Introduction to sequential quadratic programming

Mark S. Gockenbach

1 Introduction

Sequential quadratic programming (SQP) methods attempt to solve a nonlinear program directly rather than convert it to a sequence of unconstrained minimization problems. To make this introduction as simple as possible, I will begin by discussing the SQP framework for equality-constrained NLPs. The basic idea is analogous to Newton's method for unconstrained minimization: At each step, a local model of the optimization problem is constructed and solved, yielding a step (hopefully) toward the solution of the original problem. In unconstrained minimization, only the objective function must be approximated, and the local model is quadratic. In the NLP

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

both the objective function and the constraint must be modeled. An SQP method uses a quadratic model for the objective function and a linear model of the constraint. A nonlinear program in which the objective function is quadratic and the constraints are linear is called a *quadratic program* (QP). An SQP method solves a QP at each iteration.

Given an current estimate $x^{(k)}$ of a solution x^*, g can be approximated by

$$g(x^{(k)} + p) \doteq \nabla g(x^{(k)})^T p + g(x^{(k)}),$$

and so the constraint

$$g(x) = 0$$

is replaced by

$$\nabla g(x^{(k)})^T p + g(x^{(k)}) = 0.$$

At first glance, one would expect that the quadratic objective function for the model problem would be the Taylor approximation to f:

$$f(x^{(k)} + p) \doteq f(x^{(k)}) + \nabla f(x^{(k)}) \cdot p + \frac{1}{2} p \cdot \nabla^2 f(x^{(k)}) p.$$

However, this would be the wrong choice, because the curvature of the constraints must be captured by the model problem. I demonstrate this by an example below.

If λ^* is the Lagrange multiplier corresponding to a local minimizer x^* of

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

then the Lagrangian $\ell(\cdot; \lambda^*)$ has the property that $\ell(x; \lambda^*) = f(x)$ for all feasible x. It follows that

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

also has x^* as a local minimizer. Of course, λ^* is typically not known, but an algorithm can approximate λ^* as it approximates x^* (as the augmented Lagrangian method does, for example). Given $x^{(k)}$ and $\lambda^{(k)}$,

$$\ell(x^{(k)}+p;\lambda^{(k)}) \doteq \frac{1}{2}p\cdot\nabla^2\ell(x^{(k)};\lambda^{(k)})p+\nabla\ell(x^{(k)};\lambda^{(k)})\cdot p+\ell(x^{(k)};\lambda^{(k)}) \ \ \text{(for } p \text{ near 0).}$$

I will show below that solving

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

yields improved values of $x^{(k)}$ and $\lambda^{(k)}$ (at least when $x^{(k)}$ and $\lambda^{(k)}$ are close to x^* and λ^*, respectively). First, however, I give an example that illustrates the necessity of using the Lagrangian to define the quadratic programming subproblems.

Example 1.1 *Define* $f:\mathrm{R}^2\to\mathrm{R}$ *and* $g:\mathrm{R}^2\to\mathrm{R}$ *by*

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

Then the solution of the nonlinear program

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

is $x^*=(0,1)$*, and the corresponding Lagrange multiplier is* $\lambda^*=-1$*, as can be easily checked. Taking* $x^{(0)}=(1/10,3/2)$*, I form the QP*

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

or

$$\begin{array}{ll} \min & -p_1^2+p_2^2-\frac{1}{5}p_1-p_2+\frac{6}{25} \\ s.t. & \frac{4}{5}p_1+3p_2+\frac{129}{100}=0, \end{array}$$

in hopes that the solution $p^{(0)}$ *will define a step towards* x^**, allowing me to define* $x^{(1)}=x^{(0)}+p^{(0)}$*. However, the quadratic program is unbounded below and hence has no solution. This is clearly seen in Figure 1, which shows the contours of the quadratic objective function together with the linearized constraint and the original nonlinear constraint. This example shows that the quadratic objective function in the QP subproblem cannot simply approximate* f*.*

By contrast, if $\lambda^{(0)}=-1/2$*, then*

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

simplifies to

$$\begin{array}{ll} \min & 2p_1^2+3p_2^2+\frac{1}{5}p_1+\frac{1}{2}p_2+\frac{177}{200} \\ s.t. & \frac{4}{5}p_1+3p_2+\frac{129}{100}=0. \end{array}$$

This QP is well-posed; it is illustrated in Figure 2, which also shows $x^{(1)}=x^{(0)}+p^{(0)}$*, where* $p^{(0)}$ *is the solution to the QP. As Figure 2 shows,* $p^{(0)}$ *is a good step towards* x^**.*

