Sequential quadratic programming for the general nonlinear program

Mark S. Gockenbach

1 Introduction

I will now discuss sequential quadratic programming algorithms for the general NLP

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & g(x) = 0 \\ & h(x) \geq 0. \end{aligned}$$

Assuming x^* is a local minimizer and λ^*, θ^* are the corresponding Lagrange multipliers for the equality and inequality constraints, respectively, an SQP method will produce sequences $\{x^{(k)}\}$, $\{\lambda^{(k)}\}$, and $\{\theta^{(k)}\}$ by solving a sequence of quadratic programs. Based on my discussion of equality-constrained NLPs, it would seem natural to base the iteration on either of the following QPs:

$$\begin{aligned} \min \quad & \frac{1}{2} p \cdot \nabla^2 \ell(x^{(k)}; \lambda^{(k)}; \theta^{(k)}) p + \nabla \ell(x^{(k)}; \lambda^{(k)}; \theta^{(k)}) \cdot p \\ s.t. \quad & \nabla g(x^{(k)})^T p + g(x^{(k)}) = 0 \\ & \nabla h(x^{(k)})^T p + h(x^{(k)}) \geq 0, \end{aligned} \tag{1}$$

$$\begin{aligned} \min \quad & \frac{1}{2} p \cdot \nabla^2 \ell(x^{(k)}; \lambda^{(k)}; \theta^{(k)}) p + \nabla f(x^{(k)}) \cdot p \\ s.t. \quad & \nabla g(x^{(k)})^T p + g(x^{(k)}) = 0 \\ & \nabla h(x^{(k)})^T p + h(x^{(k)}) \geq 0. \end{aligned} \tag{2}$$

However, QP (1) turns out not to be suitable, because x^* need not be a local minimizer of

$$\begin{aligned} \min \quad & \ell(x; \lambda^*; \theta^*) \\ s.t. \quad & g(x) = 0 \\ & h(x) \geq 0. \end{aligned}$$

The following calculation shows that $\ell(x^*; \lambda^*; \theta^*) = f(x^*)$:

$$\begin{aligned} \ell(x^*; \lambda^*; \theta^*) &= f(x^*) - \lambda^* \cdot g(x^*) - \theta^* \cdot h(x^*) \\ &= f(x^*) - \sum_{i=1}^{p} \theta_i^* h_i(x^*) \quad (\text{since } g(x^*) = 0) \\ &= f(x^*) \quad (\text{since, for each } i, \text{ either } \theta_i^* = 0 \text{ or } h_i(x^*) = 0). \end{aligned}$$

However, if x is feasible, then

$$\ell(x; \lambda^*; \theta^*) = f(x) - \lambda^* \cdot g(x) - \theta^* \cdot h(x)$$

$$
\begin{aligned}
&= f(x) - \sum_{i=1}^{p} \theta_i^* h_i(x) \quad (\text{since } g(x) = 0) \\
&= f(x) - \sum_{i \in \mathcal{A}(x^*)} \theta_i^* h_i(x) \quad (\text{since } \theta_i^* = 0 \text{ for } i \notin \mathcal{A}(x^*)) \\
&\leq f(x) \quad (\text{since } \theta_i^* \geq 0 \text{ and } h_i(x) \geq 0 \text{ for all } i),
\end{aligned}
$$

and this inequality is strict for strictly feasible x as long as there are inactive constraints and strict complementarity holds. Therefore, there is no reason to believe that x^* is a local minimizer of $\ell(\cdot;\lambda^*;\theta^*)$ subject to the constraints $g(x) = 0$ and $h(x) \geq 0$.

