

On Secant Updates for Use in General Constrained Optimization

By Richard Tapia*

Abstract. In this paper we present two new classes of successive quadratic programming (SQP) secant methods for the equality-constrained optimization problem. One class of methods uses the SQP augmented Lagrangian formulation, while the other class uses the SQP Lagrangian formulation. We demonstrate, under the standard assumptions, that in both cases the BFGS and DFP versions of the algorithm are locally q -superlinearly convergent. To our knowledge this is the first time that either local or q -superlinear convergence has been established for an SQP Lagrangian secant method which uses either the BFGS or DFP updating philosophy and assumes no more than the standard assumptions. Since the standard assumptions do not require positive definiteness of the Hessian of the Lagrangian at the solution, it is no surprise that our BFGS and DFP updates possess the hereditary positive definiteness property only on a proper subspace.

1. Introduction. By a *successive quadratic programming (SQP) quasi-Newton method* for the constrained optimization problem

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

where $f: R^n \rightarrow R$ and $g: R^n \rightarrow R^m$ ($m \leq n$), we mean the iterative process

$$(1.2) \quad \begin{aligned} x_+ &= x + s, \\ \lambda_+ &= \lambda + \Delta\lambda, \\ B_l^+ &= \mathbf{B}(x, s, \lambda_+, B_l), \end{aligned}$$

where \mathbf{B} is an update function and s and $\Delta\lambda$ are respectively the solution and the multiplier associated with the solution of the quadratic program

$$(1.3) \quad \begin{aligned} & \text{minimize } \nabla_x l(x, \lambda)^T s + \frac{1}{2} s^T B_l s \\ & \text{subject to } \nabla g(x)^T s + g(x) = 0. \end{aligned}$$

In (1.3), $\nabla_x l(x, \lambda)$ is the gradient (with respect to x) of the Lagrangian

$$(1.4) \quad l(x, \lambda) = f(x) + \lambda^T g(x)$$

evaluated at the current iterate (x, λ) . In (1.3), B_l is intended to be an approximation to $\nabla_x^2 l(x, \lambda)$, the Hessian of the Lagrangian at (x, λ) . We denote by x_* a solution of problem (1.1) with associated multiplier λ_* , i.e., λ_* is such that $\nabla_x l(x_*, \lambda_*) = 0$.

Received October 9, 1984; revised November 24, 1986.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 49D15, 65K05.

Key words and phrases. Quasi-Newton, nonlinear programming, superlinear convergence.

*Research sponsored by DOE DE-FG05-86ER25017, ARO DAAG-03-86-K-0113, and AFOSR 85-0243.

We call (1.2)–(1.3) an *SQP Lagrangian secant method* if B_i^+ satisfies the *Lagrangian secant equation*

$$(1.5) \quad B_i^+ s = y_i,$$

where

$$(1.6) \quad y_i = \nabla_x l(x_+, \lambda_+) - \nabla_x l(x, \lambda_+)$$

and x , x_+ , s and λ_+ are as in (1.2)–(1.3). See Algorithm A1 of Appendix A.

It is common practice to use $\nabla f(x)$ instead of $\nabla_x l(x, \lambda)$ in (1.3). With this change, λ_+ and not $\Delta\lambda$ will be the multiplier associated with the solution of the quadratic program. The use of $\nabla_x l(x, \lambda)$ provides a more convenient framework for our current investigations. Strictly speaking, we should probably refer to λ as the vector of multipliers or the multiplier vector; however, we will take the liberty of merely using the term multiplier.

We are interested in the BFGS and the DFP secant updates. However, we feel that our understanding will be enhanced if we parallel our considerations with an update which (at least in theory) is effective without the assumption that $\nabla_x^2 l(x_*, \lambda_*)$ is positive definite. Toward this end, the obvious choice is the PSB secant update.

Consider the following class of secant updates from unconstrained optimization that can be used with the SQP secant method (1.2)–(1.6):

$$(1.7a) \quad B_+ = B + SECANT(s, y, B, v),$$

where

$$(1.7b) \quad SECANT(s, y, B, v) = \frac{(y - Bs)v^T + v(y - Bs)^T}{v^T s} - \frac{(y - Bs)^T s v v^T}{(v^T s)^2}.$$

Following Dennis and Walker [9], we call v the *scale of the secant update* (1.7). The following are well-known choices of scale:

$$(1.8) \quad PSB: v = s$$

$$(1.9) \quad DFP: v = y$$

$$(1.10) \quad BFGS: v = y + \sigma Bs, \quad \sigma = \sqrt{\frac{y^T s}{s^T Bs}}.$$

For further details, see Dennis and Moré [7] (in particular Theorem 7.3) and also Dennis and Schnabel [8]. The scale v is often a function of s , y or B , as is the case in (1.8)–(1.10). When the need arises, we will write $v(s, y, B)$. Also, the notation $SECANT(s, y, B, v)$ implies that the values of the arguments of v are s , y , and B , respectively, i.e., $v = v(s, y, B)$. If other argument values are desired, or if we wish to emphasize the arguments used, then we will use different notation.

Much is known about the theoretical and numerical properties of these updates when used in the context of unconstrained optimization. For details, see Dennis and Moré [7], Fletcher [10] and Dennis and Schnabel [8]. It is known that all three updates give local q -superlinear convergence. The BFGS and the DFP updates require a symmetric and positive definite Hessian, while the PSB update only requires symmetry and invertibility of the Hessian. It is generally felt that the BFGS has superior numerical properties, and it is the preferred update for use in unconstrained optimization.

The following are *standard assumptions* in the theory of quasi-Newton methods for problem (1.1):

- A1: $f, g_i, i = 1, \dots, m$, have second derivatives which are Lipschitz continuous in an open neighborhood Ω of the local solution x_* .
- A2: $\nabla^2 l(x_*, \lambda_*)$ is nonsingular.

A direct calculation shows that

$$\nabla^2 l(x_*, \lambda_*) = \begin{pmatrix} \nabla_x^2 l(x_*, \lambda_*) & \nabla g(x_*) \\ \nabla g(x_*)^T & 0 \end{pmatrix}.$$

Moreover, it is well known that Assumption A2 is equivalent to the two assumptions

- A2'a: $\nabla g(x_*)$ has full rank;
- A2'b: $\nabla_x^2 l(x_*, \lambda_*)$ is positive definite on $S(x_*) = \{\eta : \nabla g(x_*)^T \eta = 0\}$.

