An Analysis of Newton's Method for Equivalent Karush–Kuhn–Tucker Systems*

Luís N. Vicente

Departamento de Matemática
Universidade de Coimbra
3000 Coimbra, Portugal
lvicente@mat.uc.pt

Abstract
In this paper we analyze the application of Newton's method to the solution of systems of nonlinear equations arising from equivalent forms of the first–order Karush–Kuhn–Tucker necessary conditions for constrained optimization. The analysis is carried out by using an abstract model for the original system of nonlinear equations and for an equivalent form of this system obtained by a reformulation that appears often when dealing with first–order Karush–Kuhn–Tucker necessary conditions. The model is used to determine the quantities that bound the difference between the Newton steps corresponding to the two equivalent systems of equations. The model is sufficiently abstract to include the cases of equality–constrained optimization, minimization with simple bounds, and also a class of discretized optimal control problems.

Keywords: Nonlinear programming, Newton's method, first–order Karush–Kuhn–Tucker necessary conditions.

1 Introduction

A popular technique to solve constrained optimization problems is to apply Newton's method to the system of nonlinear equations arising from the first–order necessary conditions. For instance, the system

$$\begin{aligned} &\nabla_x \ell(x,y) = \nabla f(x) + \nabla g(x) y = 0 \quad \text{and} \\ &g(x) = 0\,, \end{aligned} \tag{1}$$

*Support for this author has been provided by Instituto de Telecomunicações, JNICT, and Praxis XXI 2/2.1/MAT/346/94.

corresponds to the first–order necessary conditions of the equality–constrained optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix} = 0\,. \end{array}$$

(We assume that $x \in \mathbb{R}^n$ with $n > m$ and f and g are twice continuously differentiable functions with Lipschitz second derivatives. The Lagrangian function $\ell(x,y)$ is defined as $\ell(x,y) = f(x) + y^\top g(x)$.) The Newton step associated with the system (1) is given by:

$$\begin{pmatrix} \nabla^2_{xx}\ell(x,y) & \nabla g(x) \\ \nabla g(x)^\top & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -\begin{pmatrix} \nabla_x \ell(x,y) \\ g(x) \end{pmatrix}, \tag{2}$$

where $\nabla g(x)^\top$ represents the transpose of the Jacobian matrix $\frac{dg}{dx}(x)$ of $g(x)$, and $\nabla_x \ell(x,y)$ and $\nabla^2_{xx}\ell(x,y)$ are the gradient and the Hessian of the Lagrangian with respect to x, respectively.

There are cases in constrained optimization where the system of first–order necessary conditions is reformulated by eliminating variables and/or equations. For the example given above, we know that (1) is equivalent (with $x = \bar{x}$) to:

$$\begin{aligned} & Z(\bar{x})^\top \nabla f(\bar{x}) = 0 \quad \text{and} \\ & g(\bar{x}) = 0\,, \end{aligned} \tag{3}$$

where the columns of the orthogonal matrix $Z(\bar{x})$ form a basis for the null space of $\nabla g(\bar{x})^\top$. The matrix $Z(\bar{x})$ should be computed as described in [7] so that it can be extended smoothly in a neighborhood of $\bar{x}$ (see [7, Lemma 2.1]). The Newton step associated with (3) is defined by:

$$\begin{pmatrix} Z(\bar{x})^\top \nabla^2_{xx}\ell(\bar{x}, y(\bar{x})) \\ \nabla g(\bar{x})^\top \end{pmatrix} \overline{\Delta x} = -\begin{pmatrix} Z(\bar{x})^\top \nabla f(\bar{x}) \\ g(\bar{x}) \end{pmatrix}, \tag{4}$$

where $y(\bar{x})$ is the vector of least squares multipliers obtained by solving

$$\text{minimize} \quad \|\nabla g(\bar{x}) y + \nabla f(\bar{x})\|_2^2\,,$$

with respect to y. See [7].

Two equivalent forms of the necessary conditions gave rise to two different Newton methods (computational issues related to these methods are described, e.g., in the books [6] and [8]). It is natural to ask how do these two methods compare, in other words how close Δx and $\overline{\Delta x}$ are from each other. The goal of this work is not to provide the answer only for this particular reformulation. We propose a model where the answer can be given for the equality–constrained case but also for other cases.

