The Local Convergence of Sequential Quadratic Programming Methods

by

J. Stoer[1]

and

R.A. Tapia[2]

Technical Report 87-4, March 1987

[1]Institut für Angewandte Mathematik und Statistik, Universität Würzburg, Am Hubland, D-8700 Würzburg, W. Germany.

[2]Mathematical Sciences Department, Rice University, Houston, Texas 77251-1892. Research sponsored by DOE DE-FG05-86ER25017, ARO DAAG-03-86-K-0113, and AFOSR 85-0243.

The Local Convergence of Sequential Quadratic Programming Methods for Solving Nonlinear Programs

J. Stoer[1] and R.A. Tapia[2]

Summary. Sequential quadratic programming methods for solving constrained nonlinear optimization problems (P) generate iterates x_k, $x_{k+1} := \Phi_k(x_k)$, by means of a certain iteration function Φ_k, which has any Kuhn-Tucker point x_* of (P) as fixed point. By studying the differentiability properties of $\Phi_k(x)$ close to x_*, we obtain easy and straightforward proofs for some fundamental results on the convergence speed of the iterates x_k, e.g. the Boggs-Tolle-Wang characterization of their Q-superlinear convergence, and Powell's criterion for their 2-step Q-superlinear convergence.

Subject classification: AMS(MOS): 49D15, 65K05.

[1] Institut für Angewandte Mathematik und Statistik, Universität Würzburg, Am Hubland, D-8700 Würzburg, W.- Germany.

[2] Mathematical Sciences Department, Rice University, Houston, Texas 77251-1892. Research sponsored by DOE DE-FG05-86ER25017, ARO DAAG-03-86-K-0113, and AFOSR 85-0243.

1. Introduction

In this paper we are concerned with some theoretical aspects of sequential quadratic programming (SQP) methods for solving the following standard problem of nonlinear programming

$$(P) \qquad \begin{array}{ll} \min & f(x) \\ x \in \mathbf{R}^n : & g(x) = 0. \end{array}$$

Here $f: \mathbf{R}^n \to \mathbf{R}$, and $g: \mathbf{R}^n \to \mathbf{R}^m$, $g(x) = (g_1(x), \ldots, g_m(x))^T$, are sufficiently smooth functions. (For the ease of exposition, we consider only equality constrained problems, even though essentially all results of this paper carry over to the case of problems with both inequality and equality constraints). By $L(x,u) := f(x) + u^T g(x)$ we denote the Lagrangian of (P). We denote by $Df(x) = \nabla f(x)^T$ the transpose of the gradient of f, by $Dg(x)$ the Jacobian of g, by $D^2 f(x) = f''(x)$ the Hessian of f and by $W(x,u) := D_x^2 L(x,u)$ the Hessian of the Lagrangian $L(x,u)$ with respect to x.

We are interested in the convergence behaviour of the familiar SQP-methods (see Han [6], Powell [8]) close to a Kuhn-Tucker point x_* of (P), where f and g satisfy the following standard assumptions

(A) 1) $f, g \in C^2(\mathbf{R}^n)$.

2) rank $Dg(x_*) = m$, and

3) there is a Kuhn-Tucker multiplier $u_* \in \mathbf{R}^m$ with

$$\nabla L(x_*, u_*) = \begin{pmatrix} \nabla_x L(x_*, u_*) \\ g(x_*) \end{pmatrix} = \begin{pmatrix} \nabla f(x_*) + Dg(x_*)^T u_* \\ g(x_*) \end{pmatrix} = 0,$$

that is, x_* satisfies the *first order necessary conditions* for a solution of (P).

4) The Hessians $f''(x)$, $g_i''(x)$ are Lipschitz continuous at x_*:

$$\|f''(x) - f''(x_*)\|, \quad \|g_i''(x) - g_i''(x_*)\| \leq L\|x - x_*\|$$

for x close to x_*.

(x_*, u_*) is called a K.-T. pair of (P). A K.-T. point x_* (resp. a K.-T. pair (x_*, u_*)) satisfies the *second order sufficient condition* for a solution of (P), if, with the abbreviation $W_* := W(x_*, u_*)$,

$$(1.1) \qquad s^T W_* s > 0 \quad \text{for all} \quad s \neq 0 \quad \text{with} \quad Dg(x_*)s = 0,$$

in which case x_* is a strict local solution of (P). It satisfies the *second order necessary condition* for a solution of (P), if

$$(1.2) \qquad s^T W_* s \geq 0 \quad \text{for all} \quad s \quad \text{with} \quad Dg(x_*)s = 0.$$

As is well-known, SQP-methods generate a sequence $\{x_k\}$ according to the following iterative scheme. Given x_k and a symmetric $n \times n$-matrix B_k, one first finds a K.-T. point s_k of the quadratic program

$$(P_k) \qquad \begin{array}{ll} \min & Df(x_k)s + \frac{1}{2} s^T B_k s \\ s : & g(x_k) + Dg(x_k)s = 0 \end{array}$$

and defines x_{k+1} by

$$(1.3) \qquad x_{k+1} := x_k + \lambda_k s_k.$$

Here $\lambda_k > 0$ is a stepsize chosen by some line search technique involving merit functions, which is usually such that it ensures $\lambda_k = 1$ for x_k close to a local solution x_* of (P).