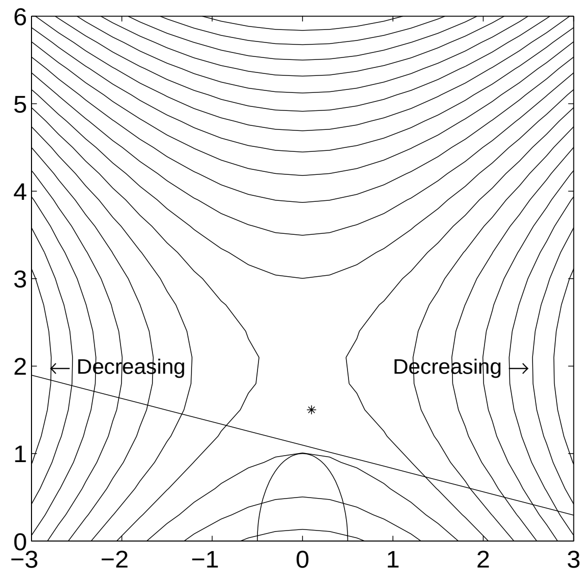

Figure 1: An illustration of the first proposed quadratic program in Example 1.1. The nonlinear constraint, the linearized constraint, and the contours of the quadratic objective function are shown. The asterisk indicates $x^{(0)} = (1/10, 3/2)$.

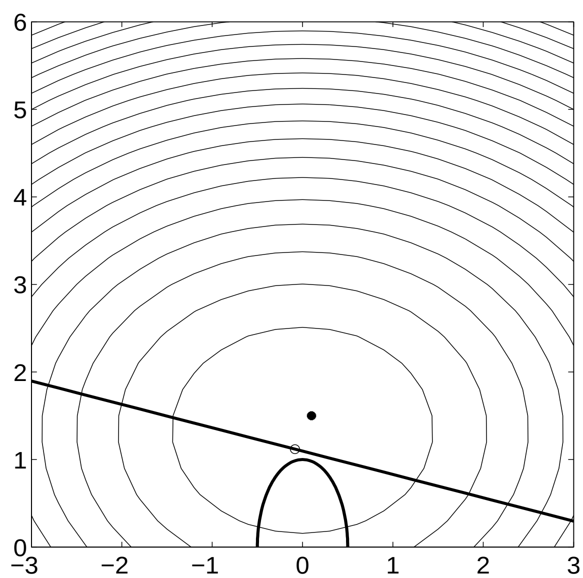

Figure 2: An illustration of the second proposed quadratic program in Example 1.1. The nonlinear constraint, the linearized constraint, and the contours of the quadratic objective function are shown. The asterisk indicates $x^{(0)} = (1/10, 3/2)$, while the small circle indicates $x^{(1)} \doteq (-0.0855, 1.1195)$.

2 Relationship of Newton's method to SQP

If x^* is a local minimizer and nonsingular point of

$$\min \quad f(x)$$

$$s.t. \quad g(x) = 0,$$

and λ^* is the corresponding Lagrange multiplier, then

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

Moreover, the Jacobian matrix

$$\begin{bmatrix} \nabla^2 \ell(x^*; \lambda^*) & -\nabla g(x^*) \\ -\nabla g(x^*)^T & 0 \end{bmatrix}$$

is nonsingular. It is therefore reasonable to try to compute x^*, λ^* by applying Newton's method to the system

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

Given an estimate $(x^{(k)}, \lambda^{(k)})$ of (x^*, λ^*), Newton's method defines $(x^{(k+1)}, \lambda^{(k+1)}) = (x^{(k)}, \lambda^{(k)}) + (p^{(k)}, \omega^{(k)})$, where the step $(p^{(k)}, \omega^{(k)})$ is defined by the linear system

$$\begin{bmatrix} \nabla^2 \ell(x^{(k)}; \lambda^{(k)}) & -\nabla g(x^{(k)}) \\ -\nabla g(x^{(k)})^T & 0 \end{bmatrix} \begin{bmatrix} p^{(k)} \\ \omega^{(k)} \end{bmatrix} = - \begin{bmatrix} \nabla \ell(x^{(k)}; \lambda^{(k)}) \\ -g(x^{(k)}) \end{bmatrix}. \tag{1}$$

On the other hand, assuming $(x^{(k)}, \lambda^{(k)})$ is sufficiently close to (x^*, λ^*) that $\nabla^2 \ell(x^{(k)}; \lambda^{(k)})$ is positive definite on the null space of $\nabla g(x^{(k)})^T$, then the quadratic program