One case in which we are also interested is minimization with simple bounds:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \geq 0 \,. \end{array}$$

The first–order necessary conditions are:

$$\nabla f(x) - y = 0 \,, \tag{5}$$

$$x^\top y = 0 \,, \quad \text{and} \tag{6}$$

$$x, y \geq 0 \,. \tag{7}$$

Since this problem involves inequality constraints, we will rather call conditions (5)–(7), first–order Karush–Kuhn–Tucker (KKT) necessary conditions. A step of Newton's method applied to (5)–(6) is the solution of

$$\begin{pmatrix} \nabla^2 f(x) & -I \\ Y & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -\begin{pmatrix} \nabla f(x) - y \\ XYe \end{pmatrix} , \tag{8}$$

where $X = diag(x)$ and $Y = diag(y)$. (Given a vector u in $\mathbb{R}^n$, $diag(u)$ represents the diagonal matrix of order n whose i–th diagonal element is u_i. Also, e represents a vector of ones. We omit the dimension of e since that will be determined from the context.) The application of Newton's method is made by using an interior–point approach, where x and y are positive and Δx and Δy are scaled by α_x and α_y so that $x + \alpha_x \Delta x$ and $y + \alpha_y \Delta y$ are also positive. The equivalent form of (5)–(7) that we would like to consider (with $x = \bar{x}$) is given by

$$D(\bar{x})^2 \nabla f(\bar{x}) = 0 \quad \text{and} \tag{9}$$

$$\bar{x} \geq 0 \,,$$

where $D(\bar{x})$ is the diagonal matrix of order n with i–th diagonal element given by:

$$\Big(D(\bar{x})\Big)_{ii} = \begin{cases} (\bar{x}_i)^{\frac{1}{2}} & \text{if} \quad (\nabla f(\bar{x}))_i \geq 0 \,, \\ 1 & \text{if} \quad (\nabla f(\bar{x}))_i < 0 \,. \end{cases}$$

This simple fact is proved in [1]. The diagonal element $\Big(D(\bar{x})^2\Big)_{ii}$ might not be differentiable, or even continuous, but if $\Big(D(\bar{x})^2\Big)_{ii}$ is differentiable for all i, then

$$\frac{d}{dx}\left(D(\bar{x})^2 \nabla f(\bar{x})\right) = D(\bar{x})^2 \nabla^2 f(\bar{x}) + \left(\frac{d}{dx}(D(\bar{x})^2 e)\right) \nabla f(\bar{x}) \,.$$

The diagonal element $\left(D(\bar{x})^2\right)_{ii}$ is not differentiable if $(\nabla f(\bar{x}))_i = 0$, in which case the i–th component of the (diagonal) Jacobian matrix $\frac{d}{dx}(D(\bar{x})^2 e)$ is artificially set to zero. So, it makes sense to define $E(\bar{x})$ as the diagonal matrix of order n whose i–th component is given by

$$\left(E(\bar{x})\right)_{ii} = \begin{cases} (\nabla f(\bar{x}))_i & \text{if} \quad (\nabla f(\bar{x}))_i > 0\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

and to define the Newton step corresponding to the equation (9) as

$$\left(D(\bar{x})^2 \nabla^2 f(\bar{x}) + E(\bar{x})\right) \overline{\Delta x} \;=\; -D(\bar{x})^2 \nabla f(\bar{x})\,. \tag{10}$$

See [1] and [2] for more details.

For these and other examples, we are interested in comparing the two alternative Newton approaches by looking at the distance between the two alternative Newton steps Δx and $\overline{\Delta x}$. In the next section, we derive an abstract model to provide a general answer that can be particularized for the different examples. The analysis will consider a *relating condition* $U(x, \bar{x}, y) = 0$, and will establish that the norm of the difference of the steps is bounded by a constant times $\|U(x, \bar{x}, y)\|$. We will observe in the equality-constrained example (Section 3) that $\|U(x, \bar{x}, y)\|$ converges to zero if both Newton sequences are converging to a point satisfying the first-order necessary conditions. As a consequence of our analysis, the alternative Newton steps tend to be the same. However, this is not true in the second type of examples (Sections 4 and 5) because $\|U(x, \bar{x}, y)\|$ does not necessarily converge to zero even if both Newton sequences are converging to the same stationary point.

