}(y_*) \quad \text{and} \\ \Big(C_u(x_*)^\top s_y\Big)_i = 0 \quad &\text{for all} \quad i \in \mathcal{N}(u_*)\,, \end{aligned} \tag{3.3}$$

where the sets $\mathcal{N}(y_*)$ and $\mathcal{N}(u_*)$ are defined in (2.3). Now we define

$$\begin{aligned} r_{\mathcal{A}(y_*)} &= -\Big(C_y(x_*)^\top s_y\Big)_{\mathcal{A}(y_*)} \in \mathbb{R}^{|\mathcal{A}(y_*)|} \quad \text{and} \\ r_{\mathcal{A}(u_*)} &= -\Big(C_u(x_*)^\top s_y\Big)_{\mathcal{A}(u_*)} \in \mathbb{R}^{|\mathcal{A}(u_*)|}\,. \end{aligned} \tag{3.4}$$

From (3.3) and (3.4), we obtain

$$\begin{pmatrix} C_y(x_*)^\top & I^\top_{\mathcal{A}(y_*)} & 0 \\ C_u(x_*)^\top & 0 & I^\top_{\mathcal{A}(u_*)} \end{pmatrix} \begin{pmatrix} s_y \\ r_{\mathcal{A}(y_*)} \\ r_{\mathcal{A}(u_*)} \end{pmatrix} \;=\; 0\,,$$

which in turn, by using the definition of regularity given in Definition 2.1, implies that $s_y = 0$. Thus, we have proved that $D(x_*, \lambda_*)J(x_*)^\top$ has full column rank.

Since $D(x_*, \lambda_*)\nabla^2_{xx}\ell(x_*, \lambda_*)D(x_*, \lambda_*) + E(x_*, \lambda_*)$ is positive definite on the null–space of the matrix $J(x_*)D(x_*, \lambda_*)$ and its transpose, $D(x_*, \lambda_*)J(x_*)^\top$, has full column rank, we conclude that the matrix (3.2) is nonsingular (see [26]). □

If $x_*$ is a regular point then the optimality conditions can be stated by using a null–space basis representation of $J(x_*)D(x_*, \lambda_*)$.

PROPOSITION 3.4. *Let $x_*$ be a regular point and $W(x_*, \lambda_*)$ a matrix whose columns form a basis for the null space of $J(x_*)D(x_*, \lambda_*)$, where $\lambda_*$ is a vector of Lagrange multipliers.*

- *The point $(x_*, \lambda_*)$ satisfies the first–order optimality conditions if and only if*

$$W(x_*, \lambda_*)^\top D(x_*, \lambda_*)^2 \nabla_x \ell(x_*, \lambda_*) = 0\,,$$
$$C(x_*) = 0, \quad x_* \geq 0\,.$$

- *The point $(x_*, \lambda_*)$ satisfies the second–order necessary optimality conditions if and only if it satisfies the first–order optimality conditions and*

$$W(x_*, \lambda_*)^\top D(x_*, \lambda_*)\nabla^2_{xx}\ell(x_*, \lambda_*)D(x_*, \lambda_*)W(x_*, \lambda_*)$$

  *is positive semi–definite.*
  *If the constraint qualification (2.8) holds, these conditions are necessary for $x_*$ to be a local minimizer of (1.1).*
- *The nondegenerate point $(x_*, \lambda_*)$ satisfies the second–order sufficient optimality conditions if and only if it satisfies the first–order optimality conditions and*

$$W(x_*, \lambda_*)^\top \Big(D(x_*, \lambda_*)\nabla^2_{xx}\ell(x_*, \lambda_*)D(x_*, \lambda_*) + E(x_*, \lambda_*)\Big)W(x_*, \lambda_*)$$

  *is positive definite.*

**4. An affine–scaling interior–point Newton algorithm.** To motivate the affine–scaling interior–point Newton algorithm described in this paper we consider the application of Newton's method to the system of nonlinear equations in $x$ and $\lambda$

$$D(x, \lambda)^2 \nabla_x \ell(x, \lambda) = 0\,, \tag{4.1}$$

$$C(x) = 0\,. \tag{4.2}$$

This algorithm is of the interior–point type, meaning that $x$ is required always to be strictly feasible with respect to the bound constraints, in other words $x$ is such that $x > 0$. We will assume throughout this section that $x > 0$.

The diagonal element functions in $D(x, \lambda)^2$ are typically discontinuous at points where the equality $(\nabla_x \ell(x, \lambda))_i = 0$ is satisfied, but the vector–valued function $D(x, \lambda)^2 \nabla_x \ell(x, \lambda)$ is continuous (but not differentiable) at such type of points. The application of Newton's method to this type of systems of nonlinear equations has first been suggested in [6] in the context of nonlinear minimization problems with simple bounds. It is shown in [6] that this type of nondifferentiability still allows the Newton process to achieve local q–quadratic convergence. This has also been shown in [10] for problems of the type (1.1) but with no bounds on the state variables $y$.

A linearization of (4.1)–(4.2) is of the form

$$\begin{pmatrix} D(x,\lambda)^2\nabla^2_{xx}\ell(x,\lambda)+E(x,\lambda) & D(x,\lambda)^2J(x)^\top \\ J(x) & 0 \end{pmatrix} \begin{pmatrix} s \\ \Delta\lambda \end{pmatrix} = -\begin{pmatrix} D(x,\lambda)^2\nabla_x\ell(x,\lambda) \\ C(x) \end{pmatrix}. \tag{4.3}$$

The second linear equation in (4.3) is the linearized state equation. The first linear equation in (4.3) has been derived by applying the product rule to (4.1) in the cases where we have differentiability. The diagonal elements of $E(x,\lambda)$ are the product of the derivative of the diagonal elements of $D(x,\lambda)^2$ and the components of the $\nabla_x\ell(x,\lambda)$. The derivative of $(D(x,\lambda)^2)_{ii}$ does not exist if $(\nabla_x\ell(x,\lambda))_i = 0$. In this case we set the corresponding quantities in the Jacobian to zero (see references [6], [7], [10]). This gives the first linear equation in (4.3).