We note that the K.-T. point s_k and the associated Lagrange multiplier v_k of (P_k) can be obtained by solving the linear equations

$$(1.4)\qquad \begin{pmatrix} B_k & Dg(x_k)^T \\ Dg(x_k) & 0 \end{pmatrix} \begin{pmatrix} s_k \\ v_k \end{pmatrix} = - \begin{pmatrix} \nabla f(x_k) \\ g(x_k) \end{pmatrix}.$$

(1.3) and (1.4) determine a mapping $x_k \mapsto x_{k+1} =: \Phi_k(x_k)$, where the mapping function Φ_k depends on k via the choice of B_k and the step λ_k. It is easily seen, that x_k is a fixed point of the mapping, iff x_k is a K.-T. point of (P). Within this paper, we plan to study the convergence behaviour of the iteration (1.3), (1.4), $x_{k+1} = \Phi_k(x_k)$, near a fixed point x_* by studying $\phi_k(x)$ near x_*. In particular, we will exploit the differentiability properties of the function $\Phi_k(x)$ if x is close to x_* and $B_k = B_k^T$ satisfies some mild boundedness conditions. Moreover we will give explicit formulae for the Jacobian $D\Phi_k(x_*)$ of $\Phi_k(x)$ at x_*.

In this way we will discover the following results:

If the iteration function $\Phi_k \equiv \Phi$ does not depend on k (i.e. $B_k \equiv B$ and $\lambda_k \equiv \lambda \geq 0$) and $B = B^T$ is positive definite, then any K.-T. point x_* of (P) satisfying the second order *sufficient* condition (1.1), is an *attractive* fixed point of the iteration, and any K.-T. point which does *not* satisfy the second order *necessary* condition (1.2) is a *repulsive* fixed point of $\Phi_k \equiv \Phi$.

This describes a natural stability property of SQP-methods and explains why it is reasonable to always choose B_k to be positive definite: Loosely speaking, in this case the method is attracted by local minimizers of (P) , but avoids automatically all wrong K.-T. points of (P), which are not local minimizers.

For variable Φ_k (B_k variable, but fixed $\lambda_k \equiv 1$) , we obtain, as a byproduct, another very short proof (and a slight generalization) of the necessary and sufficient conditon of Boggs, Tolle and Wang [1] for the Q-superlinear convergence of $\{x_k\}$

$$\lim_{k\to\infty} \|x_{k+1} - x_*\| \,/\, \|x_k - x_*\| = 0$$

towards a local minimizer x_* of (P) satisfying (1.1) (see also [9]).

In the same way, also a short proof is obtained for Powell's [8] condition ensuring the 2-step Q-superlinear convergence of $\{x_k\}$

$$\lim_{k\to\infty} \|x_{k+1} - x_*\| \,/\, \|x_{k-1} - x_*\| = 0,$$

where again x_* is a local minimizer of (P) satisfying (1.1).

2. An Expression for the Jacobian of Φ_k.

In this section we derive an explicit analytic expression for the Jacobian $D\Phi_k(x_*)$ of the iteration function Φ_k defined by (1.3), (1.4) at a K.-T. point x_* of (P). We assume that (A) holds at x_*. Consider the $m \times n$-matrix $Dg(x)$ for x close to x_*.

We show first

Lemma 1. *Under assumption (A) there is an $\epsilon_1 > 0$ and a $L_1 > 0$ such that the following holds: For every unitary $n \times n$-matrix $Q_* = (Y_*, Z_*)$, with*

$$(2.1)\qquad Dg(x_*)Q_* = Dg(x*)(Y_*, Z_*) = (R_*^T, 0),$$

where Y_* and R_* have m columns and Z_* has n columns, there are continuously differentiable matrix functions $Q(x) = \big(Y(x), Z(x)\big)$, $R(x)$, which are defined for all $\|x - x*\| \leq \epsilon_1$ and satisfy

$$(2.2)\qquad \begin{aligned} &Dg(x)Q(x) = Dg(x)\big(Y(x), Z(x)\big) = (R(x)^T, 0) \\ &Q(x)^T Q(x) = I, \qquad R(x) \quad \textit{is nonsingular} \end{aligned}$$

and

$$Q(x_*) = \big(Y(x_*), Z(x_*)\big) = Q_* = (Y_*, Z_*), \quad R(x_*) = R_*.$$

Moreover, their first derivatives $DQ(x)$ and $DR(x)$ are Lipschitz continuous at x_*,

$$\|DQ(x) - DQ(x_*)\|, \quad \|DR(x) - DR(x_*)\| \leq L_1 \|x - x_*\|$$

for all $\|x - x_*\| \leq \epsilon_1$.