2 The Abstract Reformulation

The original system of nonlinear equations is defined in the variables x and y:

$$F(x, y) \;=\; 0\,. \tag{11}$$

The equivalent form of this system that we consider is based on the equation

$$G(\bar{x}) H(\bar{x}) \;=\; 0\,, \tag{12}$$

where G satisfies at a solution (x_*, y_*) a *defining condition* of the form

$$G(x_*) F(x_*, y_*) \;=\; G(x_*) H(x_*)\,. \tag{13}$$

The equivalence relates the variables x, $\bar{x}$, and y through the *relating condition*

$$U(x, \bar{x}, y) \;=\; 0\,,$$

so that

$$F(x,y) = 0 \quad \text{is equivalent to} \quad G(x)H(x) = 0 \ \text{ and } \ U(x,x,y) = 0\,.$$

The entrances in F, G, and H are assumed to have Lipschitz first derivatives. Their meaning is clear for the examples we have given before but we postpone this for a while to analyze Newton's method when applied to (11) and (12).

A step of Newton's method when applied to (11) is the solution of

$$\frac{\partial F}{\partial x}(x,y)\Delta x + \frac{\partial F}{\partial y}(x,y)\Delta y \;=\; -F(x,y)\,, \tag{14}$$

whereas a Newton's step for system (12) is given by:

$$\left[\frac{d}{dx}\left(G(\bar{x})H(\bar{x})\right)\right]\overline{\Delta x} \;=\; -G(\bar{x})H(\bar{x})\,. \tag{15}$$

Our goal is to bound $\|\Delta x - \overline{\Delta x}\|$ in terms of $\|U(x,\bar{x},y)\|$. First, we multiply (14) by $G(x)$:

$$G(x)\frac{\partial F}{\partial x}(x,y)\Delta x + G(x)\frac{\partial F}{\partial y}(x,y)\Delta y \;=\; -G(x)F(x,y)\,. \tag{16}$$

By subtracting (15) to (16), we obtain

$$-\left[\frac{d}{dx}\left(G(\bar{x})H(\bar{x})\right)\right]\overline{\Delta x} \;=\; -G(x)\frac{\partial F}{\partial x}(x,y)\Delta x - G(x)\frac{\partial F}{\partial y}(x,y)\Delta y + \Big(G(\bar{x})H(\bar{x}) - G(x)F(x,y)\Big)\,.$$

Finally, we add $\left(\frac{d}{dx}G(\bar{x})H(\bar{x})\right)\Delta x$ to both sides, to get

$$\begin{aligned}\left[\tfrac{d}{dx}\left(G(\bar{x})H(\bar{x})\right)\right](\Delta x - \overline{\Delta x}) \;&=\; \left(\tfrac{d}{dx}\left(G(\bar{x})H(\bar{x})\right) - G(x)\tfrac{\partial F}{\partial x}(x,y)\right)\Delta x - \\ &\quad\; G(x)\tfrac{\partial F}{\partial y}(x,y)\Delta y + \Big(G(\bar{x})H(\bar{x}) - G(x)F(x,y)\Big)\,.\end{aligned}$$

Using the definitions

$$R(\bar{x}) \;=\; \frac{d}{dx}\left(G(\bar{x})H(\bar{x})\right) \quad \text{and} \quad S(x,y) \;=\; G(x)\frac{\partial F}{\partial x}(x,y)\,,$$

we derive an upper bound for $\|\Delta x - \overline{\Delta x}\|$:

$$\begin{aligned}\|\Delta x - \overline{\Delta x}\| \;&\leq\; \|R(\bar{x})^{-1}\|\left(\|R(\bar{x}) - S(x,y)\| + \|G(x)\tfrac{\partial F}{\partial y}(x,y)\|\right)\|\nabla F(x,y)^{-\top}\|\,\|F(x,y)\| \\ &\quad + \|R(\bar{x})^{-1}\|\,\|G(\bar{x})H(\bar{x}) - G(x)F(x,y)\|.\end{aligned}$$

If β_{GH}, β_F, and α_F are positive constants such that

$$\begin{aligned}\|R(\bar{x})^{-1}\| \;&\leq\; \beta_{GH}\,, \\ \|\nabla F(x,y)^{-\top}\| \;&\leq\; \beta_F\,, \quad \text{and} \\ \|F(x,y)\| \;&\leq\; \alpha_F\,,\end{aligned}$$

then

$$\|\Delta x-\overline{\Delta x}\| \ \leq\ \alpha_F\beta_F\beta_{GH}\left(\|R(\bar{x})-S(x,y)\|+\left\|G(x)\frac{\partial F}{\partial y}(x,y)\right\|\right)+\beta_{GH}\|G(\bar{x})H(\bar{x})-G(x)F(x,y)\|\,.$$

We observe from this inequality that $\|\Delta x-\overline{\Delta x}\|$ is bounded above by three important terms.