Under these standard assumptions the local q -superlinear convergence in (x, λ) of the SQP PSB secant method was established by Han [15] and Tapia [19]. Tapia included the Broyden update and Han [15], [16] considered other updates.

From the statements made above, it follows that the use of the BFGS and DFP updates is precluded unless one makes the stronger assumption that $\nabla_x^2 l(x_*, \lambda_*)$ is positive definite on the entire space. This unfortunate circumstance has been the subject of numerous research papers, including Powell [18], Coleman and Conn [6], Fontecilla [11], Nocedal and Overton [17], Bartholomew-Biggs [2], and the present work. Assuming the positive definiteness of $\nabla_x^2 l(x_*, \lambda_*)$, Han [15], [16] established the local q -superlinear convergence in (x, λ) of the SQP Lagrangian BFGS and DFP secant methods. Under this same positive definiteness assumption Boggs, Tolle and Wang [4] and Fontecilla, Steihaug and Tapia [12] established the local q -superlinear convergence in x of the SQP Lagrangian BFGS and DFP secant methods.

Working with the SQP Lagrangian BFGS secant method, Powell [18] proposed modifying y_l in (1.6) in the following manner

$$(1.11a) \quad \hat{y}_l = \theta y_l + (1 - \theta) B_l s,$$

where

$$(1.11b) \quad \theta = \begin{cases} 1 & \text{if } y_l^T s \geq \varepsilon s^T B_l s, \\ \frac{(1 - \varepsilon) s^T B_l s}{s^T B_l s - y_l^T s} & \text{otherwise} \end{cases}$$

and ε is a small positive constant, e.g., $\varepsilon = 0.2$. The update B_l^+ is then obtained as a BFGS secant update using \hat{y}_l instead of y_l . This modification guarantees that $\hat{y}_l^T s > 0$ and therefore allows the BFGS secant update to maintain hereditary positive definiteness; however, the Lagrangian secant equation (1.5) is lost. Assuming that the iteration sequence $\{x_k\}$ generated by the modified algorithm converged, Powell was able to show that the convergence was r -superlinear. Bartholomew-Biggs [2] suggested various changes to the Powell modification and presented some numerical experimentation.

Recently, Boggs and Tolle [3] demonstrated that if in the SQP Lagrangian BFGS or DFP secant methods, x_k converged q -linearly to x_* , λ_k converged q -linearly to λ_* , and the convergence of x_k to x_* satisfied an additional condition (tangential

convergence), then the convergence of x_k to x_* would be q -superlinear even if $\nabla_x^2 l(x_*, \lambda_*)$ is not positive definite.

An often-used alternative to the restrictive assumption of requiring $\nabla_x^2 l(x_*, \lambda_*)$ to be positive definite is to work with the augmented Lagrangian

$$(1.12) \quad L(x, \lambda) = l(x, \lambda) + \frac{1}{2} C g(x)^T g(x) \quad (C \geq 0)$$

instead of the Lagrangian $l(x, \lambda)$ given by (1.4). In this case, B_l in (1.3) is replaced with B_L , an approximation to $\nabla_x^2 L(x, \lambda)$, and $\nabla_x l$ in (1.3) and (1.6) is replaced with $\nabla_x L$. This approach leads to what we will call the *SQP augmented Lagrangian quasi-Newton methods* (see Algorithm A2 of Appendix A). The motivation for this consideration comes from the fact that under the standard assumptions, for sufficiently large C , $\nabla_x^2 L(x_*, \lambda_*)$ is positive definite (see Corollary 12.9 and Theorem 12.10 of Avriel [1]) and the BFGS and DFP secant updates can be used in (1.2). Local q -superlinear convergence in (x, λ) of the SQP augmented Lagrangian BFGS and DFP secant methods was established by Han [15], Tapia [19] and Glad [14]. Local q -superlinear convergence in x was established by Fontecilla, Steihaug and Tapia [12].

The SQP augmented Lagrangian quasi-Newton methods have the serious drawback of having to choose C in the absence of adequate guidelines. Moreover, large C seems to present severe numerical problems; see the examples given in Tapia [19] and Nocedal and Overton [17]. We feel that the various components in this approach do not fit together well, and this bad fit could account for some of the numerical difficulties. The present work is concerned with effectively combining the theoretical advantages of the augmented Lagrangian path with the BFGS or DFP secant update in order to obtain a better fit of the two components.

In Section 2 we recall two quasi-Newton formulations for problem (1.1) which are equivalent to the SQP augmented Lagrangian quasi-Newton formulation. These equivalent formulations will be used in other sections. In Section 3 we take a very close look at the structure present in the Hessian of the augmented Lagrangian and make several observations concerning the role of the penalty constant.

In Section 4 we utilize structure to derive what we call the SQP augmented scale secant methods. These algorithms are SQP Lagrangian secant methods (as opposed to SQP augmented Lagrangian secant methods) in that the subproblem is of the form (1.3) and the Hessian approximation satisfies the Lagrangian secant equation (1.5). We will see that a complete use of structure leads to an algorithm where only the scale in the secant update is affected. This change of scale is important for the BFGS and DFP updates and will not be active for the PSB update (where it is not needed). In the case of the BFGS and DFP updates, the updates will possess the hereditary positive definiteness property only on an appropriate proper subspace. However, the change of scale allows us to view the augmented scale BFGS and DFP updates as least-change secant updates, even when $y_l^T s$ is negative.

In Section 5 we utilize structure to present the SQP structured augmented Lagrangian secant methods. These algorithms possess the flavor of the modified BFGS algorithm which Powell [18] proposed. Specifically, structure leads us to suggest a modification of y_l so that the BFGS and DFP updates will possess the hereditary positive definiteness property. However, instead of sacrificing the Lagrangian secant

equation we replace it with a structured augmented Lagrangian secant equation. It is this latter consideration which will allow us to establish local q -superlinear convergence in Section 7.

In Section 6 we consider the effect of letting the penalty constant become arbitrarily large in the SQP augmented scale BFGS and DFP methods. Both these updates give the same limit update. It follows that any update which is a member of the Broyden convex class will produce the same limit update. The role of this limit update will be discussed in Section 8.