Note that the columns of $Z(x)$ provide an orthogonal basis of the nullspace of $Dg(x)$.

Proof: Let Q_*, R_* be any matrices satisfying (2.1). Because rank $Dg(x_*) = m$ by assumption (A), R_* is nonsingular. Then again by (A) the matrix function

$$H(x) := Q_*^T \left(Dg(x)^T R_*^{-1}, Z_*\right)$$

is continuously differentiable for all $x \in \mathbf{R}^n$ and satisfies $H(x_*) = I$.

Now every nonsingular $n \times n$-matrix H has a unique QR-factorization

$$H = A \cdot S, \qquad A^T A = I$$

where $S = (s_{ik})$ is an upper triangular matrix with $s_{ii} > 0$. The factorization can be computed by the known Gram-Schmidt orthogonalization procedure applied to the columns of H, which gives rise to a map Ψ

$$H \mapsto \Psi(H) := (A, S).$$

The function Ψ is well-defined on the open set of nonsingular $n \times n$-matrices H. It is a C^∞-function on this set, as the Gram-Schmidt procedure computes $\Psi(H)$ by finitely many additions, multiplications, divisions and square roots, where for nonsingular H, the denominators of the divisions are always nonzero and the arguments of the square roots are positive. Hence $\Psi(H)$ and all its Fréchet derivatives $D^r\Psi(H)$ are Lipschitz continuous on any compact subset of the set of nonsingular matrices H, say for $\|H - I\| \leq \frac{1}{2}$. Now by assumption $(A)1), 2)$ the function $H(x)$ is continuously differentiable for all x and $DH(x)$ is locally Lipschitz at x_*. Hence by $H(x_*) = I$ there is an $\epsilon_1 > 0$ so that

$$\|H(x) - I\| \leq \frac{1}{2} \quad \text{for} \quad \|x - x_*\| \leq \epsilon_1.$$

Therefore, also $\Psi(H(x)) = \big(A(x), S(x)\big)$ is continuously differentiable for all $\|x - x_*\| \leq \epsilon_1$ and satisfies

$$\begin{aligned} I &= A(x)^T A(x) \\ H(x) &= A(x) S(x), \quad S(x)^{-1} \text{exists} \\ H(x_*) &= A(x_*) = S(x_*) = I \end{aligned}$$

$$\|\big(DA(x) - DA(x_*), DS(x) - DS(x_*)\big)\| \leq L_1 \|x - x*\|.$$

Now, by definition of $H(x)$,

$$A(x)^T Q_*^T \left(Dg(x)^T R_*^{-1}, Z_*\right) = S(x) = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix},$$

with $S_{11} = S_{11}(x)$ a nonsingular upper triangular $m \times m$-matrix. It follows that for $\|x - x_*\| \leq \epsilon_1$ the matrix functions

$$Q(x) = \big(Y(x), Z(x)\big) := Q_* A(x), \quad R(x) := S_{11}(x) R_*$$

are continuously differentiable and have the properties required in the Lemma.

Finally, we note that the constants ϵ_1, L_1 can be chosen independently of the matrices $Q_* = (Y_*, Z_*)$ and R_* satisfying (2.1). In fact, every unitary matrix $\bar{Q} = (\bar{Y}, \bar{Z})$, and every $\bar{R}$ satisfying (2.1) have the form

$$\bar{Q} = (\bar{Y}, \bar{Z}) \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix}, \quad \bar{R}^T = R_*^T T_1$$

with unitary matrices T_1, T_2 of appropriate dimension. The matrix $\bar{H}(x)$ belonging to $\bar{Q}$, $\bar{R}$ then has the form

$$\bar{H}(x) = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix}^T H(x) \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix},$$

so that $\|\bar{H}(x)\|_2 = \|H(x)\|_2$. This implies that ϵ_1 and L_1 can be chosen independently of Q_*, R_*. This completes the proof of the lemma.

Note that also the matrices $Q(x), R(x)$ are by no means uniquely determined by (2.2). However, it is easy to see that any matrices $Q_x = (Y_x, Z_x), R_x$ satisfying (2.2) can be written in terms of the differentiable functions $Q(x)$ and $R(x)$ of the lemma in the form

$$Y_x = Y(x)T_{1x}, \quad Z_x = Z(x)T_{2x}, \quad R_x = R(x)T_{1x}$$

with suitable unitary matrices T_{ix} of appropriate dimension, where , of course, the matrices Q_x, R_x, T_{ix} need not be differentiable functions of x. But in any case, the matrices

$$Z(x)Z(x)^T = Z_x Z_x^T, \qquad Y(x)Y(x)^T = Y_x Y_x^T$$

have an invariant meaning: they are the orthogonal projections onto the nullspace of $Dg(x)$ and its orthogonal complement, respectively.