Premultiplying on the left the first equation (4.3) by $D(x,\lambda)^{-1}$ and making the change of variables $\hat{s} = D(x,\lambda)^{-1}s$ yields

$$\begin{pmatrix} D(x,\lambda)\nabla^2_{xx}\ell(x,\lambda)D(x,\lambda)+E(x,\lambda) & D(x,\lambda)J(x)^\top \\ J(x)D(x,\lambda) & 0 \end{pmatrix} \begin{pmatrix} \hat{s} \\ \Delta\lambda \end{pmatrix} = -\begin{pmatrix} D(x,\lambda)\nabla_x\ell(x,\lambda) \\ C(x) \end{pmatrix}. \tag{4.4}$$

The form of the coefficient matrix in (4.4), Assumptions 2.1, and Proposition 3.3, together imply that the affine–scaling interior–point Newton algorithm is well defined in a neighborhood of a nondegenerate regular point that satisfies the second–order sufficient optimality conditions. We will return to this point later.

For further description of the algorithm, we need to make use of our notation to point out that:

$$s = \begin{pmatrix} s_y \\ s_u \end{pmatrix},$$

$$D(x,\lambda) = \begin{pmatrix} D_y(x,\lambda) & 0 \\ 0 & D_u(x,\lambda) \end{pmatrix}, \quad E(x,\lambda) = \begin{pmatrix} E_y(x,\lambda) & 0 \\ 0 & E_u(x,\lambda) \end{pmatrix},$$

and

$$\nabla_x\ell(x,\lambda) = \begin{pmatrix} \nabla_y\ell(x,\lambda) \\ \nabla_u\ell(x,\lambda) \end{pmatrix} = \begin{pmatrix} \nabla_y f(x)+C_y(x)^\top\lambda \\ \nabla_u f(x)+C_u(x)^\top\lambda \end{pmatrix}.$$

Using this we can rewrite the Newton equation (4.3) as follows

$$\begin{pmatrix} D_y(x,\lambda)^2\nabla^2_{yy}\ell(x,\lambda)+E_y(x,\lambda) & D_y(x,\lambda)^2\nabla^2_{yu}\ell(x,\lambda) & D_y(x,\lambda)^2C_y(x)^\top \\ D_u(x,\lambda)^2\nabla^2_{uy}\ell(x,\lambda) & D_u(x,\lambda)^2\nabla^2_{uu}\ell(x,\lambda)+E_u(x,\lambda) & D_u(x,\lambda)^2C_u(x)^\top \\ C_y(x) & C_u(x) & 0 \end{pmatrix} \cdot \begin{pmatrix} \begin{pmatrix} s_y \\ s_u \end{pmatrix} \\ \Delta\lambda \end{pmatrix} = -\begin{pmatrix} D_y(x,\lambda)^2\left(\nabla_y f(x)+C_y(x)^\top\lambda\right) \\ D_u(x,\lambda)^2\left(\nabla_u f(x)+C_u(x)^\top\lambda\right) \\ C(x) \end{pmatrix}. \tag{4.5}$$

From the first linear equation in (4.5), we obtain

$$\begin{aligned}\Delta\lambda \;&=\; -\Big( C_y(x)^{-\top}\nabla^2_{yy}\ell(x,\lambda) + C_y(x)^{-\top}D_y(x,\lambda)^{-2}E_y(x,\lambda) \quad C_y(x)^{-\top}\nabla^2_{yu}\ell(x,\lambda) \Big) \begin{pmatrix} s_y \\ s_u \end{pmatrix} \\ &\quad - C_y(x)^{-\top}\nabla_y f(x) \;-\; \lambda\,.\end{aligned} \tag{4.6}$$

The third linear equation, also called the linearized state equation, can be rewritten as

$$C_y(x)s_y + C_u(x)s_u \;=\; -C(x)\,.$$

Since $C_y(x)$ is nonsingular, we obtain

$$s \;=\; s^{\mathsf{n}} + W(x)s_u\,, \tag{4.7}$$

where

$$s^{\mathsf{n}} = \begin{pmatrix} -C_y(x)^{-1}C(x) \\ 0 \end{pmatrix} \tag{4.8}$$

is a particular solution of the linearized state equation, and

$$W(x) = \begin{pmatrix} -C_y(x)^{-1}C_u(x) \\ I_{n-m} \end{pmatrix} \tag{4.9}$$

is a matrix whose columns form a basis for the null space of the Jacobian matrix $J(x)$.

The second linear equation in (4.5) can now be rewritten only in the variables $s_u$ by using the formulae (4.6), (4.7), (4.8), and (4.9) given above. The form for this linear equation involves only the variables $s_u$ and is the following:

$$\Big(W(x)^\top H^a(x,\lambda)W(x)\Big)s_u \;=\; -W(x)^\top\Big(H^a(x,\lambda)s^{\mathsf{n}} + \nabla f(x)\Big)\,, \tag{4.10}$$

where

$$H^a(x,\lambda) \;=\; \begin{pmatrix} \nabla^2_{yy}\ell(x,\lambda) + D_y(x,\lambda)^{-2}E_y(x,\lambda) & \nabla^2_{yu}\ell(x,\lambda) \\ \nabla^2_{uy}\ell(x,\lambda) & \nabla^2_{uu}\ell(x,\lambda) + D_u(x,\lambda)^{-2}E_u(x,\lambda) \end{pmatrix} \tag{4.11}$$

is an augmentation of the Hessian $\nabla^2_{xx}\ell(x,\lambda)$ of the Lagrangian function $\ell(x,\lambda)$. The augmented term

$$G(x,\lambda) \;=\; \begin{pmatrix} G_y(x,\lambda) & 0 \\ 0 & G_u(x,\lambda) \end{pmatrix} \;=\; \begin{pmatrix} D_y(x,\lambda)^{-2}E_y(x,\lambda) & 0 \\ 0 & D_u(x,\lambda)^{-2}E_u(x,\lambda) \end{pmatrix}$$

takes into account the presence of the bound constraints in the variables $y$ and $u$.