In Section 7 we demonstrate that both the SQP augmented scale BFGS and DFP secant methods and the SQP structured augmented Lagrangian BFGS and DFP secant methods are locally q -superlinearly convergent. In Section 8 we offer a summary of our results and some concluding remarks. For purposes of convenience and reference we offer Appendix A, where the various SQP secant methods discussed in the paper are collected and catalogued; Appendix B contains justification for the statement made in Section 3 that a particular second-order term is responsible for the bad fit in the SQP augmented Lagrangian quasi-Newton methods and consequently should be left out of any structured approximations.

2. Equivalent Formulations. The material in this section is taken from Tapia [20]. In addition to the three equivalent formulations which we are about to describe, several other interesting formulations can be found in that paper.

Suppose that we are given an iterate (x, λ) as an approximation to (x_*, λ_*) (the solution of problem (1.1) and its associated multiplier) and a positive definite matrix B_L as an approximation to $\nabla_x^2 L(x, \lambda)$. We now present three equivalent quasi-Newton formulations for constructing the subsequent iterate (x_+, λ_+) .

SQP Formulation. Recall that the SQP augmented Lagrangian quasi-Newton method consists of determining $(x_+, \lambda_+) = (x + s, \lambda + \Delta\lambda)$ by choosing s and $\Delta\lambda$ to be the solution and the multiplier associated with the solution of the quadratic program

$$(2.1) \quad \begin{aligned} & \text{minimize } \nabla_x L(x, \lambda)^T s + \frac{1}{2} s^T B_L s \\ & \text{subject to } \nabla g(x)^T s + g(x) = 0. \end{aligned}$$

Extended System Formulation. If we apply the first-order necessary conditions to the quadratic programs (2.1), we see that the SQP quasi-Newton step s and its associated multiplier λ_+ can be obtained from the following linear system:

$$(2.2) \quad \begin{pmatrix} B_L & \nabla g(x) \\ \nabla g(x)^T & 0 \end{pmatrix} \begin{pmatrix} s \\ \lambda_+ \end{pmatrix} = - \begin{pmatrix} \nabla f(x) + \nabla g(x) C g(x) \\ g(x) \end{pmatrix}.$$

Since (2.1) is a convex program, the first-order necessary conditions are also sufficient conditions, and (2.1) and (2.2) will have the same solutions (which by strict convexity are necessarily unique).

Multiplier Update Formulation. A straightforward calculation can be used to show that (x_+, λ_+) obtained via (2.2) can be written as

$$(2.3a) \quad \lambda_+ = (\nabla g(x)^T B_L^{-1} \nabla g(x))^{-1} (g(x) - \nabla g(x)^T B_L^{-1} (\nabla f(x) + \nabla g(x) C g(x)))$$

and

$$(2.3b) \quad x_+ = x - B_L^{-1} \nabla_x L(x, \lambda_+).$$

In Tapia [19] and Fontecilla, Steihaug and Tapia [12], (2.3) is referred to as the *diagonalized multiplier method*. In Tapia [20] and in the present work, (2.3) is called the *multiplier update method*. We prefer this latter terminology since this is the only formulation where the multiplier update is given explicitly; moreover, the term “diagonalized” has often been misunderstood.

3. Structure in the Hessian of the Augmented Lagrangian. The objective of a secant method for constrained optimization is to obtain an effective algorithm without calculating second-order information. As has been continually demonstrated (see Tapia [19], Tapia [20] and Nocedal and Overton [17] for example), it is also advisable to take advantage of structure, whenever possible, and not approximate first-order information—especially first-order information which will be calculated in another part of the algorithm. These comments serve to motivate the investigation conducted in this section and the steps taken in Section 4. Throughout the remainder of this paper, when we use S as a superscript or a subscript, it will signify that structure was utilized in forming the quantity in question.

We will be using quantities defined in the previous section. Observe that

$$(3.1) \quad \nabla_x^2 L(x, \lambda) = \nabla_x^2 l(x, \lambda) + Cg(x) \cdot \nabla^2 g(x) + C\nabla g(x)\nabla g(x)^T$$

and

$$(3.2) \quad \nabla_x^2 L(x_*, \lambda_*) = \nabla_x^2 l(x_*, \lambda_*) + C\nabla g(x_*)\nabla g(x_*)^T.$$

In (3.1) we use $g(x) \cdot \nabla^2 g(x)$ to mean $g_1(x)\nabla^2 g_1(x) + \cdots + g_m(x)\nabla^2 g_m(x)$. The structure in (3.1) and (3.2) is readily apparent. However, it is not at all clear whether an approximation of $\nabla_x^2 L(x, \lambda)$ should take into account the term $Cg(x) \cdot \nabla^2 g(x)$ which appears in (3.1), but which does not appear in (3.2) (i.e., vanishes at the solution). The standard (unstructured for the purposes of this discussion) SQP augmented Lagrangian secant methods, described in the previous section, takes this term into consideration. To see this, merely recall that the y_L used in the secant update is defined to be the difference of gradients $\nabla_x L(x_+, \lambda_+) - \nabla_x L(x, \lambda_+)$. It is our considered opinion that every effort should be made to construct approximations which ignore the term $Cg(x) \cdot \nabla^2 g(x)$. The arguments justifying this opinion are not a main part of the current research and are therefore relegated to Appendix B.

The following proposition concerning structure was given by Tapia [19]. We have slightly reworded it to fit our present needs.

PROPOSITION 3.1. *Suppose that B_l is positive definite on $S(x) = \{\eta: \nabla g(x)^T \eta = 0\}$. Then the SQP augmented Lagrangian quasi-Newton method which uses*

$$(3.3) \quad B_L^S = B_l + C\nabla g(x)\nabla g(x)^T$$

as an approximation to $\nabla_x^2 L(x, \lambda)$ generates the same iterate as the SQP Lagrangian quasi-Newton method that uses B_l (from (3.3)) as an approximation to $\nabla_x^2 l(x, \lambda)$.

Proof. The positive definiteness assumption guarantees that both iterates exist. The proof now follows from considering (2.2) and observing that the contribution

of the form

$$(3.4) \quad C\nabla g(x)(\nabla g(x)^T s + g(x))$$

must be zero. \square

In conclusion, if we remove $g(x) \cdot \nabla^2 g(x)$ from the SQP augmented Lagrangian formulation, then Proposition 3.1 seemingly removes any further dependence on C , and it appears as if structure effectively reduces the role of the augmented Lagrangian to that of the Lagrangian. In the following section we will see that this situation is quite subtle, and the structure in the augmented Lagrangian can be utilized in a most effective manner.

