Sequential Quadratic Programming Algorithm for Constrained Nonlinear Problems

Kuei-Yuan Chan, chanky@ntu.edu.tw
Professor of Mechanical Engineering,
National Taiwan University.

Let us consider continuous optimization problems of general nonlinear objective functions and nonlinear constraints in the form of Eq.(1). Let us further assume that the activity of all constraints are identified and all inactive constraints are removed from the equation. Without loss of generality, Eq.(1) only states equality constraints.

$$\begin{aligned} &\text{mimimize} f(\mathbf{x}) \\ &\text{s.t.} \quad \mathbf{h}(\mathbf{x}) = 0 \\ &\qquad \forall \mathbf{x} \in \mathcal{X} \end{aligned} \tag{1}$$

The first order necessary condition of Eq.(1) can be obtained by using the Lagrangian of the problem with respect to both $\mathbf{x}$ and λ as

$$\nabla L(\mathbf{x}_*, \lambda_*) = \mathbf{0}^T. \tag{2}$$

We may use the first order Taylor series expansion to solve this problem as

$$[\nabla L(\mathbf{x}_k + \partial \mathbf{x}_k, \lambda_k + \partial \lambda_k)] = \nabla L_k^T + \nabla^2 L_k (\partial \mathbf{x}_k, \partial \lambda_k)^T$$

and set the left-hand side $\nabla L_{k+1}^T = 0$, we have

$$\nabla^2 L_k \begin{pmatrix} \partial \mathbf{x}_k \\ \partial \lambda_k \end{pmatrix} = -\nabla L_k^T.$$

Since

$$\nabla^2 L_k = \begin{pmatrix} \nabla^2_{\mathbf{x}} L & \nabla^2_{\lambda \mathbf{x}} L \\ \nabla^2_{\mathbf{x}\lambda} L & \nabla^2_{\lambda} L \end{pmatrix}_k = \begin{pmatrix} \nabla^2 f + \lambda^T \nabla^2 \mathbf{h} & \nabla \mathbf{h}^T \\ \nabla \mathbf{h}^T & \mathbf{0} \end{pmatrix}_k \tag{3}$$

and

$$\nabla L_k = \left(\nabla f + \lambda^T \nabla \mathbf{h}\right)_k^T \tag{4}$$

To simplify our further representation, let us define

$$\mathbf{W} \triangleq \nabla^2 f + \lambda^T \nabla^2 \mathbf{h} \text{ and } \mathbf{A} \triangleq \nabla \mathbf{h} \tag{5}$$

Above equations can eventually be written as

$$\begin{pmatrix} \mathbf{W}_k & \mathbf{A}_k^T \\ \mathbf{A}_k & \mathbf{0} \end{pmatrix} \begin{pmatrix} \partial \mathbf{x}_k \\ \partial \lambda_k \end{pmatrix} = \begin{pmatrix} -\nabla f_k^T - \mathbf{A}_k \lambda_k \\ -\mathbf{h}_k \end{pmatrix} \tag{6}$$

Setting $\partial \mathbf{x}_k = \mathbf{s}_k$ and $\partial \lambda = \lambda_{k+1} - \lambda_k$, we can rewrite Eq.(6) as

$$\begin{pmatrix} \mathbf{W}_k & \mathbf{A}_k^T \\ \mathbf{A}_k & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{s}_k \\ \lambda_{k+1} \end{pmatrix} = \begin{pmatrix} -\nabla f_k^T \\ -\mathbf{h}_k \end{pmatrix} \tag{7}$$

Solving Eq.(7) iteratively, we obtain the iterates $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$ and λ_{k+1}. As iteration goes, we eventually reach the optimum $\mathbf{x}_*$ and λ_*. Therefore any methods that can solve Eq.(7) can be referred to as a *Lagrange-Newton* method for solving the constrained problem in Eq.(1).

Expanding Eq.(7) we have

$$\begin{aligned} \mathbf{W}_k \mathbf{s}_k + \mathbf{A}_k^T \lambda_{k+1} + \nabla f_k^T &= \mathbf{0} \\ \mathbf{A}_k \mathbf{s}_k + \mathbf{h}_k &= \mathbf{0} \end{aligned} \tag{8}$$

We may say that Eq.(8) is the KKT conditions for the quadratic model

$$\begin{aligned} \min q(\mathbf{s}_k) = f_k + \nabla_x L_k \mathbf{s}_k &+ \frac{1}{2} \mathbf{s}_k^t \mathbf{W}_k \mathbf{s}_k \\ \text{subject to } \mathbf{A}_k \mathbf{s}_k &+ \mathbf{h}_k = \mathbf{0} \end{aligned} \tag{9}$$

where $\nabla_{\mathbf{x}} L_k = \nabla f_k + \lambda_k^T \nabla \mathbf{h}_k$ and the multipliers of problem Eq.(9) are $\partial \lambda_k$.

The Lagrangian stationary condition of Eq.(9) is

$$\begin{aligned} \nabla_{\mathbf{x}} L_k + \mathbf{s}_k^T \mathbf{W}_k + (\lambda_k)^T \mathbf{A}_k &= \mathbf{0}^T \\ \rightarrow \nabla f_k^T + \mathbf{A}_k^T \lambda_k + \mathbf{W}_k \mathbf{s}_k + \mathbf{A}_k^T (\lambda_{k+1} - \lambda_k) &= 0 \end{aligned} \tag{10}$$

which is reduced to the first of Eq.(8).

Thus,. solving the *quadratic programming subproblem* Eq.(9) gives $\mathbf{s}_k$ and λ_{k+1} exactly as solving Eq.(8). IF the formulations in Eq.(8) is selected for solving the Lagrange-Newton equations, the values of $\mathbf{x}_*$ and λ_* will be obtained from solving a sequence of quadratic programming (QP) subproblems; hence the relevant algorithms are known as *sequential quadratic programming* (SQP) methods.

A simple SQP algorithm has the following general structure.

1. Select initial point $\mathbf{x}_0, \lambda_0$; let $k = 0$
2. For k=k+1, solve the QP subproblem and determine $\mathbf{s}_k$ and λ_{k+1}
3. Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{s}_k$
4. If termination criteria are not satisfy, return to 2.