In summary, the affine–scaling interior–point Newton algorithm is the following. (The step can be calculated by either 2.1a or 2.1b. The matrix $X$ is the diagonal matrix of order $n$ where the diagonal elements are the components of $x$.)

ALGORITHM 4.1 (AFFINE–SCALING INTERIOR–POINT NEWTON ALGORITHM).

1. Choose an initial point $(x, \lambda)$ with $x > 0$.
2. Until convergence do

2.1a Compute $(s, \Delta\lambda)$ by solving (4.4) for $(\hat{s}, \Delta\lambda)$ and then by setting $s = D(x,\lambda)\hat{s}$.

**or, equivalently,**

2.1b Compute $s^{\mathsf{n}}$ as in (4.8).
Compute $s_u$ by solving (4.10).
Compute $s_y = (s^{\mathsf{n}})_y - C_y(x)^{-1}C_u(x)s_u$.
Compute $\Delta\lambda$ by (4.6).

2.2 Set $\alpha = \tau\frac{-1}{\min(X^{-1}s,-1)}$, where $\tau \in (0,1)$.
Set the new iterate $(x,\lambda)$ to $(x,\lambda) + (\alpha s, \Delta\lambda)$.

In the next section, we will give an interpretation of the diagonal matrices $D(x,\lambda)$ and $E(x,\lambda)$ and show how to incorporate a centralization term in this algorithm (see Remark 5.1).

It is important to confirm that this algorithm is locally well–defined around a point that satisfies the standard Newton assumptions, i.e., a nondegenerate regular point for which the second–order sufficient optimality conditions hold. If Step 2.1a is used, then the well–posedness of the algorithm is a direct consequence of Assumptions 2.1 and Proposition 3.3. So, let us consider Step 2.1b. Since $C_y(x)$ is assumed to be nonsingular in $\Omega$, the calculations of $s^{\mathsf{n}}$ and $\Delta\lambda$ are well defined. We show in the next result that it is always possible to compute $s_u$ in (4.10).

PROPOSITION 4.1. *Let Assumptions 2.1 hold. If $(x,\lambda)$ is sufficiently close to a nondegenerate regular point $(x_*,\lambda_*)$ for which the second–order sufficient optimality conditions hold and if $x$ is strictly feasible, then the matrix*

$$W(x)^\top H^a(x,\lambda)W(x) \tag{4.12}$$

*is nonsingular.*

*Proof.* If the matrix (4.12) is singular, there exists an $s_u \neq 0$ such that

$$\Big(W(x)^\top H^a(x,\lambda)W(x)\Big)s_u \;=\; 0\,.$$

Let $s^{\mathsf{n}} = 0$, $s = (s_y^\top \; s_u^\top)^\top = s^{\mathsf{n}} + W(x)s_u$, and $\Delta\lambda$ be given by (4.6). Then

$$\begin{pmatrix} D(x,\lambda)\nabla^2_{xx}\ell(x,\lambda)D(x,\lambda) + E(x,\lambda) & D(x,\lambda)J(x)^\top \\ J(x)D(x,\lambda) & 0 \end{pmatrix}\begin{pmatrix} \hat{s} \\ \Delta\lambda \end{pmatrix} \;=\; 0\,, \tag{4.13}$$

where $\hat{s} = (\hat{s}_y^\top \; \hat{s}_u^\top)^\top$, $\hat{s}_y = D_y(x,\lambda)^{-1}s_y$, and $\hat{s}_u = D_u(x,\lambda)^{-1}s_u$. Since $\hat{s}_u \neq 0$, this shows that the matrix in (4.13) is singular. From Assumptions 2.1, Proposition 3.3, and the proximity of $(x,\lambda)$ and $(x_*,\lambda_*)$, we know that the matrix in (4.13) is nonsingular, which establishes a contradiction. □

It is important to remark that the matrix (4.12) is not defined at $(x_*,\lambda_*)$ if any of the diagonal elements of $D(x_*,\lambda_*)$ is zero, meaning that one of the components of $x_*$ is also zero. In this case the condition number of (4.12) is arbitrarily large for points close to $(x_*,\lambda_*)$. This statement could lead to the impression that the q–quadratic convergence of the algorithm is jeopardized if Step 2.1b is used. However, we point out that we only suggest the use of (4.10) as an alternative and equivalent way to compute the step (4.4), and we stress that the matrix in (4.13) is nonsingular at a point satisfying the standard Newton assumptions.

Now, we can guarantee the q–quadratic rate of convergence of Algorithm 4.1.

THEOREM 4.1. *Let Assumptions 2.1 hold and consider a sequence in the pair $(x, \lambda)$ generated by Algorithm 4.1 converging to a nondegenerate regular point $(x_*, \lambda_*)$ that satisfies the second–order sufficient optimality conditions. If $\tau$ is chosen so that $|\tau - 1| = \mathcal{O}(\|F_1(x, \lambda)\|)$, where*

$$F_1(x, \lambda) \;=\; \begin{pmatrix} D(x, \lambda)^2 \nabla_x \ell(x, \lambda) \\ C(x) \end{pmatrix},$$

*then the sequence converges with a q–quadratic rate.*

*The corresponding sequence in $z$ generated by $z = z(x, \lambda) = \nabla_x \ell(x, \lambda)$ converges r–quadratically to $z_* = z(x_*, \lambda_*)$.*

*Proof.* As we have seen before the algorithm is well defined. The q–quadratic rate follows by using an argument similar to the one used in [7, Theorem 11] or [10, Corollary 9.1] for simpler classes of problems. For the r–quadratic rate, we point out that Assumptions 2.1 imply

$$\|z(x, \lambda) - z(x_*, \lambda_*)\| \;\leq\; \gamma_1 \left\| \begin{pmatrix} x - x_* \\ \lambda - \lambda_* \end{pmatrix} \right\|,$$

for some positive constant $\gamma_1$. Since the pair $(x, \lambda)$ converges q–quadratically to $(x_*, \lambda_*)$, $z(x, \lambda)$ converges r–quadratically to $z(x_*, \lambda_*)$. ∎