Example 1.1 *Let $f : \mathrm{R}^2 \to \mathrm{R}$ and $h : \mathrm{R}^2 \to \mathrm{R}$ be defined by*

$$
\begin{aligned}
f(x) &= x_1^2 - x_2^2 + 4x_2, \\
h(x) &= x_2 - 1.
\end{aligned}
$$

Then the NLP

$$
\begin{aligned}
\min \quad & f(x) \\
s.t. \quad & h(x) \geq 0
\end{aligned}
$$

has a unique local minimizer, $x^* = (0,1)$, with corresponding Lagrange multiplier $\theta^* = 2$ (x^* is not a global minimizer since f is unbounded below in the feasible region). It is easy to verify that x^* is a nonsingular point of the NLP. The Lagrangian, using the exact Lagrange multiplier, is

$$
\ell(x;\theta^*) = x_1^2 - x_2^2 + 4x_2 - 2(x_2 - 1) = x_1^2 - (x_2 - 1)^2 + 3,
$$

and this formula shows that x^* is not a local minimizer of

$$
\begin{aligned}
\min \quad & \ell(x;\theta^*) \\
s.t. \quad & h(x) \geq 0.
\end{aligned}
$$

The following theorem shows that formulation (2) of the QP subproblem is suitable for an SQP algorithm.

Theorem 1.2 *Suppose $f : \mathrm{R}^n \to \mathrm{R}$, $g : \mathrm{R}^n \to \mathrm{R}^m$, and $h : \mathrm{R}^n \to \mathrm{R}$ are twice continuously differentiable, and that $x^* \in \mathrm{R}^n$ is a local minimizer and nonsingular point of the NLP*

$$
\begin{aligned}
\min \quad & f(x) \\
s.t. \quad & g(x) = 0 \\
& h(x) \geq 0.
\end{aligned}
$$

Let λ^* and θ^* be the corresponding Lagrange multipliers for the equality and inequality constraints, respectively. Then x^* is also a local minimizer of

$$
\begin{aligned}
\min \quad & \frac{1}{2}(x - x^*) \cdot \nabla^2 \ell(x^*;\lambda^*;\theta^*)(x - x^*) + \nabla f(x^*) \cdot (x - x^*) \\
s.t. \quad & \nabla g(x^*)^T (x - x^*) + g(x^*) = 0 \\
& \nabla h(x^*)^T (x - x^*) + h(x^*) \geq 0.
\end{aligned}
\tag{3}
$$

Proof: I will show that x^* together with λ^*, θ^* satisfy the sufficient conditions for a local minimizer. For the sake of notation, I define

$$\begin{aligned} q(x) &= \frac{1}{2}(x-x^*)\cdot\nabla^2\ell(x^*;\lambda^*;\theta^*)(x-x^*)+\nabla f(x^*)\cdot(x-x^*), \\ c(x) &= \nabla g(x^*)^T(x-x^*)+g(x^*), \\ d(x) &= \nabla h(x^*)^T(x-x^*)+h(x^*). \end{aligned}$$

Then

$$\begin{aligned} \nabla q(x^*) &= \nabla f(x^*), \\ \nabla c(x^*) &= \nabla g(x^*), \\ \nabla d(x^*) &= \nabla h(x^*), \end{aligned}$$

and the first-order optimality conditions are satisfied by x^* together with λ^* and θ^*:

$$\begin{aligned} \nabla q(x^*)-\nabla c(x^*)\lambda^*-\nabla d(x^*)\theta^* &= \nabla f(x^*)-\nabla g(x^*)\lambda^*-\nabla h(x^*)\theta^*=0, \\ c(x^*) &= g(x^*)=0, \\ d(x^*) &= h(x^*)\geq 0, \\ \theta^* &\geq 0, \\ \theta^* d(x^*) &= \theta^* h(x^*)=0. \end{aligned}$$

Moreover, if $z\in \mathrm{R}^n$ satisfies $z\neq 0$ and

$$\begin{aligned} c(x^*)^T z &= 0, \\ d_i(x^*)\cdot z &= 0 \text{ for all } i\in\mathcal{A}(x^*) \text{ such that } \lambda_i^*>0, \\ d_i(x^*)\cdot z &\geq 0 \text{ for all } i\in\mathcal{A}(x^*) \text{ such that } \lambda_i^*=0, \end{aligned}$$

then z also satisfies

$$\begin{aligned} g(x^*)^T z &= 0, \\ h_i(x^*)\cdot z &= 0 \text{ for all } i\in\mathcal{A}(x^*) \text{ such that } \lambda_i^*>0, \\ h_i(x^*)\cdot z &\geq 0 \text{ for all } i\in\mathcal{A}(x^*) \text{ such that } \lambda_i^*=0, \end{aligned}$$

and hence, since x^* is a nonsingular point for the original NLP,

$$z\cdot\ell(x^*;\lambda^*;\theta^*)x>0.$$

But, since the QP (3) has linear constraints, the Hessian of the Lagrangian of (3) is simply

$$\nabla^2 q(x^*)=\nabla^2\ell(x^*;\lambda^*;\theta^*).$$

This shows that x^* satisfies the sufficient conditions for a local minimizer of (3). QED