Henceforth, $Q(x) = (Y(x), Z(x))$ and $R(x)$ denote the particular matrices constructed above from Q_*, which depend differentiably on x for $\|x - x_*\| \le \epsilon_1$ and on the initial choice of Q_*. One should carefully check in every particular case that all hypotheses and results formulated in terms of the matrices $Q(x)$, $R(x)$ have an invariant meaning. As an example we note that the sufficient condition (1.1) is equivalent to the positive definiteness of the matrix $Z_*^T W_* Z_*$, as the columns of Z_* span the nullspace of $Dg(x_*)$. Here, of course, one gets the same condition if one replaces Z_* by any other matrix Z whose columns span the same nullspace.

In order to investigate the convergence behaviour of SQP-methods near a K.-T. point x_* of (P) we now assume that (see Lemma 1) $\|x - x_*\| \le \epsilon_1$ and that $B = B^T$ is a symmetric matrix. Then any vector $s = s(x, B)$ (see(1.4)) such that

$$\begin{pmatrix} B & Dg(x)^T \\ Dg(x) & 0 \end{pmatrix} \cdot \begin{pmatrix} s \\ v \end{pmatrix} = - \begin{pmatrix} \nabla f(x) \\ g(x) \end{pmatrix} \tag{2.3}$$

can be obtained in the following standard way: Introduce new variables s_Y,s_Z by

$$s = Y(x)s_Y + Z(x)s_Z,$$

and multiply the first set of equations (2.3) by $Z(x)^T$. Then by (2.2)

$$\begin{pmatrix} Z(x)^T B \\ Dg(x) \end{pmatrix} s \equiv \begin{pmatrix} Z(x)^T BY(x) & Z(x)^T BZ(x) \\ R(x)^T & 0 \end{pmatrix} \cdot \begin{pmatrix} s_Y \\ s_Z \end{pmatrix} = - \begin{pmatrix} Z(x)^T \nabla f(x) \\ g(x) \end{pmatrix}. \tag{2.4}$$

We see that (2.3), (2.4) has a unique solution $s = s(x, B)$ iff

$$Z(x)^T BZ(x) \quad \text{is nonsingular,} \tag{2.5}$$

a condition, which is independent of the choice of $Z(x)$. Again, noting that $Dg(x)Z(x) = 0$, we have

$$Z(x)^T \nabla f(x) = Z(x)^T \nabla_x L(x, u)$$

for arbitrary u, hence also for $u = u_*$. Therefore, if (2.5) holds, the solution $s = s(x, B)$ of (2.3) is given by the function $\Xi(.,.)$

$$\begin{aligned} s &= -Q(x) \begin{pmatrix} Z(x)^T BY(x) & Z(x)^T BZ(x) \\ R(x)^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} Z(x)^T \nabla_x L(x, u_*) \\ g(x) \end{pmatrix} \\ &\equiv: \Xi(x; B), \end{aligned} \tag{2.6}$$

for which we prove the following lemma:

Lemma 2. *Assume that (A) holds and let $c > 0$ be an arbitrary positive number. Then there exist positive numbers $0 < \epsilon = \epsilon(c) \le \epsilon_1$ and $C = C(c) > 0$ so that the following holds:*
Let $B = B^T$ be any $n \times n$-matrix such that

$$\|B\|, \quad \|\left(Z(y)^T BZ(y)\right)^{-1}\| \le c \quad \text{for some} \quad \|y - x_*\| \le \epsilon. \tag{2.7}$$

Then the function $\Xi(x; B)$ is continuously differentiable in x for all $\|x - x_*\| \le \epsilon$ and its Jacobian $D_x\Xi(x; B)$ is Lipschitz continuous in x at x_* with Lipschitz constant C:

$$\|D_x\Xi(x, B) - D_x\Xi(x_*; B)\| \le C\|x - x_*\|$$

holds for all $\|x - x_*\| \le \epsilon$ and all $B = B^T$ satisfying (2.7). Moreover, $D_x\Xi(x_*; B)$ is given by

$$D_x\Xi(x_*; B) = -Q_* \begin{pmatrix} Z_*^T BY_* & Z_*^T BZ_* \\ R_*^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} Z_*^T W_* \\ Dg(x_*) \end{pmatrix}. \tag{2.8}$$

Note that the formulae (2.6),(2.8) and the hypothesis (2.7) of the lemma do not depend on the choice of $Q(x)$: they are invarianr under any change $Y(x) \to Y(x)T_{1x}$, $Z(x) \to Z(x)T_{2x}$ with unitary T_{ix}.