First, $\|\Delta x-\overline{\Delta x}\|$ depends on how close the values for the functions $G(\bar{x})H(\bar{x})$ and $G(x)F(x,y)$ are from each other. It does not matter how small the residuals $G(\bar{x})H(\bar{x})$ and $F(x,y)$ are, but rather how close the function value $G(\bar{x})H(\bar{x})$ is from the value of $F(x,y)$ reduced by $G(x)$.

The dependence on $R(\bar{x})-S(x,y)$ is about the consistency of the derivatives $\frac{d}{dx}\left(G(\bar{x})H(\bar{x})\right)$ and $G(x)\frac{\partial F}{\partial x}(x,y)$, the former being the derivative of $G(\bar{x})H(\bar{x})$ and the latter the derivative of $F(x,y)$ with respect to x reduced by the operator $G(x)$.

We conclude also that the norm of $\Delta x-\overline{\Delta x}$ depends on the norm of $G(x)\frac{\partial F}{\partial y}(x,y)$ which is a quite interesting aspect of the analysis. In the examples given later, the term $\|G(x)\frac{\partial F}{\partial y}(x,y)\|$ is either zero or bounded by $\|U(x,\bar{x},y)\|$ with $\bar{x}=x$. One can see that $G(x)\frac{\partial F}{\partial y}(x,y)$ is the derivative of $F(x,y)$ with respect to y reduced by the operator $G(x)$, and its norm influences the difference between Δx and $\overline{\Delta x}$.

From the inequality given above we can easily prove the following theorem.

Theorem 2.1: Consider a Newton step (14) for $F(x,y)=0$, where $\nabla F(x,y)$ is nonsingular. Consider a Newton step (15) for $G(\bar{x})H(\bar{x})=0$, where $R(\bar{x})$ is nonsingular.

If there exist positive constants γ_1, γ_2, and γ_3 such that

$$\begin{aligned}
\|R(\bar{x})-S(x,y)\| &\ \leq\ \gamma_1\|U(x,\bar{x},y)\|\,,\\
\left\|G(x)\frac{\partial F}{\partial y}(x,y)\right\| &\ \leq\ \gamma_2\|U(x,\bar{x},y)\| \ \text{ (for some } \bar{x}), \ \text{ and}\\
\|G(\bar{x})H(\bar{x})-G(x)F(x,y)\| &\ \leq\ \gamma_3\|U(x,\bar{x},y)\|\,,
\end{aligned}$$

then

$$\left\|\Delta x-\overline{\Delta x}\right\| \ \leq\ \gamma\ \|U(x,\bar{x},y)\|$$

for some positive constant γ.

The constant γ in Theorem 2.1 depends on x, $\bar{x}$, and y since γ_1, γ_2, and γ_3 depend also on these points. We can assume that the standard Newton assumptions [4] hold for the Newton methods defined by (14) and (15) at the points (x_*,y_*) and x_*, respectively. Then, if x and y are sufficiently closed to x_* and y_*, and $\bar{x}$ is sufficiently close to x_*, the constants γ_1, γ_2, γ_3, and γ do not depend on the points x, $\bar{x}$, and y.

We move rapidly to the examples to illustrate our analysis.

3 Equality–constrained Optimization

In this case

$$F(x,y) \;=\; \begin{pmatrix} \nabla f(x) + \nabla g(x) y \\ g(x) \end{pmatrix}$$

and

$$G(\bar{x})H(\bar{x}) \;=\; \begin{pmatrix} Z(\bar{x})^\top \nabla f(\bar{x}) \\ g(\bar{x}) \end{pmatrix}.$$

The choices for $G(\bar{x})$ and $H(\bar{x})$ are

$$G(\bar{x}) \;=\; \begin{pmatrix} Z(\bar{x})^\top & 0 \\ 0 & I \end{pmatrix} \quad \text{and} \quad H(\bar{x}) \;=\; \begin{pmatrix} \nabla f(\bar{x}) \\ g(\bar{x}) \end{pmatrix}.$$

Since

$$G(\bar{x})H(\bar{x}) - G(x)F(x,y) \;=\; \begin{pmatrix} Z(\bar{x})^\top \nabla f(\bar{x}) - Z(x)^\top \nabla f(x) \\ g(\bar{x}) - g(x) \end{pmatrix},$$

the defining condition (13) is satisfied for any pair (x,y), even if it does not verify the first-order necessary conditions.