It follows from the previous theorem that, for $x^{(k)}$, $\lambda^{(k)}$, $\theta^{(k)}$ sufficiently close to x^*, λ^*, θ^*, respectively, the QP (2) has a local minimizer $p^{(k)}$. Moreover, this solution eventually identifies the correct active set of constraints, as the following theorem shows:

Theorem 1.3 *Suppose $f:\mathrm{R}^n\to\mathrm{R}$, $g:\mathrm{R}^n\to\mathrm{R}^m$, and $h:\mathrm{R}^n\to\mathrm{R}$ are all twice continuously differentiable and x^* is a local minimizer and nonsingular point of the NLP*

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & g(x)=0 \\ & h(x)\geq 0. \end{aligned}$$

Suppose further that λ^* and θ^* are the corresponding Lagrange multipliers. Then there exist neighborhoods N_x of x^*, N_λ of λ^*, and N_θ of θ^* and functions

$$\begin{aligned} p &: N_x \times N_\lambda \times N_\theta \to \mathrm{R}^n, \\ \lambda^+ &: N_x \times N_\lambda \times N_\theta \to \mathrm{R}^m, \\ \theta^+ &: N_x \times N_\lambda \times N_\theta \to \mathrm{R}^p \end{aligned}$$

such that, for all $x \in N_x$, $\lambda \in N_\lambda$, $\theta \in N_\theta$, $p(x, \lambda, \theta)$ is the (locally unique) solution of the QP

$$\begin{aligned} \min \quad & \frac{1}{2} p \cdot \nabla^2 \ell(x; \lambda; \theta) p + \nabla f(x) \cdot p \\ s.t. \quad & \nabla g(x)^T p + g(x) = 0 \\ & \nabla h(x)^T p + h(x) \geq 0 \end{aligned} \tag{4}$$

and $\lambda^+(x, \lambda, \theta)$, $\theta^+(x, \lambda, \theta)$ are the corresponding Lagrange multipliers. Moreover, p, λ^+, and θ^+ are continuously differentiable and

$$\begin{aligned} p(x^*, \lambda^*, \theta^*) &= 0, \\ \lambda^+(x^*, \lambda^*, \theta^*) &= \lambda^*, \\ \theta^+(x^*, \lambda^*, \theta^*) &= \theta^*. \end{aligned}$$

Finally, the neighborhoods can be chosen sufficiently small that, for all $x \in N_x$, $\lambda \in N_\lambda$, $\theta \in N_\theta$, the active set of (linearized) constraints determined by $p(x, \lambda, \theta)$ is the same as the active set determined by x^* for the nonlinear program. In other words,

$$\{i \,:\, \nabla h_i(x) \cdot p(x, \lambda, \theta) + h_i(x) = 0\} = \mathcal{A}(x^*).$$

Proof: The proof is by the implicit function theorem. Consider the QP

$$\begin{aligned} \min \quad & \frac{1}{2} p \cdot \nabla^2 \ell(x^*; \lambda^*; \theta^*) p + \nabla f(x^*) \cdot p \\ s.t. \quad & \nabla g(x^*)^T p + g(x^*) = 0 \\ & \nabla h(x^*)^T p + h(x^*) \geq 0, \end{aligned} \tag{5}$$

which, by the previous theorem, has solution $p^* = 0$ with Lagrange multipliers λ^*, θ^*. The vectors p^*, λ^*, θ^* satisfy the first-order necessary conditions

$$\begin{aligned} \nabla^2 \ell(x^*; \lambda^*; \theta^*) p^* + \nabla f(x^*) - \nabla g(x^*) \lambda^* - \nabla h(x^*) \theta^* &= 0, \\ \nabla g(x^*)^T p^* + g(x^*) &= 0, \\ \theta^* \left(\nabla h(x^*)^T p^* + h(x^*) \right) &= 0, \\ \nabla h(x^*)^T p^* + h(x^*) &\geq 0, \\ \theta^* &\geq 0. \end{aligned}$$