Proof: Assume $c > 0$ and (2.7). By Lemma 1, we have

$$\begin{aligned} \|Z(x) - Z_*\| &\le L_1\|x - x_*\| \\ \|R(x) - R_*\| &\le L_1\|x - x_*\| \end{aligned}$$

for all $\|x - x_*\| \le \epsilon_1$. We define

$$\epsilon := \min\{\epsilon_1, 1/(8c^2 L_1)\}.$$

Then by (2.7) the following holds for all $\|x - x_*\| \le \epsilon$:

$$\begin{aligned} \|x - y\| &\le 2\epsilon \\ Z(x)^T BZ(x) &= Z(y)^T BZ(y)[I + \Delta] \end{aligned}$$

where

$$\begin{aligned} \Delta &:= \left(Z(y)^T Z(y)\right)^{-1} \left[\left(Z(x) - Z(y)\right)^T BZ(x) + Z(y)^T B\left(Z(x) - Z(y)\right)\right] \\ \|\Delta\| &\le 2c^2 L_1\|x - y\| \le 4c^2 L_1 \epsilon \le \tfrac{1}{2}. \end{aligned}$$

Therefore $\left(Z(x)^T BZ(x)\right)^{-1}$ exists and is bounded:

$$\|\left(Z(x)^T BZ(x)\right)^{-1}\| \le 2\|\left(\left(Z(y)^T BZ(y)\right)^{-1}\| \le 2c. \tag{2.9}$$

Also, as $\epsilon \le \epsilon_1$, $R(x)^{-1}$ exists and is bounded

$$\max_{\|x-x_*\|\le\epsilon} \|R(x)^{\pm 1}\| < +\infty.$$

In order to abbreviate the following arguments, let us say that a function $\phi(x)$ which is defined for $\|x - x_*\| \le \epsilon$ has *property (S)*, if $\phi(x)$ is continuously differentiable and its derivative $D\phi(x)$ is Lipschitz continuous at x_*,

$$\|D\phi(x) - D\phi(x_*)\| \le C\|x - x_*\| \quad \text{for} \|x - x_*\| \le \epsilon$$

with a Lipschitz constant C which only depends on the problem (P) and the constant c of this lemma, but not on B.

Then by (A), $g(x)$ and $\nabla_x L(x, u_*)$ have property (S), and, by Lemma 1, also the other elementary building stones $Q(x) = (Y(x), Z(x))$, $R(x)$ of (2.6). Next, using the familiar differentiation rules, (2.9) and the boundedness of $\|R(x)^{\pm 1}\|$ for $\|x - x_*\| \le \epsilon$ one can prove property (S) for more and more complicated expressions appearing during the evaluation of (2.6), like

$$Z(x)^T B, \quad Z(x)^T BZ(x), \quad R(x)^{-1}, \quad \left(Z(x)^T BZ(x)\right)^{-1}$$
$$T(x) := \begin{pmatrix} Z(x)^T BY(x) & Z(x)^T BZ(x) \\ R(x)^T & 0 \end{pmatrix}, \quad T(x)^{-1}, \quad Q(x)T(x)^{-1}$$

and finally for $\Xi(x; B)$ itself. The explicit formula (2.8) for $D_x\Xi(x_*; B)$ follows immediately from (2.6) because $\nabla_x L(x_*, u_*) = 0$.

3. Local Convergence Analysis

We now apply Lemma 2 to analyse the local convergence behaviour of SQP-methods near a K.-T. point x_* of (P) under assumption (A). In terms of the function $\Xi(x; B)$, (2.6), the basic iteration (1.3), (1.4) can be written as follows

$$x_{k+1} = x_k + \lambda_k s_k = \Phi_k(x_k)$$

where

(3.1) $$\Phi_k(x) := x + \lambda_k \Xi(x; B_k).$$

Assume first that in all iterations we use the same positive definite $B_k \equiv B = B^T$ and the same step $\lambda_k \equiv \lambda$ for all $k > 0$,

$$x_{k+1} = \Phi(x_k), \qquad \Phi(x) := x + \lambda\Xi(x; B).$$

Also the matrix $Z_*^T BZ_*$ is required to be nonsingular so that we may apply Lemma 1. Then the convergence behaviour of this method is determined by the spectral radius $\rho(D\Phi(x_*))$ of the Jacobian $D\Phi(x_*)$, which, by (2.8), is given by

(3.2)
$$\begin{aligned} D\Phi(x_*) &= I - \lambda Q_* \begin{pmatrix} Z_*^T BY_* & Z_*^T BZ_* \\ R_*^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} Z_*^T W_* \\ Dg(x_*) \end{pmatrix} \\ &= Q_* \left[I - \lambda \begin{pmatrix} I & 0 \\ 0 & (Z_*^T BZ_*)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ Z_*^T (W_* - B)Y_* & Z_*^T W_* Z_* \end{pmatrix} \right] Q_*^T. \end{aligned}$$

Here we have used that

$$\begin{pmatrix} Z_*^T W_* \\ Dg(x_*) \end{pmatrix} = \begin{pmatrix} Z_*^T W_* Y_* & Z_*^T W_* Z_* \\ R_*^T & 0 \end{pmatrix} Q_*^T.$$