From the theory presented in [7],

$$R(\bar{x}) \;=\; \frac{d}{dx}\left(G(\bar{x})H(\bar{x})\right) \;=\; \begin{pmatrix} Z(\bar{x})^\top \nabla^2_{xx}\ell(\bar{x}, y(\bar{x})) \\ \nabla g(\bar{x})^\top \end{pmatrix}.$$

Also,

$$S(x,y) \;=\; \begin{pmatrix} Z(x)^\top \nabla^2_{xx}\ell(x,y) \\ \nabla g(x)^\top \end{pmatrix},$$

and the Lipschitz continuity of the second derivatives of $\ell(x,y)$ imply

$$\|R(\bar{x}) - S(x,y)\| \;\le\; \gamma_1 \left\| \begin{pmatrix} x - \bar{x} \\ y - y(\bar{x}) \end{pmatrix} \right\|,$$

for some positive constant γ_1.

Moreover,

$$G(x)\frac{\partial F}{\partial y}(x,y) \;=\; 0\,.$$

It is natural to define the relating condition as

$$U(x,\bar{x},y) \;=\; \left(\begin{array}{c} x-\bar{x} \\ y-y(\bar{x}) \end{array}\right) \;=\; 0\,.$$

Theorem 2.1 assures the existence of a positive constant γ such that

$$\|\Delta x - \overline{\Delta x}\| \;\leq\; \gamma \left\| \left(\begin{array}{c} x-\bar{x} \\ y-y(\bar{x}) \end{array}\right) \right\|\,,$$

where Δx and $\overline{\Delta x}$ are given by (2) and (4). It is obvious that $F(x,y)=0$ is equivalent to $G(x)H(x)=0$ and $U(x,x,y)=0$.

Also, if (x_*,y_*) is a stationary point then

$$\lim_{\substack{(x,y)\longrightarrow(x_*,y_*) \\ \bar{x}\longrightarrow x_*}} \|U(x,\bar{x},y)\| \;=\; 0\,, \tag{17}$$

and the Newton steps Δx and $\overline{\Delta x}$ tend in this situation to be the same step.

4 Minimization with Simple Bounds

In this case, the equivalent KKT systems are:

$$F(x,y) \;=\; \left(\begin{array}{c} \nabla f(x)-y \\ XYe \end{array}\right) \;=\; 0$$

and

$$G(\bar{x})H(\bar{x}) \;=\; D(\bar{x})^2\nabla f(\bar{x}) \;=\; 0\,.$$

Of course, we have excluded from the KKT systems the nonnegativity of x, $\bar{x}$, and y. The choices for $G(\bar{x})$ and $H(\bar{x})$ are

$$G(\bar{x}) \;=\; \left(\begin{array}{cc} D(\bar{x})^2 & I \end{array}\right) \quad\text{and}\quad H(\bar{x}) \;=\; \left(\begin{array}{c} \nabla f(\bar{x}) \\ 0 \end{array}\right).$$

In this case the defining condition (13) is not satisfied unless we are at a point that verifies the first–order KKT necessary conditions. In fact, we have

$$G(\bar{x})H(\bar{x}) - G(x)F(x,y) \;=\; D(\bar{x})^2\nabla f(\bar{x}) - D(x)^2\nabla f(x) \;-\; \left(X-D(x)^2\right)Ye\,.$$

$R(\bar{x})$ and $S(x,y)$ are given by:

$$R(\bar{x}) \;=\; D(\bar{x})^2\nabla^2 f(\bar{x}) + E(\bar{x}) \quad \text{and} \quad S(x,y) \;=\; D(x)^2\nabla^2 f(x) + Y\,.$$