Defining $F : \mathrm{R}^{n+m+p} \times \mathrm{R}^{n+m+p} \to \mathrm{R}^{n+m+p}$ by

$$F(p, \lambda^+, \theta^+; x, \lambda, \theta) = \begin{bmatrix} \nabla^2 \ell(x; \lambda; \theta) p + \nabla f(x) - \nabla g(x) \lambda^+ - \nabla h(x) \theta^+ \\ \nabla g(x)^T p + g(x) \\ \theta^+ \left(\nabla h(x)^T p + h(x) \right) \end{bmatrix},$$

it follows that

$$F(p^*, \lambda^*, \theta^*; x^*, \lambda^*, \theta^*) = 0.$$

The Jacobian of F (with respect to the variables (p, λ^+, θ^+)) is

$$J(p, \lambda^+, \theta^+; x, \lambda, \theta) = \begin{bmatrix} \nabla^2 \ell(x; \lambda; \theta) & -\nabla g(x) & -\nabla h(x) \\ \nabla g(x)^T & 0 & 0 \\ \Theta^+ \nabla h(x)^T & 0 & \operatorname{diag}\big(\nabla h(x)^T p + h(x)\big) \end{bmatrix}.$$

In particular,

$$J^* = J(p^*, \lambda^*, \theta^*; x^*, \lambda^*, \theta^*) = \begin{bmatrix} \nabla^2 \ell(x^*; \lambda^*; \theta^*) & -\nabla g(x^*) & -\nabla h(x^*) \\ \nabla g(x^*)^T & 0 & 0 \\ \Theta^* \nabla h(x^*)^T & 0 & H(x^*) \end{bmatrix}.$$

(I have used the convention that if a lowercase letter, for example, v, represents a vector, then the corresponding uppercase letter, for example, V, represents the diagonal matrix whose diagonal entries are the components of the vector.) As discussed earlier, the matrix J^* is nonsingular, and therefore the implicit function theorem guarantees the existence of neighborhoods N_x of x^*, N_λ of λ^*, and N_θ of θ^* and functions

$$\begin{aligned} p &: N_x \times N_\lambda \times N_\theta \to \mathrm{R}^n, \\ \lambda^+ &: N_x \times N_\lambda \times N_\theta \to \mathrm{R}^m, \\ \theta^+ &: N_x \times N_\lambda \times N_\theta \to \mathrm{R}^p \end{aligned}$$

such that, for all $x \in N_x$, $\lambda \in N_\lambda$, $\theta \in N_\theta$, $p = p(x, \lambda, \theta)$, $\lambda^+ = \lambda^+(x, \lambda, \theta)$, $\theta^+ = \theta^+(x, \lambda, \theta)$ form the unique solution of

$$F(p, \lambda^+, \theta^+; x, \lambda, \theta) = 0. \tag{6}$$

Moreover, p, λ^+, θ^+ are continuously differentiable and

$$\begin{aligned} p(x^*, \lambda^*, \theta^*) &= 0, \\ \lambda^+(x^*, \lambda^*, \theta^*) &= \lambda^*, \\ \theta^+(x^*, \lambda^*, \theta^*) &= \theta^*. \end{aligned}$$

Next I will show that $p(x, \lambda, \theta)$ is the solution of QP (4) and $\lambda^+(x, \lambda, \theta)$, $\theta^+(x, \lambda, \theta)$ are the corresponding Lagrange multipliers. The equation (6) shows that three of the first-order necessary conditions (the Lagrange multiplier equation, feasiblity with respect to the equality constraint, and complementarity) hold for $p(x, \lambda, \theta)$ and the multipliers $\lambda^+(x, \lambda, \theta)$, $\theta^+(x, \lambda, \theta)$. Since x^* is a nonsingular point, strict complementarity holds for x^* and θ^*. It then follows, from the continuity of p and θ^+, that if (x, λ, θ) is sufficiently close to $(x^*, \lambda^*, \theta^*)$, then

$$i \in \mathcal{A}(x^*) \;\Rightarrow\; \theta^+(x, \lambda, \theta) > 0. \tag{7}$$