It follows from (3.2) immediately that

$$\rho(D\Phi(x_*)) = \max\Big[|1-\lambda|, \rho(I - \lambda(Z_*^T B Z_*)^{-1}(Z_*^T W_* Z_*))\Big].$$

Assume that B has been chosen such that B or at least $Z_*^T B Z_*$ is positive definite. Since the matrix

$$(Z_*^T B Z_*)^{-1}(Z_*^T W_* Z_*)$$

is similar to

$$(Z_*^T B Z_*)^{-1/2}(Z_*^T W_* Z_*)(Z_*^T B Z_*)^{-1/2},$$

the eigenvalues of $(Z_*^T B Z_*)^{-1}(Z_*^T W_* Z_*)$ are real and have the same signs as the eigenvalues of $Z_*^T W_* Z_*$. Therefore, if $Z_*^T W_* Z_*$ is positive definite, that is (1.1) is satisfied, there is a $\bar{\lambda}$ with $0 < \bar{\lambda} < 2$, such that

$$\rho(D\Phi(x_*)) < 1$$

for all $\lambda_k \equiv \lambda$ with $0 < \lambda \leq \bar{\lambda}$, which proves

Theorem 1. *Let x_* be a K.-T. point of (P) satisfying the 2nd order sufficient condition (1.1), and let B be any symmetric matrix, such that $Z_*^T B Z_*$ is positive definite. Then there is a $\bar{\lambda} > 0$ such that the sequence $\{x_k\}$ given by the SQP-method with fixed $B_k \equiv B$ and fixed step $\lambda_k \equiv \lambda$, $0 < \lambda \leq \bar{\lambda}$, converges locally (if $\|x_0 - x_*\|$ is sufficiently small) at least linearly to x_*.*

On the other hand, if x_* violates (1.2), that is if $Z_*^T B Z_*$ has at least one negative eigenvalue, then so has $(Z_*^T B Z_*)^{-1}(Z_*^T W_* Z_*)$, and therefore

$$\rho(D\Phi(x_*)) > 1$$

for any choice of $\lambda > 0$, however small. Hence

Theorem 2. *Let x_* be a K.-T. point of (P) violating the 2nd order necessary condition (1.2) for a minimizer of (P), and let B be such that $Z_*^T B Z_*$ is positive definite. Then for every fixed stepsize $\lambda_k \equiv \lambda > 0$, and matrices $B_k \equiv B$, x_* is a repulsive fixed point of the SQP-method.*

We now treat the case of variable B_k, but with fixed stepsizes $\lambda_k \equiv 1$, which is typical for the later stages of the SQP-method in applications. In this case (3.2) (if Lemma 1 is applicable, e.g. if $Z_* B_k Z_*$ is nonsingular) simplifies to

$$\begin{aligned} D\Phi_k(x_*) &= Q_* \begin{pmatrix} I & 0 \\ 0 & (Z_*^T B_k Z_*)^{-1} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ Z_*^T (B_k - W_*) Y_* & Z_*^T (B_k - W_*) Z_* \end{pmatrix} Q_*^T \\ I - D\Phi_k(x_*) &= Q_* \begin{pmatrix} I & 0 \\ 0 & (Z_*^T B_k Z_*)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ Z_*^T (W_* - B_k) Y_* & Z_*^T W_* Z_* \end{pmatrix} Q_*^T. \end{aligned} \tag{3.5}$$

Now, it is easy to prove the following theorem of Boggs, Tolle and Wang [1]:

Theorem 3. *Assume that x_* is a K.-T. point of (P) for which assumption (A) holds, and that the SQP-method (1.3), (1.4) with $\lambda_k \equiv 1$ corresponding to the matrices $B_k = B_k^T$ generates vectors $x_k \neq x_*$, $k \geq 0$, which converge to x_*, $\lim x_k = x_*$. Moreover let the matrices B_k be such that for all $k = 0, 1, \ldots$, $Z_k^T B_k Z_k$ is nonsingular, where Z_k is any matrix whose columns provide an orthogonal basis of the nullspace of $Dg(x_k)$. Then the sequence $\{x_k\}$ converges Q-superlinearly to x_*,*

$$\lim_k \|x_{k+1} - x_*\| / \|x_k - x_*\| = 0, \tag{3.6}$$

if and only if

$$\lim_k \|Z_k^T (B_k - W_*) s_k\| / \|s_k\| = 0. \tag{3.7}$$

Remarks: 1. Note that the hypotheses and the conclusion of the theorem do not depend of the choice of the matrices Z_k. Because of the assumption $\lim x_k = x_*$, we may therefore assume that for all $k \geq 0$

$$\|x_k - x_*\| \leq \epsilon_1$$

and $Z_k = Z(x_k)$, where the differentiable matrix function $Z(x)$ and ϵ_1 are as in Lemma 1.