Moreover,

$$G(x)\frac{\partial F}{\partial y}(x,y) \;=\; -D(x)^2 + X\,.$$

Thus, if the relating condition is defined as

$$U(x,\bar{x},y) \;=\; \begin{pmatrix} x-\bar{x} \\ \left(X - D(x)^2\right)e \\ \left(\bar{X} - D(\bar{x})^2\right)e \\ \left(Y - E(x)\right)e \end{pmatrix} \;=\; 0\,,$$

then there exist positive constants γ_1, γ_2, and γ_3 satisfying the assumptions of Theorem 2.1. This theorem assures the existence of a positive constant γ such that

$$\|\Delta x - \overline{\Delta x}\| \;\leq\; \gamma \left\| \begin{pmatrix} x-\bar{x} \\ \left(X - D(x)^2\right)e \\ \left(\bar{X} - D(\bar{x})^2\right)e \\ \left(Y - E(x)\right)e \end{pmatrix} \right\|\,,$$

where Δx and $\overline{\Delta x}$ are given by (8) and (10). Note that $F(x,y)=0$ is equivalent to $G(x)H(x)=0$ and $U(x,x,y)=0$ provided $\nabla f(x)\geq 0$. However, in this example, if (x_*,y_*) is a point that satisfies the first-order KKT necessary conditions, the limit (17) might not hold because

$$\lim_{x\longrightarrow x_*}\left\|\left(X - D(x)^2\right)e\right\| \qquad \text{and} \qquad \lim_{\bar{x}\longrightarrow x_*}\left\|\left(\bar{X} - D(\bar{x})^2\right)e\right\|$$

do not necessarily exist or equal zero.

5 Discretized Optimal Control Problems with Bounds on the Control Variables

In this section, we consider the class of nonlinear programming problems analyzed in [3]. See also [5]. A nonlinear programming problem of this class is formulated as

$$\begin{array}{ll} \text{minimize} & f(x_1,x_2) \\ \text{subject to} & g(x_1,x_2)=0 \\ & x_2 \geq 0\,, \end{array}$$

where x_1 is in $\mathbb{R}^m$ and x_2 is in $\mathbb{R}^{n-m}$. In this class of problems, $\nabla g(x)$ is partitioned as

$$\nabla g(x) \;=\; \begin{pmatrix} \nabla_{x_1} g(x) \\ \nabla_{x_2} g(x) \end{pmatrix}\,,$$

where $\nabla_{x_1} g(x)$ is nonsingular. The first–order KKT necessary conditions are:

$$\begin{aligned}
&\nabla f(x) + \nabla g(x) y_1 - \begin{pmatrix} 0 \\ y_2 \end{pmatrix} = 0\,, \\
&g(x) = 0\,, \\
&x_2^\top y_2 = 0\,, \quad \text{and} \\
&x_2, y_2 \geq 0\,.
\end{aligned}$$

A basis for the null space of $\nabla g(\bar{x})^\top$ is given by the columns of

$$W(\bar{x}) \;=\; \begin{pmatrix} -\nabla_{x_1} g(\bar{x})^{-\top} \nabla_{x_2} g(\bar{x})^\top \\ I \end{pmatrix}.$$

The equivalent KKT system that we consider is:

$$D(\bar{x})^2 W(\bar{x})^\top \nabla f(\bar{x}) = 0\,, \tag{18}$$

$$g(\bar{x}) = 0\,, \quad \text{and} \tag{19}$$

$$\bar{x}_2 \geq 0\,,$$

where $D(\bar{x})$ is the diagonal matrix of order $n-m$ with i–th diagonal element given by:

$$\Big(D(\bar{x})\Big)_{ii} = \begin{cases} (\bar{x}_2)_i^{\frac{1}{2}} & \text{if} \quad \big(W(\bar{x})^\top \nabla f(\bar{x})\big)_i \geq 0\,, \\ 1 & \text{if} \quad \big(W(\bar{x})^\top \nabla f(\bar{x})\big)_i < 0\,. \end{cases}$$

The Newton step associated with (18)–(19) is the solution of

$$\big(D(\bar{x})^2 W(\bar{x})^\top \nabla^2_{xx} \ell(\bar{x}, y_1(\bar{x})) + E(\bar{x})\big)\, \overline{\Delta x} \;=\; -D(\bar{x})^2 W(\bar{x})^\top \nabla f(\bar{x})\,,$$

where $y_1(\bar{x}) = -\nabla_{x_1} g(\bar{x})^{-1} \nabla_{x_1} f(\bar{x})$ and $E(\bar{x})$ is a diagonal matrix of order $n-m$ with i–th diagonal element given by:

$$\Big(E(\bar{x})\Big)_{ii} = \begin{cases} \big(W(\bar{x})^\top \nabla f(\bar{x})\big)_i & \text{if} \quad \big(W(\bar{x})^\top \nabla f(\bar{x})\big)_i > 0\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

See [3] for more details.