4. The Augmented Scale Secant Update. Basically, what the reader should expect in this section is that (even though Proposition 3.1 has led us to believe that structure removes dependence on the penalty constant) the secant update will not be necessarily independent of the penalty constant. The penalty constant will have the effect of changing the scale in the secant update (see (1.7)). This change of scale seems to be important for updates which in unconstrained optimization require a positive definite Hessian, e.g., the BFGS and DFP updates, and seems not to change the scale in updates which do not require a positive definite Hessian, e.g., the PSB update.

Suppose we decide to use a secant method for problem (1.1). If we use the Lagrangian formulation (see Algorithm A1 of Appendix A), the choice for y in (1.7) is

$$(4.1) \quad y_l = \nabla_x l(x_+, \lambda_+) - \nabla_x l(x, \lambda_+);$$

while if we use the augmented Lagrangian formulation (see Algorithm A2 of Appendix A), the choice for y is

$$(4.2) \quad y_L = \nabla_x L(x_+, \lambda_+) - \nabla_x L(x, \lambda_+),$$

which can be rewritten as

$$(4.3) \quad y_l + C(\nabla g(x_+)g(x_+) - \nabla g(x)g(x)).$$

By subtracting and adding $C\nabla g(x_+)g(x)$ we can write (4.3) as

$$(4.4) \quad y_l + C[\nabla g(x_+)(g(x_+) - g(x)) + (\nabla g(x_+) - \nabla g(x))g(x)].$$

By performing two obvious one-term Taylor expansions in (4.4) we interpret (4.4) as an approximation to

$$(4.5) \quad y_l + C[\nabla g(x_+)\nabla g(x_+)^T s + g(x) \cdot \nabla g^2(x_+)s].$$

Taking into account the comments made in Section 3, we are led to define

$$(4.6) \quad y_L^S = y_l + C\nabla g(x_+)\nabla g(x_+)^T s$$

as our structured form of (4.2). Suppose that we have obtained s , x_+ and λ_+ from the SQP Lagrangian quasi-Newton method using B_l as an approximation to $\nabla_x^2 l(x, \lambda)$. Then from Section 3 we have

$$(4.7) \quad B_L^S = B_l + C\nabla g(x_+)\nabla g(x_+)^T$$

as a structured approximation to $\nabla_x^2 L(x, \lambda)$ and may consider the secant update of B_L^S :

$$(4.8) \quad (B_L^S)^+ = B_L^S + SECANT(s, y_L^S, B_L^S, v_L),$$

where $v_L = v(s, y_L^S, B_L^S)$ and v is the scale of the particular update in question (see (1.7)).

Proposition 3.1 says that if we now use $(B_L^S)^+$ in the SQP augmented Lagrangian method, we are doing no more and no less than using the SQP Lagrangian method with

$$(4.9) \quad B_l^+ = B_l + SECANT(s, y_L^S, B_L^S, v_L).$$

The following proposition shows that (4.9) is merely a Lagrangian secant update. Consequently, a complete use of structure takes an SQP augmented Lagrangian secant method and reduces it to an SQP Lagrangian secant method.

PROPOSITION 4.1. *If y_l, y_L^S, B_l and B_L^S are as above, then*

$$(4.10) \quad SECANT(s, y_L^S, B_L^S) = SECANT(s, y_l, B_l, v_L).$$

Consequently, the update (4.9) can be written as

$$(4.11) \quad B_l^+ = B_l + SECANT(s, y_l, B_l, v_L)$$

and satisfies the Lagrangian secant equation

$$B_l^+ s = y_l.$$

Proof. The proof is straightforward. \square

We call (4.11) an *augmented scale secant update* and the corresponding SQP Lagrangian secant method an *SQP Lagrangian augmented scale secant method* (see Algorithm A3 of Appendix A). In the case of the BFGS and DFP updates it is this change of scale that compensates for the fact that $\nabla_x^2 l(x_*, \lambda_*)$ is not necessarily positive definite. For this reason we should choose C so that $s^T y_L^S > 0$. It will then follow that even though $y_l^T s$ may be negative, the update will be a least-change secant update (see Theorem 7.3 of Dennis and Moré [7]).

We now describe two situations in which the augmented scale update will be independent of C ; but in both of these cases the augmentation is not needed.

PROPOSITION 4.2. *The augmented scale PSB secant update is independent of the penalty constant C and therefore coincides with the usual PSB secant update.*

Proof. The result follows directly from the fact that the scale v is equal to s . \square

Proposition 4.2 is satisfying since it says that the augmentation that was introduced to compensate for the lack of positive definiteness is not needed. Indeed, we knew that the PSB update enjoyed this luxury.

PROPOSITION 4.3. *If $\nabla g(x_+)^T s = 0$, then the augmented scale BFGS and DFP secant updates are independent of the penalty constant C and therefore coincide with the nonaugmented forms of the updates.*

Proof. The proof is straightforward. \square

Again, the situation in Proposition 4.3 is as it should be. Specifically, Assumption A2 says that $\nabla_x^2 l(x_*, \lambda_*)$ is positive definite on $S(x_*)$, where

$$(4.12) \quad S(x) = \{\eta : \nabla g(x)^T \eta = 0\}.$$

Consequently, for x and x_+ near x_* , and s such that $\nabla g(x_+)^T s = 0$, we can expect $y_i^T s > 0$ and the augmentation which was introduced to compensate for the lack of positive definiteness is not needed. In the case of linear constraints we will always have $\nabla g(x_+)^T s = 0$. However, it is well known that in this case the SQP Lagrangian BFGS and DFP secant methods exhibit the proper theoretical and numerical properties since the iterates are always contained in $S(x_*)$.

Recall that the *Broyden convex class of secant updates* consists of all updates which are a convex combination of the BFGS secant update and the DFP secant update; see Dennis and Schnabel [8] or Fletcher [10]. It is well known in the theory of unconstrained optimization that any update in the Broyden convex class possesses the hereditary positive definiteness property, i.e., if B is positive definite, $y^T s > 0$ and v corresponds to the scale from a member of the Broyden convex class, then B_+ as given by (1.7) is positive definite.

In the case of the augmented scale secant update for constrained optimization we have the following appropriate analog of hereditary positive definiteness. Recall that $S(x)$ has been defined in (4.12).

