On the Characterization of q-Superlinear Convergence of Quasi-Newton Methods for Constrained Optimization

By J. Stoer and R. A. Tapia*

Abstract. In this paper we present a short, straightforward and self-contained derivation of the Boggs-Tolle-Wang characterization of those quasi-Newton methods for equality-constrained optimization which produce iterates which are q-superlinearly convergent.

1. Introduction. We begin by considering the equality-constrained optimization problem

$$
\text{(1.1)} \qquad
\begin{aligned}
&\text{minimize} \quad f(x) \\
&\text{subject to} \quad g(x) = 0,
\end{aligned}
$$

where $f\colon R^n \to R$ and $g\colon R^n \to R^m$ $(m \leqslant n)$. Along with problem (1.1) we consider the Lagrangian $l(x,\lambda) = f(x) + \lambda^T g(x)$, a local solution x_* and its associated multiplier λ_* (i.e., λ_* is such that $\nabla_x l(x_*, \lambda_*) = 0$).

On occasion we will denote an operator evaluated at x_k or x_* by deleting the argument but instead using the subscript k or $*$ as the case may be, e.g., $g_* = g(x_*)$ or $f_k = f(x_k)$. We also denote the Hessian of the Lagrangian at (x_*, λ_*) by w_* (i.e., $w_* = \nabla_x^2 l(x_*, \lambda_*)$).

By a successive quadratic programming (SQP) quasi-Newton method for problem (1.1) we mean the iterative procedure

$$
\text{(1.2)} \qquad x_{k+1} = x_k + s_k, \qquad k = 0, 1, \ldots,
$$

where s_k solves the quadratic program

$$
\text{(1.3)} \qquad
\begin{aligned}
&\text{minimize} \quad \nabla f_k^T s + \tfrac{1}{2} s^T B_k s \\
&\text{subject to} \quad \nabla g_k^T s + g_k = 0
\end{aligned}
$$

for given B_k.

In the analysis of convergence rates for the SQP method the following assumptions are standard:

A1: $f, g \in C^2(D)$ where D is an open neighborhood of x_*,

A2: $x_k \in D$ and $\{x_k\}$ converges to x_*,

A3: $\nabla g(x)$ has full rank $\forall\, x \in D$,
A4: $\eta^T w_* \eta > 0 \;\forall\, \eta \neq 0$ such that $\nabla g_*^T \eta = 0$,
A5: $\eta^T B_k \eta > 0 \;\forall\, \eta \neq 0$ such that $\nabla g_k^T \eta = 0$ and $\forall\, k$.

Assumption A4 is second-order sufficiency for problem (1.1) and Assumption A5 is second-order sufficiency for the SQP subproblem (1.3). It follows that our subproblem is convex, has a unique solution and our iterative procedure is well defined. Moreover, for $x \in D$, Assumption A3 allows us to consider the projection operator

$$(1.4) \qquad P(x) = I - \nabla g(x)\big(\nabla g(x)^T \nabla g(x)\big)^{-1} \nabla g(x)^T.$$

Clearly, $P(x)$ projects onto the null space of $\nabla g(x)^T$.

Suppose that $\{x_k\}$ has been generated by the SQP method. Boggs, Tolle and Wang [1] show that, under the assumption that the convergence of $\{x_k\}$ to x_* is q-linear, the convergence will also be q-superlinear,

$$(1.5) \qquad \lim_{k \to \infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0,$$

if and only if

$$(1.6) \qquad \lim_{k \to \infty} \frac{\|P_k[B_k - w_*]s_k\|}{\|s_k\|} = 0.$$

This characterization result is a nice extension to constrained optimization of the Dennis-Moré [2] characterization for unconstrained optimization. Recently, Fontecilla, Steihaug and Tapia [4] derived the Boggs-Tolle-Wang characterization without the q-linear convergence assumption. Even more recently, Nocedal and Overton [7] also derived this characterization without the q-linear convergence assumption.