In this case, the KKT systems are represented by:

$$F(x,y) \;=\; \begin{pmatrix} \nabla f(x) + \nabla g(x) y_1 - \begin{pmatrix} 0 \\ y_2 \end{pmatrix} \\ g(x) \\ X_2 Y_2 e \end{pmatrix}$$

and

$$G(\bar{x})H(\bar{x}) \;=\; \begin{pmatrix} D(\bar{x})^2 W(\bar{x})^\top \nabla f(\bar{x}) \\ g(\bar{x}) \end{pmatrix}.$$

(We have excluded the nonnegativity of x_2, $\bar{x}_2$, and y_2.) The choices for $G(\bar{x})$ and $H(\bar{x})$ are

$$G(\bar{x}) \;=\; \begin{pmatrix} D(\bar{x})^2 W(\bar{x})^\top & 0 & I \\ 0 & I & 0 \end{pmatrix} \quad \text{and} \quad H(\bar{x}) \;=\; \begin{pmatrix} \nabla f(\bar{x}) \\ g(\bar{x}) \\ 0 \end{pmatrix}.$$

The defining condition (13) is not satisfied unless we are at a point that verifies the first–order KKT necessary conditions:

$$G(\bar{x})H(\bar{x}) - G(x)F(x,y) \;=\; \begin{pmatrix} D(\bar{x})^2 W(\bar{x})^\top \nabla f(\bar{x}) - D(x)^2 W(x)^\top \nabla f(x) - \left(X_2 - D(x)^2\right) Y_2 e \\ g(\bar{x}) - g(x) \end{pmatrix}.$$

Also, $R(\bar{x})$ and $S(x,y)$ are given by:

$$R(\bar{x}) \;=\; \begin{pmatrix} D(\bar{x})^2 W(\bar{x})^\top \nabla^2_{xx} \ell(\bar{x}, y_1(\bar{x})) + E(\bar{x}) \\ \nabla g(\bar{x})^\top \end{pmatrix} \quad \text{(see [3]) and}$$

$$S(x,y) \;=\; \begin{pmatrix} D(x)^2 W(x)^\top \nabla^2_{xx} \ell(x, y_1) + Y_2 \\ \nabla g(x)^\top \end{pmatrix}.$$

Moreover,

$$G(x)\frac{\partial F}{\partial y}(x,y) \;=\; \begin{pmatrix} 0 & -D(x)^2 + X_2 \\ 0 & 0 \end{pmatrix}.$$

So, the relating condition is defined as

$$U(x,\bar{x},y) \;=\; \begin{pmatrix} x - \bar{x} \\ y_1 - y_1(\bar{x}) \\ \left(X_2 - D(x)^2\right) e \\ \left(\bar{X}_2 - D(\bar{x})^2\right) e \\ \left(Y_2 - E(x)\right) e \end{pmatrix} \;=\; 0\,,$$

assuring the existence of the positive constants γ_1, γ_2, and γ_3 in Theorem 2.1, which in turn guarantees the existence of a positive constant γ such that

$$\|\Delta x - \overline{\Delta x}\| \;\le\; \gamma \left\| \begin{pmatrix} x - \bar{x} \\ y_1 - y_1(\bar{x}) \\ \left(X_2 - D(x)^2\right) e \\ \left(\bar{X}_2 - D(\bar{x})^2\right) e \\ \left(Y_2 - E(x)\right) e \end{pmatrix} \right\|.$$

Note that $F(x,y) = 0$ is equivalent to $G(x)H(x) = 0$ and $U(x,x,y) = 0$ provided $W(x)^\top \nabla f(x) \geq 0$. In this example, as in the previous one, even if (x_*, y_*) satisfies the first-order KKT conditions, there is no guarantee that the limit (17) is true.

A similar analysis for the nonlinear programming problem is also of interest. For instance, we could consider the primal–dual, the affine–scaling, and the reduced primal–dual algorithms described in [9].

References

[1] 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.

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

[3] 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.

[4] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice–Hall, Englewood Cliffs, New Jersey, 1983. Republished by SIAM, Philadelphia, 1996.

[5] J. E. Dennis and L. N. Vicente, *On the convergence theory of general trust–region–based algorithms for equality–constrained optimization*, SIAM J. Optim., 7 (1997), pp. 927–950.

[6] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Chichester, second ed., 1987.

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

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

[9] L. N. Vicente, *On interior-point Newton algorithms for discretized optimal control problems with state constraints*, Optimization Methods & Software, 8 (1998), pp. 249–275.