Moreover, since $p(x, \lambda, \theta)$ is arbitrarily small and $h_i(x)$ is positive for $i \notin \mathcal{A}(x^*)$ and (x, λ, θ) sufficiently close to $(x^*, \lambda^*, \theta^*)$, it follows that

$$i \notin \mathcal{A}(x^*) \;\Rightarrow\; \nabla h_i(x) \cdot p + h_i(x) > 0 \tag{8}$$

I will now assume that N_x, N_λ, and N_θ are chosen small enough that (7) and (8) hold for all $x \in N_x$, $\lambda \in N_\lambda$, and $\theta \in N_\theta$. Since (6) implies

$$\theta^+(x, \lambda, \theta)\,(\nabla h_i(x) \cdot p + h_i(x)) = 0, \; i = 1, 2, \ldots, p,$$

(7) implies that

$$\nabla h_i(x) \cdot p(x, \lambda, \theta) + h_i(x) = 0 \text{ for all } i \in \mathcal{A}(x^*),$$

while (8) implies that

$$\theta_i^+(x, \lambda, \theta) = 0 \text{ for all } i \notin \mathcal{A}(x^*).$$

Therefore, for all $i = 1, 2, \ldots, p$,

$$\theta_i^+(x, \lambda, \theta) \geq 0 \text{ and } \nabla h_i(x) \cdot p(x, \lambda, \theta) + h_i(x) \geq 0.$$

I have now shown that $p(x, \lambda, \theta)$, $\lambda^+(x, \lambda, \theta)$, and $\theta^+(x, \lambda, \theta)$ satisfy the first-order necessary conditions for the QP (4), and also that $p(x, \lambda, \theta)$ defines the same active set (relative to the QP) as does x^* (relative to the original nonlinear program).

Since (4) is not necessarily a convex program (although the feasible region is a convex set, there is no guarantee that the quadratic objective function is a convex function), it is necessary to show that the second-order sufficiency condition holds. However, this follows by continuity, since the objective function of the limit problem (5) has the necessary positive curvature. Therefore, $p(x, \lambda, \theta)$ is a solution of (4). QED

Here is the local SQP algorithm for solving the general NLP

$$\begin{array}{ll} \min & f(x) \\ s.t. & g(x) = 0 \\ & h(x) \geq 0. \end{array}$$

Choose $x^{(0)} \in \mathrm{R}^n$, $\lambda^{(0)} \in \mathrm{R}^m$, and $\theta^{(0)} \in \mathrm{R}^p$, $\theta^{(0)} \geq 0$.

For $k = 1, 2, 3, \ldots$

Solve

$$\begin{array}{ll} \min & \frac{1}{2} p \cdot \nabla^2 \ell(x^{(k)}; \lambda^{(k)}; \theta^{(k)}) p + \nabla f(x^{(k)}) \cdot p \\ s.t. & \nabla g(x^{(k)})^T p + g(x^{(k)}) = 0 \\ & \nabla h(x^{(k)})^T p + h(x^{(k)}) \geq 0 \end{array}$$

to get the solution $p^{(k)}$ and Lagrange multipliers $\lambda^{(k+1)}$, $\theta^{(k+1)}$.

Set $x^{(k+1)} = x^{(k)} + p^{(k)}$.

Example 1.4 *Consider the NLP*

$$\begin{array}{ll} \min & f(x) \\ s.t. & g(x) = 0 \\ & h(x) \geq 0, \end{array}$$

where

$$\begin{aligned} f(x) &= x_1^2 + 2(x_2 - 2)^2, \\ g(x) &= x_1^2 - x_2, \\ h(x) &= 1 - x_1^2 - x_2^2. \end{aligned}$$

The feasible set is the part of the parabola $x_2 = x_1^2$ lying inside the circle $x_1^2 + x_2^2 = 1$. I apply the SQP algorithm described above, using $x^{(0)} = (-3, 2)$, $\lambda^{(0)} = 0$, $\theta^{(0)} = 0$. The results of the first four iterations are illustrated in Figure 1, while the first size iterations are recorded in Table 1. The exact solution is

$$x^* = \begin{bmatrix} -\left(\frac{-1+\sqrt{5}}{2}\right)^2 \\ \frac{-1+\sqrt{5}}{2} \end{bmatrix}.$$