PROPOSITION 4.4. *Suppose that s, x_+, y_l and B_l are such that B_l is positive definite on $S(x_+)$ and $y_l^T s > 0$ if $s \in S(x_+)$. Then for all C sufficiently large, B_l^+ , the augmented scale secant update of B_l obtained from (4.11), using any update from the Broyden convex class, is also positive definite on $S(x_+)$.*

Proof. Choose C so large that B_L^S given in (4.7) is positive definite and $s^T y_L^S > 0$, where y_L^S is given by (4.6). It follows from the hereditary positive definiteness property enjoyed by secant updates from the Broyden convex class that $(B_L^S)^+$ as given by (4.8) will be positive definite. Substituting (4.10) and (4.11) into (4.8) allows us to write

$$(4.13) \quad (B_L^S)^+ - B_l^+ = C \nabla g(x_+) \nabla g(x_+)^T.$$

If $\eta \in S(x_+)$, then $\nabla g(x_+)^T \eta = 0$; hence $(B_L^S)^+$ and B_l^+ coincide on $S(x_+)$. It follows that B_l^+ is positive definite on $S(x_+)$. \square

We would like to make some comments concerning our way of motivating the choice of y_L^S given by (4.6). A short and direct argument for this choice would be as follows. An obvious implication of (3.2) is to choose B_L^S given by (4.7) as a structured Hessian approximation. It then readily follows from (4.7) that the choice for y_L^S should be (4.6). However, in the following section we will employ structure in our choice for y_L^S , but not in our choice for the Hessian approximation. Hence, an argument which does not use the choice of B_L^S to motivate (4.6) as the choice for y_L^S is probably of value.

In summary, then, it seems as if the complete use of structure has allowed us to identify and eliminate the undesirable by-products given by the SQP augmented Lagrangian secant formulation and pose an SQP Lagrangian secant formulation which maintains only the key desirable feature of a change of scale in the secant update.

5. The Structured Augmented Lagrangian Secant Update. It may be that in the SQP augmented Lagrangian secant method (Algorithm A2 of Appendix A) we wish to introduce structure in y_L in (A.6) but not in B_L . Toward this end, suppose that we have obtained s , x_+ and λ_+ from the SQP augmented Lagrangian secant method, using B_L as an approximation to $\nabla_x^2 L(x, \lambda)$. We may consider the structured augmented Lagrangian update of B_L ,

$$(5.1) \quad B_L^+ = B_L + SECANT(s, y_L^S, B_L, v_L),$$

where y_L^S is given by (4.6) or (A.11), $v_L = v(s, y_L^S, B_L)$ and v is the scale function of the particular secant update in question (see (1.7)). The Broyden convex class of secant updates was introduced in the previous section. See Proposition 4.4 in particular.

PROPOSITION 5.1. *If B_L is positive definite and $C \geq 0$ has been chosen so that $s^T y_L^S > 0$, then the structured augmented Lagrangian secant update of B_L obtained from (5.1), using any update from the Broyden convex class, is also positive definite.*

Proof. This proposition is a consequence of the well-known hereditary positive definiteness property of secant updates from the Broyden convex class (see Dennis and Schnabel [8] or Dennis and Moré [7] for details). \square

The following proposition is also straightforward.

PROPOSITION 5.2. *The structured augmented Lagrangian secant update (5.1) satisfies the structured augmented Lagrangian secant equation*

$$(5.2) \quad B_L^+ s = y_L^S.$$

This algorithm has the flavor of the modified SQP Lagrangian BFGS secant method proposed by Powell (see (1.11)). In both instances, y_l has been modified. However, while Powell loses the Lagrangian secant equation, we replace it with a structured form of the augmented Lagrangian secant equation. It is this consideration which will allow us to prove local q -superlinear convergence in Section 7.

6. Limit Updates. Since the SQP augmented Lagrangian secant method exhibits such poor performance for large C , it is natural to consider the asymptotic properties of the augmented scale secant update and the structured augmented Lagrangian secant update. We will be considering Algorithms A3 and A4 from Appendix A.

If we let $\theta = C/(1+C)$, $y_\theta = (1-\theta)y_l + \theta \nabla g(x_+) \nabla g(x_+)^T s$, $B_\theta = (1-\theta)B_l + \theta \nabla g(x_+) \nabla g(x_+)^T$ and observe that the secant update (1.7) has the property

$$(6.1) \quad SECANT(s, y, B, v) = SECANT(s, y, B, \alpha v)$$

for any $\alpha \neq 0$, then we can write the scale in the augmented scale DFP update (see (1.9) and (A.9)) as

$$(6.2) \quad v_\theta(\text{DFP}) = y_\theta,$$

and we can write the scale in the augmented scale BFGS update (see (1.10) and (A.9)) as

$$(6.3) \quad v_\theta(\text{BFGS}) = y_\theta + \sqrt{\frac{y_\theta^T s}{s^T B_\theta s}} B_\theta s.$$

Observe that $C = 0$ corresponds to $\theta = 0$ and $C = +\infty$ corresponds to $\theta = 1$. Also observe that for $\theta = 1$ we have

$$(6.4) \quad v_\theta(\text{DFP}) = \frac{1}{2}v_\theta(\text{BFGS}) = \nabla g(x_+) \nabla g(x_+)^T s.$$

These observations have led us to the following proposition.

PROPOSITION 6.1. *If $\nabla g(x_+)^T s \neq 0$ and we let $C \rightarrow \infty$ in either the augmented scale DFP secant update or the augmented scale BFGS secant update, then we obtain the limit update*

$$(6.5a) \quad B_l^+ = B_l + \text{SECANT}(s, y_l, B_l, v_*)$$

with

$$(6.5b) \quad v_* = \nabla g(x_+) \nabla g(x_+)^T s.$$

Since the Broyden convex class of updates consists of all updates which are a convex combination of the BFGS update and the DFP update, we have the following corollary.

COROLLARY 6.2. *Any member of the Broyden convex class of updates gives rise to the limit update (6.5).*

PROPOSITION 6.3. *The limit update B_l^+ given by (6.5) satisfies the Lagrangian secant equation $B_l^+ s = y_l$ and gives the following extreme form of hereditary positive definiteness. For $\eta \in S(x_+) = \{\eta \in \mathbb{R}^n : \nabla g(x_+)^T \eta = 0\}$*

$$(6.6) \quad \eta^T B_l^+ \eta = \eta^T B_l \eta.$$

Proof. The proof is straightforward. \square

If we now look at the asymptotic properties of the structured augmented Lagrangian secant updates, then we see from the discussion above that the scale is well-behaved and approaches (6.5b), the scale of the limit update. However, the quantity $y_L^S - B_L s = y_l + C \nabla g(x_+) \nabla g(x_+)^T s - B_L s$ has no limit; so the update has no limit.