The update (4.6) can be rewritten as follows:

$$\Delta\lambda \;=\; -C_y(x)^{-\top} \Big( (H^a(x, y)s)_y + \nabla_y f(x) \Big) - \lambda \,.$$

If we set $\lambda_{new} = \lambda + \Delta\lambda$, then we have

$$\lambda_{new} \;=\; -C_y(x)^{-\top} \Big( \Big( \nabla^2_{xx} \ell(x, \lambda)s \Big)_y + \nabla_y f(x) \Big) - C_y(x)^{-\top} \left( G(x, \lambda)s \right)_y \,. \tag{4.14}$$

This suggests that we consider the following update:

$$\lambda_{new} \;=\; -C_y(x + s)^{-\top} \Big( \nabla_y f(x + s) + G_y(x, \lambda)s_y \Big) \,. \tag{4.15}$$

We can see that the update (4.15) is related with the update (4.14) is the same way that the least squares multipliers are related with the Newton multipliers for equality constrained optimization (see [23, Formulae (2.7e) and (2.9)]). Later, when we introduce the reduced primal–dual interior–point Newton algorithm, we will see clearly the role of the term $-C_y(x + s)^{-\top} G_y(x, \lambda)s_y$. This is the topic of Remark 5.2, where we will see for instance why in (4.15) we have $G(x, \lambda)$ rather than $G(x + s, \lambda + \Delta\lambda)$. One of the advantages of (4.15) is that it is not affected by possible inaccuracies in $\nabla^2_{xx} \ell(x, \lambda)$. Also, if there are no bounds on $y$, then the update (4.15) reduces to the update $\lambda_{new} = -C_y(x + s)^{-\top} \nabla_y f(x + s)$ that is used by other algorithms in this situation (see [10], [16]).

**5. Primal–dual interior–point Newton algorithms.** The primal–dual interior–point Newton algorithm is derived by applying Newton's method to the following system of nonlinear equations in $x$, $\lambda$, and $z$

$$\nabla_x \ell(x, \lambda) - z = 0 \,, \tag{5.1}$$

$$C(x) = 0 \,, \tag{5.2}$$

$$XZe = \mu e \,, \tag{5.3}$$

where $\mu = \sigma \min\{x_i z_i,\ i = 1, \ldots, n\}$ is the perturbation parameter[1], $\sigma \in (0, 1)$ is the centralization parameter, $X$ and $Z$ are diagonal matrices where the diagonal elements are the components of $x$ and $z$ respectively, and $e$ is a vector of ones with $n$ components. The equation $XZe = \mu e$ is a relaxation of the complementarity condition $x^\top z = 0$ that includes a perturbation term $\mu e$ (see [12] and [32] for more details). This algorithm is also of interior–point type, meaning that $x$ and $z$ are required always to be strictly feasible with respect to the bound constraints, i.e., $x$ and $z$ have to satisfy $x > 0$ and $z > 0$. We will assume that $x > 0$ and $z > 0$ throughout this section.

The linearization of (5.1)–(5.3) yields

$$\begin{pmatrix} \nabla^2_{xx}\ell(x,\lambda) & J(x)^\top & -I_n \\ J(x) & 0 & 0 \\ Z & 0 & X \end{pmatrix} \begin{pmatrix} s \\ \Delta\lambda \\ \Delta z \end{pmatrix} = - \begin{pmatrix} \nabla_x \ell(x,\lambda) - z \\ C(x) \\ XZe - \mu e \end{pmatrix}. \tag{5.4}$$

We can eliminate $\Delta z$ from the third equation

$$\Delta z = -X^{-1}Zs - z + \mu X^{-1} e \tag{5.5}$$

and replace it in the first equation:

$$\Big(\nabla^2_{xx}\ell(x,\lambda) + X^{-1}Z\Big)s + J(x)^\top \Delta\lambda = -\nabla_x \ell(x,\lambda) + \mu X^{-1} e\,.$$

Thus, the calculation of the steps $s$ and $\Delta\lambda$ can be done by solving

$$\begin{pmatrix} X^{\frac{1}{2}}\nabla^2_{xx}\ell(x,\lambda)X^{\frac{1}{2}} + Z & X^{\frac{1}{2}}J(x)^\top \\ J(x)X^{\frac{1}{2}} & 0 \end{pmatrix} \begin{pmatrix} \hat{s} \\ \Delta\lambda \end{pmatrix} = - \begin{pmatrix} X^{\frac{1}{2}}\left(\nabla_x \ell(x,\lambda) - \mu X^{-1} e\right) \\ C(x) \end{pmatrix} \tag{5.6}$$

and setting $s = X^{\frac{1}{2}}\hat{s}$. By computing the step in this way, we obtain a symmetric matrix where $X$ is not inverted.

We already know that the second equation in (5.4) leads to $s = s^{\mathsf{n}} + W(x)s_u$, where $s^{\mathsf{n}}$ and $W(x)$ are given in (4.8) and (4.9), respectively. Thus, we can find $s_u$ by solving

$$\Big(W(x)^\top H^a(x,\lambda)W(x)\Big)s_u = -W(x)^\top \Big(H^a(x,\lambda)s^{\mathsf{n}} + \nabla f(x) - \mu X^{-1} e\Big)\,, \tag{5.7}$$

where

$$H^a(x,\lambda) = \begin{pmatrix} \nabla^2_{yy}\ell(x,\lambda) + X_y^{-1}Z_y & \nabla^2_{yu}\ell(x,\lambda) \\ \nabla^2_{uy}\ell(x,\lambda) & \nabla^2_{uu}\ell(x,\lambda) + X_u^{-1}Z_u \end{pmatrix}. \tag{5.8}$$

REMARK 5.1. *The similarity between the formula (5.7) to compute $s_u$ and the formula (4.10) used in the affine–scaling algorithm is evident. A comparison of the formulas (4.10)–(4.11) and (5.7)–(5.8) indicates that the pair $D(x,\lambda)$ and $E(x,\lambda)$ plays in the affine–scaling algorithm a role identical to the one that $X^{\frac{1}{2}}$ and $Z$ plays in the primal–dual algorithm. In fact at a point $(x_*, \lambda_*)$ satisfying the first–order optimality conditions, the following equalities hold*

$$D(x_*,\lambda_*) = X_*^{\frac{1}{2}} \quad \textit{and} \quad E(x_*,\lambda_*) = Z_*\,.$$

[1] A more general form for the perturbation parameter is given in [31].