2. The matrix $P_k := Z_k Z_k^T$ is the orthogonal projection of the nullspace of $Dg(x_k)$, which does not depend on the choice of Z_k; a condition in terms of P_k, which is equivalent to (3.7), is

$$\lim_k \|P_k(B_k - W_*)s_k\| / \|s_k\| = 0. \tag{3.7$'$}$$

Proof: According to the above remarks we may assume

$$\|x_k - x_*\| \leq \epsilon_1, \quad Z_k := Z(x_k)$$

for all $k \geq 0$. By (2.4), (2.6), s_k is the solution of

$$\begin{pmatrix} Z_k^T B_k \\ Dg(x_k) \end{pmatrix} s_k = -r(x_k), \tag{3.8}$$

where the function

$$r(x) := \begin{pmatrix} Z(x)^T \nabla_x L(x, u_*) \\ g(x) \end{pmatrix} \equiv \begin{pmatrix} Z(x)^T \nabla f(x) \\ g(x) \end{pmatrix}$$

is differentiable for all $\|x - x_*\| \leq \epsilon_1$ according to Lemma 1 and has the root x_*, $r(x_*) = 0$. Its derivative $Dr(x)$ is locally Lipschitz continuous at x_*, and $Dr(x_*)$ is given by

$$Dr(x_*) = \begin{pmatrix} Z_*^T W_* \\ Dg(x_*) \end{pmatrix}. \tag{3.9}$$

Note also that (3.8) is solvable for s_k, as $Z_k^T B_k Z_k$ is nonsingular by hypothesis.

Therefore, a well-known result of Dennis and Moré [2] characterizing the superlinear convergence of quasi-Newton methods, shows thas the Q-superlinear convergence of $\{x_k\}$ towards x_* is equivalent to

$$\lim_k \left\| \left[\begin{pmatrix} Z_k^T B_k \\ Dg(x_k) \end{pmatrix} - Dr(x_*) \right] s_k \right\| / \|s_k\| = 0.$$

Because of (3.9) the latter condition holds iff

$$\lim_k \left\| \begin{pmatrix} Z_k^T B_k - Z_*^T W_* \\ Dg(x_k) - Dg(x_*) \end{pmatrix} s_k \right\| / \|s_k\| = 0,$$

which by $\lim x_k = x_*$, $\lim Z_k = Z_*$, is equivalent to (3.7). (This simple argument was essentially used by Nocedal and Overton [7], however their proof of (3.9) was not given directly, but was based on a complicated result of Goodman [5]. Another direct proof of theorem 3 which avoids the use of the not uniquely defined differentiable functions $Q(x) = (Y(x), Z(x)$ is given in [9].)

Our results on the iteration function $\Phi_k(x)$ (3.1), in particular the explicit formulae (3.5) for $D\Phi(x_*)$ also permit a simple proof of another central result on the convergence of SQP-methods, which was shown by Powell [8]. Within this theorem, Z_k again denotes any matrix, whose columns form an orthogonal basis of the nullspace of $Dg(x_k)$, and $P_k := Z_k Z_k^T$ denotes the corresponding orthogonal projection on this nullspace.

Theorem 4. *Assume that x_* is a K.-T. point of (P) for which assumption (A) and the 2nd order sufficient condition (1.1) holds, so that x_* is a local minimiser of (P). Suppose further that the SQP-method (1.3),(1.4) with $\lambda_k \equiv 1$ belonging to the matrices $B_k = B_k^T$ generates vectors x_k, $k = 0, 1, \ldots$, that converge to x_*. Moreover , let the matrices B_k satisfy*

$$\|B_k\|, \quad \|(Z_k^T B_k Z_k)^{-1}\| \le c \tag{3.10}$$

for all $k = 0, 1, \ldots$.
If in addition, the condition

$$\lim_k \left\| P_k(B_k - W_*)P_k s_k \right\| / \|s_k\| = 0 \tag{3.11}$$

holds, then the sequence $\{x_k\}$ is 2-step Q-superlinearly convergent,

$$\lim_k \|x_{k+1} - x_*\| / \|x_{k-1} - x_*\| = 0. \tag{3.12}$$

Proof: We note again that the hypothesis (3.10) of the theorem does not depend on the particular choice of the matrices Z_k. Therefore and because of $\lim x_k = x_*$ and (A) we may assume without loss of generality that for all $k \ge 0$

$$\|x_k - x*\| \le \epsilon, \quad \text{and} \quad Z_k := Z(x_k),$$

where $\epsilon = \epsilon(c)$ and the differentiable matrix function $Z(x)$ are as in Lemma 2. In the sequel we use the notation $a_k = O(\alpha_k)$ if there is a constant $C = C(c)$ only depending on (P) and the bound c of this theorem, but not on the B_k such that