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

is a convex program and has a unique solution-Lagrange multiplier pair $(p^{(k)}, \omega^{(k)})$. This pair is determined by the first-order necessary conditions:

$$\nabla^2 \ell(x^{(k)}; \lambda^{(k)}) p^{(k)} + \nabla \ell(x^{(k)}; \lambda^{(k)}) - \nabla g(x^{(k)}) \omega^{(k)} = 0, \tag{2}$$

$$\nabla g(x^{(k)})^T p^{(k)} + g(x^{(k)}) = 0. \tag{3}$$

But the system (2–3) is the same as system (1). In other words, the SQP method is equivalent to Newton's method applied to the first-order necessary conditions!. Therefore, at least locally, the SQP method defines not only a good step from $x^{(k)}$ towards x^*, but also a good step from $\lambda^{(k)}$ towards λ^*. In fact, under the usual assumptions, $(x^{(k)}, \lambda^{(k)}) \to (x^*, \lambda^*)$ quadratically.

The advantage of the SQP framework over simply applying Newton's method to the first-order necessary conditions is that the optimization framework gives us some basis for modifying the step when $(x^{(k)}, \lambda^{(k)})$ is not sufficiently close to (x^*, λ^*) that pure Newton's method defines a good step. This is the same reason that, in unconstrained minimization, it is advantageous to think of Newton's method as repeatedly minimizing a quadratic model rather than as trying to find a zero of the gradient. Sequential quadratic programming defines a locally convergent algorithm, and it is in this sense that one can talk about *the* SQP method. However, different techniques can be used to create a globally-convergent iteration, and so there can be many SQP algorithms.

Example 2.1 *Figure 3 illustrates four iterations of the local SQP method applied to Example 1.1 with* $x^{(0)} = (2, 4)$ *and* $\lambda^{(0)} = 1/2$. *The numerical results of nine iterations are summarized in Table 1, which shows* $x^{(k)} \to x^* = (0, 1)$ *and* $\lambda^{(k)} \to \lambda^* = -1$.

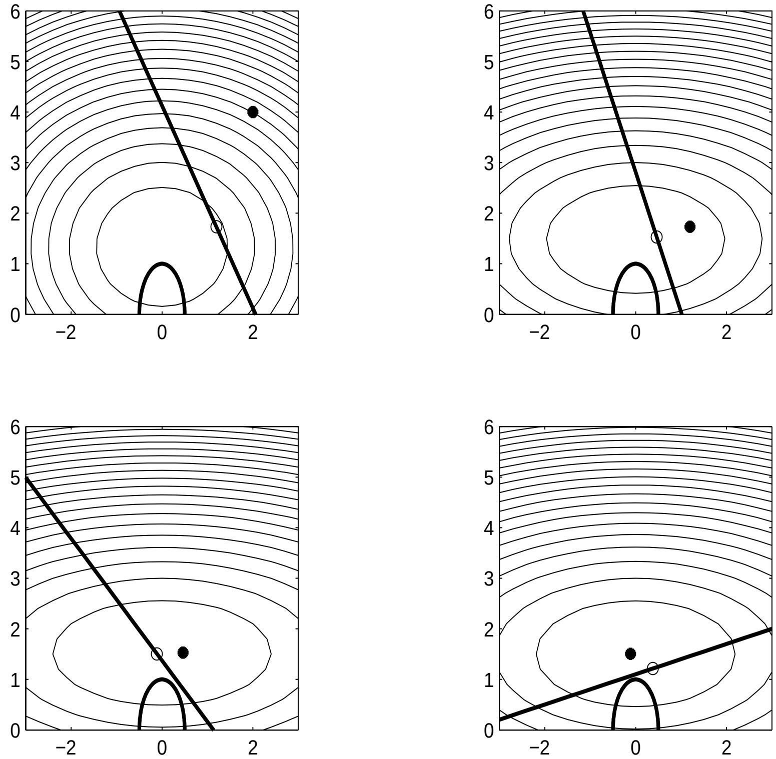

Figure 3: Four iterations of the local SQP method applied to the NLP from Example 1.1. The nonlinear constraint, the linearized constraint, and the contours of the quadratic objective function are shown. The asterisk indicates $x^{(k)}$ and the small circle $x^{(k+1)}$, $k = 0$ (upper left), $k = 1$ (upper right), $k = 2$ (lower left), and $k = 3$ (lower right).