*A comparison between the affine–scaling and the primal–dual algorithms allow us to introduce centralization in the affine–scaling algorithm. We observe from formulas (4.4) and (5.6) that centralization in the affine–scaling algorithm could be obtained if we replace* $-D(x,\lambda)\nabla_x \ell(x,\lambda)$ *by* $-D(x,\lambda)\nabla_x \ell(x,\lambda) + \mu D(x,\lambda)^{-1}e$ *in the right hand side of (4.4). The perturbation parameter would be given by* $\mu = \sigma \min\{x_i(E(x,\lambda))_{ii} : (E(x,\lambda))_{ii} > 0\,,\ i \in \{1,\ldots,n\}\}$ *with* $\sigma \in (0,1)$ *the centralization parameter.*

The formula for $\Delta\lambda$ in this case is slightly different from (4.6). The solution component $\Delta\lambda$ of (5.4) is given by formula (5.9):

$$\begin{aligned} \Delta\lambda \;=\; & -\left( C_y(x)^{-\top}\nabla^2_{yy}\ell(x,\lambda) + C_y(x)^{-\top}X_y^{-1}Z_y \quad C_y(x)^{-\top}\nabla^2_{yu}\ell(x,\lambda) \right) \begin{pmatrix} s_y \\ s_u \end{pmatrix} \\ & -\; C_y(x)^{-\top}\nabla_y f(x) \;-\; \lambda + \mu C_y(x)^{-\top}X_y^{-1}e_y\,. \end{aligned} \tag{5.9}$$

The primal–dual interior–point Newton algorithm is summarized below. We point out that the procedures described in 2.1a, 2.1b, and 2.1c correspond to equivalent forms of calculating the same step.

ALGORITHM 5.1 (PRIMAL–DUAL INTERIOR–POINT NEWTON ALGORITHM).

1. Choose an initial point $(x,\lambda,z)$ with $x > 0$ and $z > 0$.
2. Until convergence do
   - 2.1a Compute $(s,\Delta\lambda,\Delta z)$ by solving (5.4).
   - **or, equivalently,**
   - 2.1b Compute $(\hat{s},\Delta\lambda)$ by solving (5.6) and set $s = X^{\frac{1}{2}}\hat{s}$.
     Compute $\Delta z$ by (5.5).
   - **or, equivalently,**
   - 2.1c Compute $s^{\mathsf{n}}$ as in (4.8).
     Compute $s_u$ by solving (5.7).
     Compute $s_y = (s^{\mathsf{n}})_y - C_y(x)^{-1}C_u(x)s_u$.
     Compute $\Delta\lambda$ by (5.9).
     Compute $\Delta z$ by (5.5).
   - 2.2 Set $\alpha = \tau \frac{-1}{\min(X^{-1}s, Z^{-1}\Delta z, -1)}$, where $\tau \in (0,1)$.
     Set the new iterate $(x,\lambda,z)$ to $(x,\lambda,z) + (\alpha s, \Delta\lambda, \alpha\Delta z)$.

THEOREM 5.1. *Let Assumptions 2.1 hold and consider a sequence in the triple* $(x,\lambda,z)$ *generated by Algorithm 5.1 converging to a nondegenerate regular point* $(x_*,\lambda_*,z_*)$ *that satisfies the second–order sufficient optimality conditions. If* $\tau$ *and* $\sigma$ *are chosen so that* $|\tau - 1| = \mathcal{O}(\|F_2(x,\lambda,z)\|)$ *and* $\sigma = \mathcal{O}(\|F_2(x,\lambda,z)\|)$*, where*

$$F_2(x,\lambda,z) \;=\; \begin{pmatrix} \nabla_x \ell(x,\lambda) - z \\ C(x) \\ XZe \end{pmatrix},$$

*then the sequence converges with a q–quadratic rate.*

*Proof.* If $(x,\lambda,z)$ is sufficiently close to a nondegenerate regular point $(x_*,\lambda_*,z_*)$ for which the second–order sufficient optimality conditions hold and if $x$ and $z$ are strictly feasible, then the coefficient matrix in (5.4) is nonsingular. In fact, this matrix is nonsingular at $(x_*,\lambda_*,z_*)$. See [12, Proposition 4.1].

By appealing to an argument similar to the one used in Proposition 4.1 to prove that (4.12) is nonsingular, we establish that

$$W(x)^\top H^a(x,\lambda)W(x)$$

is also nonsingular, where $H^a(x,\lambda)$ is given by (5.8). Thus, the algorithm is well defined.

The proof of q–quadratic convergence is given in [12, Theorem 5.1]. □

Now we introduce a reduced primal–dual interior–point Newton algorithm for the solution of problem (1.1). First, we note that

$$\nabla_x \ell(x,\lambda) - z \;=\; \begin{pmatrix} \nabla_y f(x) + C_y(x)^\top \lambda - z_y \\ \nabla_u f(x) + C_u(x)^\top \lambda - z_u \end{pmatrix}.$$

Thus, from the first–order optimality conditions, we can obtain the following formula for the multipliers $\lambda$

$$\lambda(x,z) \;=\; -C_y(x)^{-\top}\Big(\nabla_y f(x) - z_y\Big). \tag{5.10}$$

If we set $\lambda = \lambda(x,z)$ in $\nabla_u \ell(x,\lambda) - z_u = 0$, we get

$$W(x)^\top\Big(\nabla f(x) - z\Big) = 0\,.$$

The reduced primal–dual interior–point Newton algorithm is derived by applying Newton's method to the following system of nonlinear equations in $x$ and $z$

$$W(x)^\top\Big(\nabla f(x) - z\Big) = 0\,, \tag{5.11}$$

$$C(x) = 0\,, \tag{5.12}$$

$$XZe = \mu e\,, \tag{5.13}$$

where $X$, $Z$, $e$, and $\mu$ are described after (5.1)–(5.3).