7. Local q -Superlinear Convergence. In this section we establish local q -superlinear convergence of the SQP augmented scale Lagrangian BFGS and DFP secant methods and the SQP structured augmented Lagrangian BFGS, DFP and PSB secant methods described in Sections 4 and 5, respectively.

The approach we will take to establish convergence for the structured algorithms can be thought of as a piggyback approach. The foundation of our approach will be the known convergence results for the SQP augmented Lagrangian secant methods (see Corollary 5.5 of Fontecilla, Steihaug and Tapia [12], for example). Specifically, we will derive perturbation bounds for the quantities associated with each of the structured algorithms in terms of the corresponding quantities associated with the well-known (unstructured) SQP augmented Lagrangian secant method. Bounded deterioration for the structured algorithm will then follow from the known bounded deterioration for the unstructured algorithm. Once bounded deterioration has been established, the local q -superlinear convergence of the structured algorithm will follow immediately from the Fontecilla-Steihaug-Tapia convergence theory. The Fontecilla, Steihaug and Tapia [12] convergence theory is cast in the framework

of the multiplier update formulation (see Section 2). Consequently, we will work with the equivalent multiplier update formulation of each of our structured SQP methods. In the case of the structured DFP algorithms, bounded deterioration also follows from the Dennis-Walker theory for structured updates [9].

Recall that the SQP augmented Lagrangian secant method can be found in Appendix A as Algorithm A2, the SQP Lagrangian augmented scale secant method can be found in Appendix A as Algorithm A3, and the SQP structured augmented Lagrangian secant method can be found in Appendix A as Algorithm A4. Also, recall that x_* is a solution of problem (1.1) with associated multiplier λ_* .

Suppose that we have determined x , λ , x_+ and λ_+ . Let

$$(7.1) \quad y_I = \nabla_x l(x_+, \lambda_+) - \nabla_x l(x, \lambda_+),$$

$$(7.2) \quad y_L = \nabla_x L(x_+, \lambda_+) - \nabla_x L(x, \lambda_+),$$

and

$$(7.3) \quad y_L^S = y_I + C \nabla g(x_+) \nabla g(x_+)^T s.$$

In (7.3) the superscript S is used to signify that structure in the Hessian was utilized in forming this approximation. Our convergence analysis will depend on the following important estimates. Let $S(R^{n \times n})$ denote the space of $n \times n$ symmetric matrices.

LEMMA 7.1. *Suppose that the standard assumptions given in Section 1 for problem (1.1) hold and $C \geq 0$ has been chosen so that $\nabla_x^2 L(x_*, \lambda_*) = \nabla_x^2 l(x_*, \lambda_*) + C \nabla g(x_*) \nabla g(x_*)^T$ is positive definite. Then there exist generic positive constants α and β and neighborhoods $N_1 \subset R^n$ of x_* and $N_2 \subset S(R^{n \times n})$ of $\nabla_x^2 L(x_*, \lambda_*)$ such that $N_1 \subset \Omega$ (see A1 of Section 1) and for every $(x, B_L) \in N = N_1 \times N_2$, B_L is positive definite,*

$$(7.4) \quad \|\lambda_+ - \lambda_*\| \leq \beta \|x - x_*\|,$$

where λ_+ is given by (2.3a),

$$(7.5) \quad x_+ = x - B_L^{-1} \nabla_x L(x, \lambda_+) \in \Omega,$$

$$(7.6) \quad \|y_L^S - y_L\| \leq \beta \sigma(x, x_+) \|s\|,$$

$$(7.7) \quad \|v_L^S - v_L\| \leq \beta \|y_L^S - y_L\|,$$

where $v_L^S = v(s, y_L^S, B_L)$ and $v_L = v(s, y_L, B_L)$. Moreover, for ω given by any one of the quantities y_L^S , y_L , v_L^S or v_L , we have

$$(7.8) \quad \alpha \|s\|^2 \leq \omega^T s \leq \beta \|s\|^2$$

and

$$(7.9) \quad \|\omega\| \leq \beta \|s\|,$$

and finally

$$(7.10) \quad \|\text{SECANT}(s, y_L^S, B_L, v) - \text{SECANT}(s, y_L, B_L, v)\| \leq \beta \sigma(x, x_+),$$

where the v used in the definition of any of the quantities appearing in (7.7)–(7.10) is the scale from either the BFGS, DFP or PSB secant update, the choice of norms

is arbitrary (the neighborhoods and constants depend on these choices), as usual $s = x_+ - x$ and

$$(7.11) \quad \sigma(x, x_+) = \max\{\|x - x_*\|, \|x_+ - x_*\|\}.$$

Proof. We will give the proof using the 2-norm for all vector norms and the induced 2-norm for all matrix norms. That arbitrary norms can be used follows from the equivalence of norms in finite-dimensional spaces. The proof will be divided into parts.

Proof of (7.4) and (7.5). The estimate (7.4) follows from observing that (2.3a) allows us to write

$$(7.12) \quad \lambda_+ - \lambda_* = (\nabla g(x)^T B_L^{-1} \nabla g(x))^{-1} \\ \times [(g(x) - g(x_*)) - \nabla g(x)^T B_L^{-1} (\nabla_x L(x, \lambda_*) - \nabla_x L(x_*, \lambda_*))].$$

For further detail, see the proof of Proposition 4.2 in Fontecilla, Steihaug and Tapia [12]. The statement (7.5) can be established with the aid of (7.4).

Proof of (7.6). Combining (4.4) and (4.6) leads to

$$(7.13) \quad y_L - y_L^S = C \nabla g(x_+) [g(x_+) - g(x) - \nabla g(x_+)^T s] + C [\nabla g(x_+) - \nabla g(x)] g(x).$$

The estimate (7.6) now follows from the standard assumptions and (7.13).