The following statements serve to motivate the present work. All three previous derivations of the Boggs-Tolle-Wang characterization leave something to be desired. The Boggs, Tolle and Wang [1] derivation is neither short nor direct and uses the unnecessary assumption of q-linear convergence; however we emphasize that it was the first derivation. The Fontecilla, Steihaug and Tapia [4] derivation is lengthy and not direct. This is to be expected since they solve a more difficult problem. Specifically, they obtain the Boggs-Tolle-Wang characterization as a special case of a characterization result for a more general class of quasi-Newton methods than those considered here. Members of their class need not give iterates which satisfy linearized constraints. Nocedal and Overton [7] give a short and direct derivation. However, their derivation is based on an existence theorem and a differentiation formula from differential geometry. The theorem and the formula are due to Goodman [5] and are nontrivial. It is not clear how their derivation could be given in a complete manner in an elementary presentation. The derivation of the Boggs-Tolle-Wang characterization was not the principal issue of these latter two papers.

In Section 2 we present several formulations which are equivalent to the SQP quasi-Newton formulation. In Section 3 we use one of these equivalent formulations and the Dennis-Moré characterization to derive the Boggs-Tolle-Wang characterization. Concluding remarks are given in Section 4.

2. Formulations Equivalent to SQP. The material in this section is taken from Tapia [10]. The reader interested in motivation and further detail is referred to that paper.

Extended System Formulation. If we apply the first-order necessary conditions to the quadratic programs (1.3), we see that the SQP quasi-Newton step s and its associated multiplier λ can be obtained from the following linear system:

(2.1a)
$$B_k s + \nabla g_k \lambda = -\nabla f_k,$$

(2.1b)
$$\nabla g_k^T s = -g_k.$$

By Assumption A5 we know that (1.3) is a convex program. It follows that in this case the first-order necessary conditions are also sufficient conditions. Also from A5 we know that s is unique. This means that the quadratic program (1.3) and the linear system (2.1) determine the same s and it is necessarily unique.

Multiplier Substitution Formulation. We will show that determining s from (2.1) is equivalent to determining s from the linear system

(2.2)
$$\left[P_k B_k + \nabla g_k \nabla g_k^T \right] s = -\left[P_k \nabla f_k + \nabla g_k g_k \right].$$

Toward this end, observe that if we define

(2.3)
$$\lambda' = -\left(\nabla g_k^T \nabla g_k \right)^{-1} \nabla g_k^T (B_k s + \nabla f_k),$$

then we can write

(2.4)
$$P_k [B_k s + \nabla f_k] = B_k s + \nabla f_k + \nabla g_k \lambda'.$$

Suppose that s has been obtained from (2.1). Multiplying (2.1a) by P_k, recalling that $P_k \nabla g_k = 0$ and using (2.1b), we see that s satisfies (2.2). Now, suppose s satisfies (2.2). Multiplying (2.2) by ∇g_k^T and recalling that $\nabla g_k^T P_k = 0$, we see that (2.1b) is satisfied. It follows that the left-hand side of (2.4) is zero; hence the right-hand side of (2.4) is zero. This means that (s, λ') is the unique solution of (2.1).

3. Derivation of the Boggs-Tolle-Wang Characterization. We begin with several simple observations. If P is given by (1.4) then

(3.1)
$$P(x)\nabla f(x) = \nabla f(x) + \nabla g(x)\lambda(x),$$

where

(3.2)
$$\lambda(x) = -\left(\nabla g(x)^T \nabla g(x) \right)^{-1} \nabla g(x)^T \nabla f(x).$$

It follows from (3.1) and (3.2) that if

(3.3)
$$F(x) = P(x)\nabla f(x) + \nabla g(x)g(x),$$

then

(3.4)
$$F'(x_*) = P_* w_* + \nabla g_* \nabla g_*^T.$$

Thus, we can interpret (2.2) (and therefore SQP) as a quasi-Newton method applied to the nonlinear system $F(x) = 0$, where F is given by (3.3) and the approximation to the Jacobian $F'(x_k)$ is given by $P_k B_k + \nabla g_k \nabla g_k^T$. Moreover, if $F'(x_*)$ is singular, then by the equivalence between (2.1) and (2.2) it follows that the matrix

(3.5)
$$\begin{pmatrix} B_* & \nabla g_* \\ \nabla g_*^T & 0 \end{pmatrix}$$

is singular. This in turn implies that the quadratic program (1.3) with $x_k = x_*$ and $B_k = w_*$ does not have a unique solution. This statement contradicts Assumptions A3 and A4.