To calculate the corresponding Newton step we point out that

$$\begin{aligned}
\frac{\partial}{\partial y}\lambda(x,z) &= -C_y(x)^{-1}\nabla^2_{yy}\ell(x,\lambda(x,z))\,,\\
\frac{\partial}{\partial u}\lambda(x,z) &= -C_y(x)^{-1}\nabla^2_{yu}\ell(x,\lambda(x,z))\,,\\
\frac{\partial}{\partial x}W(x)^\top\Big(\nabla f(x) - z\Big) &= W(x)^\top\nabla^2_{xx}\ell(x,\lambda(x,z))\,.
\end{aligned}$$

(A slightly different version of these formulae is proved in [10].) Thus, the Newton step corresponding to the system of nonlinear equations (5.11)–(5.13) is given by

$$\begin{pmatrix} W(x)^\top\nabla^2_{xx}\ell(x,\lambda(x,z)) & -W(x)^\top \\ J(x) & 0 \\ Z & X \end{pmatrix}\begin{pmatrix} s \\ \Delta z \end{pmatrix} \;=\; -\begin{pmatrix} W(x)^\top\Big(\nabla f(x) - z\Big) \\ C(x) \\ XZe - \mu e \end{pmatrix}. \tag{5.14}$$

Again, we know that the second equation in (5.14) leads to $s = s^{\mathsf{n}} + W(x)s_u$, where $s^{\mathsf{n}}$ and $W(x)$ are given in (4.8) and (4.9), respectively. Thus, we can eliminate the third equation by using (5.5) and reduce the linear system (5.14) to

$$\left(W(x)^\top H^a(x,\lambda(x,z))W(x)\right)s_u \;=\; -W(x)^\top\left(H^a(x,\lambda(x,z))s^{\mathsf{n}} + \nabla f(x) - \mu X^{-1}e\right), \tag{5.15}$$

where

$$H^a(x,\lambda(x,z)) \;=\; \begin{pmatrix} \nabla^2_{yy}\ell(x,\lambda(x,z)) + X_y^{-1}Z_y & \nabla^2_{yu}\ell(x,\lambda(x,z)) \\ \nabla^2_{uy}\ell(x,\lambda(x,z)) & \nabla^2_{uu}\ell(x,\lambda(x,z)) + X_u^{-1}Z_u \end{pmatrix}.$$

The reduced primal–dual interior–point Newton algorithm is summarized below. The step is given by any of the equivalent forms described in 2.2a and 2.2b.

Algorithm 5.2 (Reduced Primal–dual interior–point Newton algorithm).

1. Choose an initial point $(x,z)$ with $x > 0$ and $z > 0$.
2. Until convergence do
   - 2.1 Compute $\lambda(x,z)$ by (5.10).
   - 2.2a Compute $(s,\Delta z)$ by solving (5.14).
   - **or, equivalently,**
   - 2.2b Compute $s^{\mathsf{n}}$ as in (4.8).
     Compute $s_u$ by solving (5.15).
     Compute $s_y = (s^{\mathsf{n}})_y - C_y(x)^{-1}C_u(x)s_u$.
     Compute $\Delta z$ by (5.5).
   - 2.3 Set $\alpha = \tau\frac{-1}{\min(X^{-1}s, Z^{-1}\Delta z, -1)}$, where $\tau \in (0,1)$.
     Set the new iterate $(x,z)$ to $(x,z) + \alpha(s,\Delta z)$.

Theorem 5.2. *Let Assumptions 2.1 hold and consider a sequence in the pair $(x,z)$ generated by Algorithm 5.2 converging to a nondegenerate regular point $(x_*,z_*)$ that satisfies the second–order sufficient optimality conditions. If $\tau$ and $\sigma$ are chosen so that $|\tau - 1| = \mathcal{O}(\|F_3(x,z)\|)$ and $\sigma = \mathcal{O}(\|F_3(x,z)\|)$, where*

$$F_3(x,z) \;=\; \begin{pmatrix} W(x)^\top\left(\nabla f(x) - z\right) \\ C(x) \\ XZe \end{pmatrix},$$

*then the sequence converges with a q–quadratic rate.*

*The corresponding sequence in $\lambda$ generated by $\lambda = \lambda(x,z)$ converges r–quadratically to $\lambda_* = \lambda(x_*,z_*)$.*

*Proof.* If $(x,z)$ is sufficiently close to a nondegenerate regular point $(x_*,z_*)$ for which the second–order sufficient optimality conditions hold and if $x$ and $z$ are strictly feasible, then the matrix

$$\begin{pmatrix} W(x)^\top\nabla^2_{xx}\ell(x,\lambda(x,z)) & -W(x)^\top \\ J(x) & 0 \\ Z & X \end{pmatrix} \tag{5.16}$$

is nonsingular. To see why this is true, notice that

$$\begin{pmatrix} W(x)^\top & 0 & 0 \\ 0 & I_m & 0 \\ 0 & 0 & I_n \end{pmatrix} \begin{pmatrix} \nabla^2_{xx}\ell(x,\lambda) & J(x)^\top & -I_n \\ J(x) & 0 & 0 \\ Z & 0 & X \end{pmatrix} = \begin{pmatrix} W(x)^\top \nabla^2_{xx}\ell(x,\lambda(x,z)) & 0 & -W(x)^\top \\ J(x) & 0 & 0 \\ Z & 0 & X \end{pmatrix}.$$

Since the first matrix on the left hand side of this equality has rank $2n$ and the second matrix on the left hand side is nonsingular (see Theorem 5.1), we conclude that the matrix on the right hand side has rank $2n$. Hence the matrix (5.16) is nonsingular at the points $(x,z)$ and $(x_*,z_*)$.