Proof of (7.7). In the case of PSB and DFP, (7.7) is trivially satisfied. Before we consider the case of BFGS, we will derive some simple bounds. Let λ_{\min} , respectively λ_{\max} , denote the smallest, respectively largest, eigenvalue of the positive definite matrix $\nabla_x^2 L(x_*, \lambda_*)$ and restrict N_2 , if needed, so that for each $A \in N_2$ we have

$$(7.14) \quad \frac{\lambda_{\min}}{2} s^T s \leq s^T A s \leq 2\lambda_{\max} s^T s.$$

Again, choosing smaller neighborhoods, if necessary, we can assume that

$$\nabla_x^2 L(\theta x_+ + (1 - \theta)x, \lambda_+)$$

satisfies (7.14) whenever $(x, B_L) \in N$ and $\theta \in [0, 1]$. Now, for $(x, B_L) \in N$ we have, using the mean value theorem, since $y_L^T s = s^T \nabla_x^2 L(x + \theta(x_+ - x))s$ for some $\theta \in (0, 1)$, that

$$(7.15) \quad \frac{\lambda_{\min}}{2} s^T s \leq \omega^T s \leq 2\lambda_{\max} s^T s$$

for $\omega = y_L$. Essentially the same argument establishes (7.15) for $\omega = y_L^S$.

The mean value theorem on the function $h(y) = \sqrt{s^T y}$ combined with (7.15) for $\omega = y_L$ and $\omega = y_L^S$ gives

$$(7.16) \quad |\sqrt{s^T y_L} - \sqrt{s^T y_L^S}| \leq \|y_L - y_L^S\| / \sqrt{2\lambda_{\min}}.$$

By combining (7.14) and (7.16) we obtain the bound

$$(7.17) \quad \|v_L - v_L^S\| \leq \left[1 + 2 \frac{\lambda_{\max}}{\lambda_{\min}}\right] \|y_L - y_L^S\|$$

whenever $(x, B_L) \in N$.

Proof of (7.8). Inequality (7.8) is a restatement of (7.15) for ω equal to y_L or y_L^S . In the case of PSB and DFP, (7.8) holds trivially. For the BFGS scale $v = v(s, y, B)$ we can write

$$(7.18) \quad \frac{v^T s}{s^T s} = \frac{y^T s}{s^T s} + \sqrt{\frac{y^T s}{s^T s} \frac{s^T B s}{s^T s}}.$$

Now, by choosing B in (7.18) to be any member of the neighborhood N_2 and using (7.14) and (7.15) we see that

$$(7.19) \quad \lambda_{\min} \leq \frac{\omega^T s}{s^T s} \leq 4\lambda_{\max}$$

for ω equal to v_L or v_L^S . Thus we have established (7.8) for all three choices of scale.

Proof of (7.9). From the mean value theorem we can write

$$(7.20) \quad \|y_L\| \leq \|\nabla_x^2 L(x + \theta s, \lambda_+)\| \|s\|$$

for some $\theta \in [0, 1]$. It follows from (7.14) and the comments following (7.14) that

$$(7.21) \quad \|y_L\| \leq 2\lambda_{\max} \|s\|.$$

Thus, (7.9) holds for $\omega = y_L$. Exactly the same proof can be used to establish (7.9) for $\omega = y_L^S$.

We must now establish (7.9) for ω equal to v_L and v_L^S for the three choices of scale under consideration. Again, for the choice of scale corresponding to the PSB, update (7.9) is trivially satisfied for ω equal to v_L and v_L^S ; indeed, $\omega = s$ for this choice. Also, for DFP we know that $v_L = y_L$ and $v_L^S = y_L^S$, and (7.18) has already been established for these two choices. For the BFGS choice of scale we can write

$$(7.22) \quad v = v(s, y, B) = y + \sqrt{\frac{y^T s}{s^T s}} \sqrt{\frac{s^T s}{s^T B s}} B s.$$

For $v = v_L$, $y = y_L$, $B = B_L$ and $(x, B_L) \in N_1 \times N_2$, (7.22), (7.21), (7.15) and (7.14) allow us to write

$$(7.23) \quad \|v_L\| \leq 2\lambda_{\max} \left(1 + 2\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \right) \|s\|.$$

Exactly the same argument can be used to establish (7.23) with v_L replaced by v_L^S . We have established (7.9) for all three choices of scale.

Proof of (7.10). A lengthy but straightforward calculation which uses the bounds (7.6)–(7.9) and adds and subtracts various appropriate quantities establishes (7.10). This proves the lemma. \square

Consider the SQP augmented Lagrangian PSP, DFP and BFGS secant methods given by Algorithm A2 of Appendix A, where

$$(7.24) \quad B_L^+ = B_L + SECANT(s, y_L, B_L, v)$$

and v is the choice of scale corresponding to each of the respective algorithms, i.e., v is given by either (1.8), (1.9) or (1.10). It is well known that these algorithms are locally q -superlinearly convergent. These results have been established by several authors (see Corollary 5.5 of Fontecilla, Steihaug and Tapia [12], for example).

The form of bounded deterioration that has been established for the SQP augmented Lagrangian PSB, respectively DFP secant method is the following:

There exist positive constants α_1 and α_2 and neighborhoods $N_1 \subset R^n$ of x_* and $N_2 \subset S(R^{n \times n})$ of $\nabla_x^2 L(x_*, \lambda_*)$ such that for $(x, B_L) \in N = N_1 \times N_2$ and (x_+, λ_+) given by (2.3) we have

$$(7.25) \quad \begin{aligned} & \|B_L^+ - \nabla_x^2 L(x_*, \lambda_*)\|_M \\ & \leq [1 + \alpha_1 \sigma(x, x_+)] \|B_L - \nabla_x^2 L(x_*, \lambda_*)\|_M + \alpha_2 \sigma(x, x_+), \end{aligned}$$

where $\sigma(x, x_+)$ is given by (7.11) and B_L^+ is given by (7.24) with the scale v given by (1.8), respectively (1.9), and $\|\cdot\|_M$ denotes the Frobenius, respectively a particular weighted Frobenius norm.