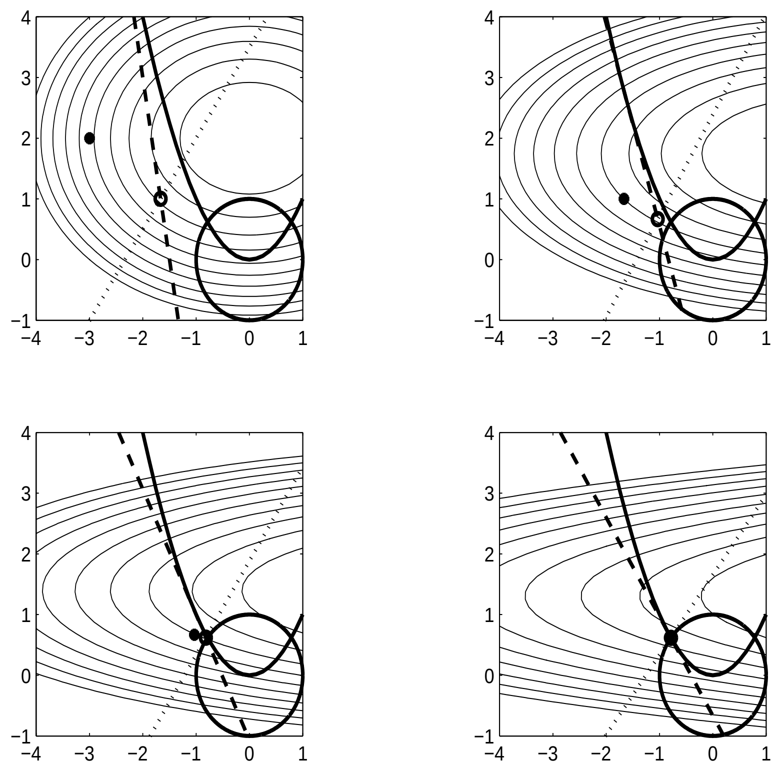

Figure 1: Four iterations of the local SQP method applied to the NLP from Example 1.4. The feasible set is the part of the parabola inside the circle. The linearized equality constraint is the dashed line, while the linear inequality constraint is indicated by the dotted line. The asterisk indicates $x^{(k)}$ and the small circle $x^{(k+1)}$, $k = 0$ (upper left), $k = 1$ (upper right), $k = 2$ (lower left), and $k = 3$ (lower right).

It remains only to analyze the rate of convergence of $x^{(k)} \to x^*$, $\lambda^{(k)} \to \lambda^*$, and $\theta^{(k)} \to \theta^*$. The vectors $p^{(k)}$, $\lambda^{(k+1)}$, and $\theta^{(k+1)}$ satisfy the optimality conditions for the QP

$$\begin{array}{ll} \min & \frac{1}{2} p \cdot \nabla^2 \ell(x^{(k)}; \lambda^{(k)}; \theta^{(k)}) p + \nabla f(x^{(k)}) \cdot p \\ s.t. & \nabla g(x^{(k)})^T p + g(x^{(k)}) = 0 \\ & \nabla h(x^{(k)})^T p + h(x^{(k)}) \geq 0, \end{array}$$

which include the equations

$$\begin{aligned} \nabla^2 \ell(x^{(k)}; \lambda^{(k)}; \theta^{(k)}) p^{(k)} + \nabla f(x^{(k)}) - \nabla g(x^{(k)}) \lambda^{(k+1)} - \nabla h(x^{(k)}) \theta^{(k+1)} &= 0, \\ \nabla g(x^{(k)})^T p^{(k)} + g(x^{(k)}) &= 0, \\ \theta^{(k+1)} \left(\nabla h(x^{(k)})^T p^{(k)} + h(x^{(k)}) \right) &= 0. \end{aligned} \tag{9}$$