By appealing to an argument similar to the one used in Proposition 4.1 to prove that (4.12) is nonsingular, we obtain that

$$W(x)^\top H^a(x,\lambda(x,z))W(x)$$

is also nonsingular at a point $(x,z)$ in the conditions mentioned at the beginning of this proof. Thus, the algorithm is well defined.

The proof of q–quadratic convergence is given in [12, Theorem 5.1]. For the r–quadratic rate, we point out that Assumptions 2.1 imply

$$\|\lambda(x,z) - \lambda(x_*,z_*)\| \quad \leq \quad \gamma_2 \left\| \begin{pmatrix} x - x_* \\ z - z_* \end{pmatrix} \right\|,$$

for some positive constant $\gamma_2$. Since the pair $(x,z)$ converges q–quadratically to $(x_*,z_*)$, $\lambda(x,z)$ converges r–quadratically to $\lambda(x_*,z_*)$. □

One of our claims in this paper is that looking at these three interior–point Newton algorithms in a comprehensive manner is a tool to gain insight about their individual components and properties. In fact, Remarks 5.1 and 5.2 try to support this claim. In Remark 5.1, we pointed out that the affine–scaling algorithm lacks of centralization, and we showed how centralization could be incorporated into the algorithm. Next, we relate Lagrange multiplier updates of different algorithms.

REMARK 5.2. *By comparing the update (4.15) suggested in the affine–scaling algorithm to compute $\lambda_{new}$ and the update*

$$\lambda(x_{new},z_{new}) \;=\; -C_y(x_{new})^{-\top}\Big(\nabla_y f(x_{new}) - (z_{new})_y\Big).$$

*that was suggested in (5.10) for the reduced primal–dual algorithm, we can identify the role that the term $-C_y^{-\top}(x+s)G_y(x,\lambda)s_y$ plays in the update (4.15).*

*In fact, if we ignore the centralization term in the update (5.5), we obtain*

$$(z_{new})_y \;=\; z_y + \Delta z_y \;=\; -X_y^{-1}Z_y s_y\,.$$

*Thus, since $D_y(x,\lambda) \simeq X_y^{\frac{1}{2}}$ and $E_y(x,\lambda) \simeq Z_y$ (see Remark 5.1), we conclude that*

$$\begin{aligned} \lambda_{new} \;&=\; -C_y(x+s)^{-\top}\Big(\nabla_y f(x+s) + G_y(x,y)s_y\Big) \\ &\simeq\; -C_y(x_{new})^{-\top}\Big(\nabla_y f(x_{new}) - (z_{new})_y\Big) \;=\; \lambda(x_{new},z_{new})\,. \end{aligned}$$

**6. General nonlinear programming.** The general nonlinear programming can be formulated as:

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & g(x) = 0\,, \\ & x \geq 0\,, \end{array} \tag{6.1}$$

where $x \in \mathbb{R}^n$, $f : \Omega \longrightarrow \mathbb{R}$, $g : \Omega \longrightarrow \mathbb{R}^m$, $n$ and $m$ are positive integers satisfying $n > m$, and where $\Omega$ is an open set of $\mathbb{R}^n$ containing $\{x : x \geq 0\}$.

The main difference between problems (1.1) and (6.1) is that the equality constraints in (6.1) are assumed to have no particular structure. For instance, for problem (1.1) the Jacobian $J(x)$ of $C(x)$ is always partitioned in the form $\begin{pmatrix} C_y(x) & C_u(x) \end{pmatrix}$, where $C_y(x)$ is nonsingular. For problem (6.1), regularity implies that the Jacobian of $g(x)$ has full row rank. So, for every $x$, there exists a partitioning of the Jacobian of $g(x)$ of the form $\nabla g(x)^\top = \begin{pmatrix} B(x) & N(x) \end{pmatrix}$, where $B(x)$ is nonsingular. The problem is that this partitioning varies for different values of $x$.

Let us consider Algorithms 4.1, 5.1, and 5.2 with their full–space versions where the step $s$ is not decomposed. Algorithms 4.1 and 5.1 are extended to problem (6.1) in a straightforward way. The extension of Algorithm 5.2 requires some attention. In the place of the matrix $W(x)$ defined in (4.9), we use the matrix $Z(x)$ that is obtained by the QR factorization of $\nabla g(x)$, and in the place of the adjoint multiplier update (5.10), we use the least squares multiplier update $\lambda(x, z) = \text{argmin}\,\{\|\nabla_x \ell(x, \lambda) - z\|\}$. Then, the analysis follows by using the result established in [14].

Now let us consider the reduced–space versions of Algorithms 4.1, 5.1, and 5.2, where the step $s$ is computed along the null space of the Jacobian of the equality constraints. The extension to problem (6.1) is carried out by considering the decomposition $s = s^{\mathsf{n}} + Z(x)\bar{s}$, where $s^{\mathsf{n}}$ is orthogonal to the null space of $\nabla g(x)^\top$, and by using least squares updates for the multipliers step $\Delta\lambda$. See [23, Formula 2.6] for more details on how this reduction is accomplished.

**7. Summary.** We believe that the framework presented in this paper can lead to practical interior–point algorithms for the solution of discretized optimal control problems with bounds on the state variables.

We studied the optimality conditions and the constraint qualifications of problem (1.1) in detail. We described an affine–scaling interior–point Newton algorithm for (1.1) that relates to a particular form of the optimality conditions. This algorithm is an extension of the one in [10] for problems with bounds only on the control variables (see also [17], [28]). We also considered two primal–dual interior–point Newton algorithms for problem (1.1). The well–posedness of these algorithms was carefully studied. Then, we showed how to use the structure of (1.1) to reduce the linear algebra of every algorithm to the null space of the Jacobian matrix. The extension for general nonlinear programming was also considered.