Now, since $F \in C^1(D)$ and $F'(x_*)$ is nonsingular, the Dennis-Moré [2] characterization applies and (2.2) (therefore SQP) generates iterates which are q-superlinearly convergent if and only if

$$(3.6) \qquad \lim_k \left\| \left[P_k B_k + \nabla g_k \nabla g_k^T - \left(P_* w_* + \nabla g_* \nabla g_*^T \right) \right] s_k \right\| / \| s_k \| = 0.$$

Finally, by adding and subtracting $P_k w_*$ in (3.6) we see that (3.6) is equivalent to the Boggs-Tolle-Wang condition (1.6).

4. Summary and Concluding Remarks. In this note we have presented what we consider to be a short, direct and self-contained derivation of the Boggs-Tolle-Wang characterization of q-superlinear convergence for quasi-Newton methods for constrained optimization. While we have stated that the three previous derivations (Boggs, Tolle and Wang [1]; Fontecilla, Steihaug and Tapia [4] and Nocedal and Overton [7]) leave something to be desired, we quickly add that the present work was strongly influenced by these three papers. Indeed, the basic idea that led to the present derivation was to attempt to parallel the Nocedal-Overton derivation using a formulation of the quasi-Newton method which possessed the attribute that all necessary differentiations could be obtained in a straightforward manner. As we have seen, one of the formulations suggested by Tapia [10] possesses this property.

Institut für Angewandte Mathematik und Statistik
Am Hubland
D-8700 Würzburg, West Germany

Department of Mathematical Sciences
Rice University
Houston, Texas 77251

1. P. T. BOGGS, J. W. TOLLE & P. WANG, "On the local convergence of quasi-Newton methods for constrained optimization," *SIAM J. Control Optim.*, v. 20, 1982, pp. 161–171.

2. J. E. DENNIS, JR. & J. J. MORÉ, "A characterization of superlinear convergence and its application to quasi-Newton methods," *Math. Comp.*, v. 28, 1974, pp. 549–560.

3. T. F. COLEMAN & A. R. CONN, "On the local convergence of a quasi-Newton method for the nonlinear programming problem," *SIAM J. Numer. Anal.*, v. 21, 1984, pp. 755–769.

4. R. FONTECILLA, T. STEIHAUG & R. A. TAPIA, *A Convergence Theory for a Class of Quasi-Newton Methods for Constrained Optimization*, Report 83-15, Department of Mathematical Sciences, Rice University, Houston, Texas, 1983. To appear in *SIAM J. Numer. Anal.*

5. J. GOODMAN, "Newton's method for constrained optimization," *Math. Programming.* (To appear.)

6. S. P. HAN, "Superlinearly convergent variable metric algorithms for general nonlinear programming problems," *Math. Programming*, v. 11, 1976, pp. 263–282.

7. J. NOCEDAL & M. OVERTON, "Projected Hessian updating algorithms for nonlinearly constrained optimization," *SIAM J. Numer. Anal.*, v. 22, 1983, pp. 821–850.

8. M. J. D. POWELL, "The convergence of variable metric methods for nonlinearly constrained optimization problems," in *Nonlinear Programming 3* (O. Mangasarian, R. Meyer and S. Robinson, eds.), Academic Press, New York, 1978, pp. 27–63.

9. R. A. TAPIA, "Diagonalized multiplier methods and quasi-Newton methods for constrained optimization," *J. Optim. Theory Appl.*, v. 22, 1977, pp. 135–194.

10. R. A. TAPIA, "Quasi-Newton methods for equality constrained optimization: Equivalence of existing methods and a new implementation", in *Nonlinear Programming 3* (O. Mangasarian, R. Meyer and S. Robinson, eds.), Academic Press, New York, 1978, pp. 125–163.