However, since I have already proved that, for $(x^{(k)}, \lambda^{(k)}, \theta^{(k)})$ sufficiently close to $(x^*, \lambda^*, \theta^*)$, the QP has the same active set of constraints as the NLP, I know that

$$\nabla h(x^{(k)})^T p^{(k)} + h(x^{(k)}) = 0, \ i \in \mathcal{A}(x^*).$$

k	$\|x^* - x^{(k)}\|$	$\lambda^{(k)}$	$\theta^{(k)}$
0	2.6098	0.0000	0.0000
1	$9.5979 \cdot 10^{-1}$	1.2444	0.68889
2	$2.5192 \cdot 10^{-1}$	2.4849	1.6538
3	$3.0072 \cdot 10^{-2}$	2.9860	2.0215
4	$5.5362 \cdot 10^{-4}$	3.0242	2.0255
5	$1.9479 \cdot 10^{-7}$	3.0249	2.0249
6	$2.6198 \cdot 10^{-12}$	3.0249	2.0249

Table 1: Six iterations of the SQP method applied to Example 1.4. See also Figure 1.

I now introduce the following notation:

$$\begin{aligned} \mathcal{A}(x^*) &= \{i_1, i_2, \ldots, i_q\}, \\ \overline{h}(x) &= \begin{bmatrix} h_{i_1}(x) \\ h_{i_2}(x) \\ \vdots \\ h_{i_q}(x) \end{bmatrix}. \end{aligned}$$

Thus $\overline{h} : \mathrm{R}^n \to \mathrm{R}^q$ is the vector-valued function whose components are the components of h defining active constraints at $x = x^*$. Then, for $(x^{(k)}, \lambda^{(k)}, \theta^{(k)})$ sufficiently close to $(x^*, \lambda^*, \theta^*)$, $p^{(k)}$, $\lambda^{(k+1)}$, $\theta^{(k+1)}$ satisfy

$$\begin{aligned} \nabla^2 \ell(x^{(k)}; \lambda^{(k)}; \theta^{(k)}) p^{(k)} + \nabla f(x^{(k)}) - \nabla g(x^{(k)}) \lambda^{(k+1)} - \nabla h(x^{(k)}) \theta^{(k+1)} &= 0, \\ \nabla g(x^{(k)})^T p^{(k)} + g(x^{(k)}) &= 0, \\ \nabla \overline{h}(x^{(k)})^T p^{(k)} + \overline{h}(x^{(k)}) &= 0. \end{aligned} \tag{10}$$

On the other hand, if $\mathcal{A}(x^*)$ were known, then $(x^*, \lambda^*, \theta^*)$ could be computed by applying Newton's method to the system

$$\begin{aligned} \nabla f(x) - \nabla g(x) \lambda - \nabla \overline{h}(x) \overline{\theta} &= 0, \\ g(x) &= 0, \\ \overline{h}(x) &= 0. \end{aligned}$$

The Newton step at $(x^{(k)}, \lambda^{(k)}, \overline{\theta}^{(k)})$ takes the form

$$\begin{bmatrix} \nabla^2 \overline{\ell}(x^{(k)}; \lambda^{(k)}; \overline{\theta}^{(k)}) & -\nabla g(x^{(k)}) & -\nabla \overline{h}(x^{(k)}) \\ \nabla g(x^{(k)})^T & 0 & 0 \\ \nabla \overline{h}(x^{(k)})^T & 0 & 0 \end{bmatrix} \begin{bmatrix} p^{(k)} \\ \lambda^{(k+1)} - \lambda^{(k)} \\ \overline{\theta}^{(k+1)} - \overline{\theta}^{(k)} \end{bmatrix}$$

$$= \begin{bmatrix} -\nabla f(x^{(k)}) + \nabla g(x^{(k)}) \lambda^{(k)} + \nabla \overline{h}(x^{(k)}) \theta^{(k)} \\ -g(x^{(k)}) \\ -\overline{h}(x^{(k)}) \end{bmatrix}.$$

It is now easy to rearrange this system and show that it is equvilent to (10). It follows that

$$x^{(k)} \to x^*, \ \lambda^{(k)} \to \lambda^*, \ \overline{\theta}^{(k)} \to \overline{\theta}^*$$

quadratically (assuming x^* is a local minimizer and nonsingular point of the original NLP). Moreover, for k sufficiently large, $\theta_i^{(k)} = \theta_i^* = 0$ if $i \notin \mathcal{A}(x^*)$, and hence $\theta^{(k)} \to \theta^*$ quadratically as well.