We are currently studying the globalization of the interior–point Newton algorithms using the trust–region technique. The local rate of convergence of these algorithms in the singular case is another topic that deserves further attention. A number of authors have studied the local rate of convergence of interior–point Newton algorithms in the absence of regularity or strict complementarity for particular classes of problems like linear programming, monotone variational inequalities, and linear complementarity (see [13], [25], [27] and the references therein). In problems of the

form (1.1), conditions like regularity and strict complementarity (or nondegeneracy) are likely to be absent and this can of course influence the local rate of convergence and global convergence of the interior–point Newton algorithms.

**Acknowledgments.** The author would like to thank John Dennis and Matthias Heinkenschloss (Rice University) for their stimulating comments and two anonymous referees for their constructive reports.

## REFERENCES

[1] A. Battermann and M. Heinkenschloss, *Preconditioners for Karush–Kuhn–Tucker systems arising in the optimal control of distributed systems*, Tech. Report TR96–34, Department of Computational and Applied Mathematics, Rice University, 1996.

[2] L. T. Biegler, C. Schmid, and D. Ternet, *A multiplier–free, reduced Hessian method for process optimization.* Preprint, 1996.

[3] M. A. Branch, T. F. Coleman, and Y. Li, *A subspace, interior, and conjugate gradient method for large–scale bound–constrained minimization problems*, Tech. Report CTC95TR217, Advancing Computing Research Institute, Cornell University, 1995.

[4] R. H. Byrd, J. C. Gilbert, and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming.* Preprint, 1996.

[5] E. M. Cliff, M. Heinkenschloss, and A. Shenoy, *An optimal control problem for flows with discontinuities*, J. Optim. Theory Appl., (1997), pp. (to appear, August).

[6] T. F. Coleman and Y. Li, *On the convergence of interior–reflective Newton methods for nonlinear minimization subject to bounds*, Math. Programming, 67 (1994), pp. 189–224.

[7] ———, *An interior trust region approach for nonlinear minimization subject to bounds*, SIAM J. Optim., 6 (1996), pp. 418–445.

[8] T. F. Coleman and J. Liu, *An interior Newton method for quadratic programming*, Tech. Report TR93–1388, Department of Computer Science, Cornell University, 1993.

[9] I. Das, *An interior point algorithm for the general nonlinear programming problem with trust region globalization*, Tech. Report TR96–17, Department of Computational and Applied Mathematics, Rice University, 1996.

[10] J. E. Dennis, M. Heinkenschloss, and L. N. Vicente, *Trust–region interior–point SQP algorithms for a class of nonlinear programming problems*, Tech. Report TR94–45, Department of Computational and Applied Mathematics, Rice University, 1994. Revised November 1995. Appeared also as Tech. Rep. 94–12–01, Interdisciplinary Center for Applied Mathematics, Virginia Polytechnic Institute and State University.

[11] J. E. Dennis and L. N. Vicente, *On the convergence theory of general trust–region–based algorithms for equality–constrained optimization*, SIAM J. Optim., (To appear).

[12] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zhang, *On the formulation and theory of the Newton interior–point method for nonlinear programming*, J. Optim. Theory Appl., 89 (1996), pp. 507–541.

[13] A. S. El-Bakry, R. A. Tapia, and Y. Zhang, *On the convergence rate of Newton interior–point methods in the absence of strict complementarity*, Tech. Report TR95–25, Department of Computational and Applied Mathematics, Rice University, 1995.

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

[15] M. Heinkenschloss, *SQP methods for the solution of optimal control problems governed by the Navier Stokes equations.* In preparation.

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

[17] M. Heinkenschloss and L. N. Vicente, *Analysis of inexact trust–region interior–point SQP algorithms*, Tech. Report TR95–18, Department of Computational and Applied Mathematics, Rice University, 1995. Revised April 1996. Appeared also as Tech. Rep. 95–06–01, Interdisciplinary Center for Applied Mathematics, Virginia Polytechnic Institute and State University.

[18] 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.

[19] Y. Li, *On global convergence of a trust region and affine scaling method for nonlinearly constrained minimization*, Tech. Report CTC94TR197, Advanced Computing Research Institute, Cornell University, 1994.

[20] ———, *A trust region and affine scaling method for nonlinearly constrained minimization*, Tech. Report CTC94TR198, Advanced Computing Research Institute, Cornell University, 1994.

[21] O. L. Mangasarian, *Nonlinear Programming*, Classics in Applied Mathematics, Vol. 10, SIAM, Philadelphia, 1994.

[22] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw–Hill, New York, 1996.

[23] J. Nocedal and M. L. Overton, *Projected Hessian updating algorithms for nonlinearly constrained optimization*, SIAM J. Numer. Anal., 22 (1985), pp. 821–850.

[24] T. Plantenga, *Large–Scale Nonlinear Constrained Optimization using Trust Regions*, PhD thesis, Northwestern University, Evanston, Illinois, 1994.

[25] D. Ralph and S. J. Wright, *Superlinar convergence of an interior–point method despite dependent constraints*, Tech. Report ANL/MCS–P622–1196, Argonne National Laboratory, 1996.

[26] R. A. Tapia, *An introduction to the algorithms and theory of constrained optimization.* Lecture Notes, 1981.

[27] R. A. Tapia and Y. Zhang, *On the quadratic convergence of the singular Newton's method*, SIAG/OPT Views–and–News, (Fall 1992), pp. 6–8.

[28] L. N. Vicente, *Trust–Region Interior–Point Algorithms for a Class of Nonlinear Programming Problems*, PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251, USA, 1996.

[29] S. J. Wright, *Interior point methods for optimal control of discrete–time systems*, J. Optim. Theory Appl., 77 (1993), pp. 161–187.

[30] H. Yamashita, *A globally convergent primal–dual interior–point method for constrained optimization*, tech. report, Mathematical Systems Institute, Japan, 1992.

[31] H. Yamashita and H. Yabe, *Superlinear and quadratic convergence of some primal–dual interior point methods for constrained optimization*, Math. Programming, 75 (1996), pp. 377–397.

[32] Y. Zhang, R. A. Tapia, and J. E. Dennis, *On the superlinear and quadratic convergence of primal–dual interior point linear programming algorithms*, SIAM J. Optim., 2 (1992), pp. 304–324.