$$\|a_k\| \le C|\alpha_k| \quad \text{for all} \quad k \ge 0.$$

Now, by (3.10) and the other hypotheses of the theorem, Lemma 2 applies showing that for all $k \ge 0$, $\Phi_k(x)$ is continuously differentiable for $\|x - x_*\| \le \epsilon$ and that

$$D\Phi_k(x) = O(\|x - x_*\|) \quad \text{for} \quad \|x - x_*\| \le \epsilon.$$

For the errors $e_k := x_k - x_*$ we therefore obtain

$$\begin{aligned} e_{k+1} &= \Phi_k(x_k) - \Phi_k(x_*) = \int_0^1 D\Phi_k(x_* + te_k)e_k\, dt \\ &= D\Phi_k(x_*)e_k + O(\|e_k\|^2) \\ &= -D\Phi_k(x_*)s_k + D\Phi_k(x_*)e_{k+1} + O(\|e_k\|^2). \end{aligned} \tag{3.13}$$

By (3.10) and (3.5), $\left(I - D\Phi_k(x_*)\right)^{-1}$ exists and is bounded for $k \ge 0$:

$$\left(I - D\Phi_k(x_*)\right)^{-1} = O(1).$$

Therefore by (3.13)

$$\begin{aligned} e_{k+1} &= -\left(I - D\Phi_k(x_*)\right)^{-1} D\Phi_k(x_*)s_k + O(\|e_k\|^2) \\ &= O(\|D\Phi_k(x_*)s_k\|) + O\|e_k\|^2). \end{aligned} \tag{3.14}$$

On the other hand, (3.5) gives by (3.10)

$$\|D\Phi_k(x_*)s_k\| = O(\|Y_*^T s_k\|) + O(\|Z_*^T(B_k - W_*)Z_* Z_*^T s_k\|). \tag{3.15}$$

But

$$\begin{aligned} s_k &= \Phi_k(x_k) - x_k = \left(D\Phi_k(x_*) - I\right)e_k + O(\|e_k\|^2) \\ &= \left(D\Phi_k(x_*) - I\right)D\Phi_{k-1}(x_*)e_{k-1} + O(\|e_{k-1}\|^2), \end{aligned}$$

which shows

$$s_k = O(\|e_{k-1}\|)$$

and also by (3.5)

$$Y_*^T s_k = 0 + O(\|e_{k-1}\|^2),$$

so that (3.14), (3.15) imply

$$\|e_{k+1}\|/\|e_{k-1}\| = O(\|e_{k-1}\|) + O(\|Z_*^T(B_k - W_*)Z_* Z_*^T s_k\|/\|s_k\|).$$

As $\lim x_k = x_*$, $\lim Z_k = Z_*$, condition (3.11) of the theorem, which is equivalent to

$$\lim_k \|(Z_k^T B_k Z_k - Z_k^T W_* Z_k) Z_k^T s_k\|/\|s_k\| = 0,$$

therefore gives the final result

$$\lim_k \|e_{k+1}\|/\|e_{k-1}\| = 0,$$

which was to be proved.

References

1. Boggs, P.T., Tolle, J.W., Wang, P.: On the local convergence of quasi- Newton methods for constrained optimization. *SIAM J. Control and Optimization* **20**, 161–171 (1982)
2. Dennis, J.E.Jr., Moré, J.J.: A characterization of superlinear convergence and its application to quasi-Newton methods. *Mathematics of Computation* **28**, 549–560 (1974)
3. Coleman, T.F., Conn, A.R.: On the local convergence of a quasi-Newton method for the nonlinear programming problem. *SIAM J. Numerical Analysis* **21**, 755–769 (1984).
4. Fontecilla, R., Steihaug, T., Tapia, R.A.: 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. Numerical Analysis*.
5. Goodman. J.: Newton's method for constrained optimization. To appear in *Math. Programming*.
6. Han, S.P.:Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Math. Programming* **11**, 263–282 (1976)
7. Nocedal, J., Overton, M.: Projected Hessian updating algorithms for nonlinearly constrained optimization. *SIAM J. Numerical Analysis* **22**, 821–850 (1985)
8. Powell, M.J.D.: The convergence of variable metric methods for nonlinearly constrained optimization problems. In: Nonlinear Programming 3. Mangasarian, O.L., Meyer, R.R., Robinson, S.M.(eds.). pp.27–63, New York: Academic Press 1978
9. Stoer, J., Tapia, R.A.: On the characterization of Q-superlinear convergence of quasi-Newton methods for constrained optimization. To appear in *Mathematics of Computation*.
10. Tapia, R.A.: Diagonalized multiplier methods and quasi-Newton methods for constrained optimization. *Journal of Optimization Theory and Applications* **22**, 135–194 (1977)
11. Tapia, R.A.: Quasi-Newton methods for equality constrained optimization: Equivalence of existing methods and a new implementation. In: Nonlinear Programming 3. Mangasarian, O.L., Meyer, R.R., Robinson, S.M.(eds.). pp.125–163, New York: Academic Press 1978