3 An alternate formulation of the local SQP method

The following calculation suggests a slightly different formulation of the SQP algorithm:

$$\begin{aligned}
\nabla \ell(x^{(k)}; \lambda^{(k)}) \cdot p &= \left(\nabla f(x^{(k)}) - \nabla g(x^{(k)})\lambda^{(k)}\right) \cdot p \\
&= \nabla f(x^{(k)}) \cdot p - \left(\nabla g(x^{(k)})\lambda^{(k)}\right) \cdot p \\
&= \nabla f(x^{(k)}) \cdot p - \lambda^{(k)} \cdot \left(\nabla g(x^{(k)})^T p\right) \\
&= \nabla f(x^{(k)}) \cdot p + \lambda^{(k)} \cdot g(x^{(k)})
\end{aligned}$$

The last equality is valid if p satisfies the constraint

$$\nabla g(x^{(k)})^T p + g(x^{(k)}) = 0. \tag{4}$$

It follows that, on the linearized feasible set defined by (4), the two quadratics

$$\frac{1}{2} p \cdot \nabla^2 \ell(x^{(k)}; \lambda^{(k)}) p + \nabla \ell(x^{(k)}; \lambda^{(k)}) + \ell(x^{(k)}; \lambda^{(k)})$$

and

$$\frac{1}{2} p \cdot \nabla^2 \ell(x^{(k)}; \lambda^{(k)}) p + \nabla f(x^{(k)}) + f(x^{(k)})$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$\lambda^{(k)}$
0	2.0000	4.0000	−0.5000
1	1.1964	1.7321	−0.35045
2	$4.6212 \cdot 10^{-1}$	1.5308	−0.31165
3	$-1.1280 \cdot 10^{-1}$	1.5072	−0.32670
4	$3.7811 \cdot 10^{-1}$	1.2154	−0.58379
5	$-4.3954 \cdot 10^{-3}$	1.2598	−0.58767
6	$3.8193 \cdot 10^{-3}$	1.0269	−0.88110
7	$-6.9643 \cdot 10^{-4}$	1.0004	−0.99617
8	$3.5686 \cdot 10^{-6}$	1.0000	−1.0000
9	$-1.7129 \cdot 10^{-11}$	1.0000	−1.0000

Table 1: Nine iterations of the SQP method applied to Example 1.1. See also Figure 3.

differ by a constant, and therefore the QPs

$$\min \quad \frac{1}{2}p \cdot \nabla^2 \ell(x^{(k)};\lambda^{(k)})p + \nabla \ell(x^{(k)};\lambda^{(k)}) + \ell(x^{(k)};\lambda^{(k)}) \tag{5}$$

$$s.t. \quad \nabla g(x^{(k)})^T p + g(x^{(k)}) = 0 \tag{6}$$

and

$$\min \quad \frac{1}{2}p \cdot \nabla^2 \ell(x^{(k)};\lambda^{(k)})p + \nabla f(x^{(k)}) + f(x^{(k)}) \tag{7}$$

$$s.t. \quad \nabla g(x^{(k)})^T p + g(x^{(k)}) = 0 \tag{8}$$

have the same solution $p^{(k)}$. However, the Lagrange multipliers for the two QPs differ since the gradient of the objective functions are different. The optimality conditions for (7–8) are

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

or, equivalently,

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

Adding $\nabla g(x^{(k)})\lambda^{(k)}$ to both sides of the first equation yields

$$\nabla^2 \ell(x^{(k)};\lambda^{(k)})p^{(k)} - \nabla g(x^{(k)})\left(\overline{\omega}^{(k)} - \lambda^{(k)}\right) = -\nabla \ell(x^{(k)};\lambda^{(k)}), \tag{9}$$

$$\nabla g(x^{(k)})^T p^{(k)} + g(x^{(k)}) = 0. \tag{10}$$

I have already observed that QPs (5–6) and (7–8) have the same solution $p^{(k)}$. Comparing the optimality conditions (2–3) and (9–10) shows that

$$\overline{\omega}^{(k)} - \lambda^{(k)} = \omega^{(k)},$$

that is,

$$\overline{\omega}^{(k)} = \lambda^{(k)} + \omega^{(k)} = \lambda^{(k+1)}.$$

Therefore, when the QP is formulated as (7–8), the Lagrange multiplier for the QP is not the step to $\lambda^{(k+1)}$, it is actually $\lambda^{(k+1)}$ itself. For equality-constrained NLPs, this is the only significant difference between the two formulations of the QP subproblem. However, as I will show, the difference is much more significant for inequality-constrained NLPs, and it is necessary to adopt version quadratic objective (7).

A final remark is that the constant term in the quadratic objective is irrelevant for determining $p^{(k)}$ and $\omega^{(k)}$ or $\overline{\omega}^{(k)}$ and so it is usually not included. Therefore the objective function for the QP is taken to be

$$\frac{1}{2}p \cdot \nabla^2 \ell(x^{(k)}; \lambda^{(k)})p + \nabla \ell(x^{(k)}; \lambda^{(k)})$$

or

$$\frac{1}{2}p \cdot \nabla^2 \ell(x^{(k)}; \lambda^{(k)})p + \nabla f(x^{(k)})$$