The form of bounded deterioration that has been established for the SQP augmented Lagrangian BFGS secant method is the following:

There exist positive constants α_1 and α_2 and neighborhoods $N_1 \subset R^n$ of x_* and $N_2 \subset S(R^{n \times n})$ of $\nabla_x^2 L(x_*, \lambda_*)$ such that for $(x, B_L) \in N = N_1 \times N_2$ and (x_+, λ_+) given by (2.3) we have B_L^+ is positive definite and

$$(7.26) \quad \begin{aligned} & \|(B_L^+)^{-1} - \nabla_x^2 L(x_*, \lambda_*)^{-1}\|_M \\ & \leq [1 + \alpha_1 \sigma(x, x_+)] \|B_L^{-1} - \nabla_x^2 L(x_*, \lambda_*)^{-1}\|_M + \alpha_2 \sigma(x, x_+), \end{aligned}$$

where $\sigma(x, x_+)$ is given by (7.11), B_L^+ is given by (7.24) with the scale v given by (1.10), and $\|\cdot\|_M$ denotes a particular weighted Frobenius norm.

We are now ready to establish our convergence results.

THEOREM 7.2. *Suppose that the standard assumptions given in Section 1 for problem (1.1) hold and $C \geq 0$ has been chosen so that $\nabla_x^2 L(x_*, \lambda_*) = \nabla_x^2 l(x_*, \lambda_*) + C \nabla g(x_*) \nabla g(x_*)^T$ is positive definite. Then there are positive constants ε, δ such that for $x_0 \in R^n$ and symmetric $B_l^0 \in R^{n \times n}$ satisfying $\|x_0 - x_*\| < \varepsilon$ and $\|B_l^0 - \nabla_x^2 l(x_*, \lambda_*)\| < \delta$ the iteration sequence $\{x_k\}$ generated by the SQP Lagrangian augmented scale BFGS or DFP secant method is q -superlinearly convergent to x_* .*

Proof. We will first establish bounded deterioration for the augmented scale DFP update. From Lemma 7.1 we have a neighborhood of x_* called N_1 , and associated with (7.25) we have another neighborhood of x_* also called N_1 . Let us redefine N_1 to be the intersection of these two neighborhoods of x_* . We do the same thing to obtain the neighborhood N_2 of $\nabla_x^2 L(x_*, \lambda_*)$. Now, choose a neighborhood $N_3 \subset S(R^{n \times n})$ of $\nabla_x^2 l(x_*, \lambda_*)$ and restrict N_1 , as needed, so that $(x, B_l) \in N_1 \times N_3$ implies that $B_l + C \nabla g(x_+) \nabla g(x_+)^T \in N_2$.

Recall that from Section 2, Proposition 3.1 and Proposition 4.1 the SQP augmented scale Lagrangian secant method can be equivalently written as

$$(7.27a) \quad x_+ = x - B_S^{-1} \nabla_x L(x, \lambda_+),$$

$$(7.27b) \quad B_S^+ = B_l^+ + C \nabla g(x_+) \nabla g(x_+)^T,$$

where

$$(7.28) \quad B_S = B_l + C \nabla g(x) \nabla g(x)^T,$$

λ_+ is given by (2.3a) with B_S playing the role of B_L ,

$$(7.29) \quad B_l^+ = B_l + SECANT(s, y_L^S, B_L^S, v),$$

where y_L^S is given by (7.3) and

$$(7.30) \quad B_L^S = B_l + C\nabla g(x_+) \nabla g(x_+)^T.$$

Define

$$T(w, z) = B_l + C\nabla g(w) \nabla g(w)^T + SECANT(s, z, B_L^S, v)$$

for $(x, B_l) \in N_1 \times N_3$, $w \in \{x, x_+\}$, and $z \in \{y_L, y_L^S\}$, where y_L is given by (7.2).

From Lemma 7.1, inequality (7.25) and an obvious bound we have

$$(7.31) \quad \begin{aligned} \|B_S^+ - B_*\|_M &= \|T(x_+, y_L^S) - B_*\|_M \\ &\leq \|T(x_+, y_L) - B_*\|_M + \|T(x_+, y_L^S) - T(x_+, y_L)\|_M \\ &\leq [1 + \alpha_1\sigma] \|B_L^S - B_*\|_M + \beta\sigma \\ &\leq [1 + \alpha_1\sigma] [\|B_S - B_*\|_M + \|B_L^S - B_S\|_M] + \beta\sigma \\ &\leq [1 + \alpha_1\sigma] \|B_S - B_*\|_M + \alpha'_2\sigma, \end{aligned}$$

where $B_* = \nabla_x^2 L(x_*, \lambda_*)$, $\sigma = \sigma(x, x_+)$ and α_1 and α'_2 are positive constants.

The bounded deterioration inequality (7.31) and the fact that the update formula (2.3a) is *x-dominated* (i.e., we can establish (7.4) with $\beta < 1$ as is done in the proof of Proposition 4.2 of Fontecilla, Steihaug and Tapia [12]) allows us to establish the existence of the constants ε and δ in the statement of this theorem and the q -linear convergence of the iteration sequence $\{x_k\}$ as a direct consequence of Theorem 3.1 in Fontecilla, Steihaug and Tapia [12].

Now, in turn, the q -linear convergence of $\{x_k\}$ allows us to use an argument identical to the one used by Broyden, Dennis and Moré [5] to show that

$$(7.32) \quad \lim_{k \rightarrow \infty} \frac{\|(B_S^k - \nabla_x^2 L(x_*, \lambda_*))s_k\|}{\|s_k\|} = 0,$$

where the interpretation of the index k in (7.32) is obvious. Having established (7.32), we return to Theorem 5.3 of Fontecilla, Steihaug and Tapia [12] to establish the q -superlinear convergence of the iteration sequence $\{x_k\}$. This proves the theorem for the case of the DFP update.

In the case of the BFGS algorithm we follow the standard approach of establishing convergence by working with the inverse update. From Lemma 7.1, inequality (7.26) and the well-known Banach Perturbation Lemma, further restricting N_1 and N_3 if necessary, we can determine a constant γ so that $T(w, z)^{-1}$ exists and

$$(7.33) \quad \|T(w, z)^{-1}\| \leq \gamma$$

whenever $(x, B_l) \in N_1 \times N_3$, $w \in \{x, x_+\}$ and $z \in \{y_L, y_L^S\}$.

Using Lemma 7.1, inequalities (7.26) and (7.33), the Banach Perturbation Lemma, working with $T(w, z)^{-1}$ instead of $T(w, z)$ and taking exactly the same steps as were taken in the derivation of (7.31), we obtain

$$(7.34) \quad \|(B_S^+)^{-1} - \nabla_x^2 L(x_*, \lambda_*)^{-1}\|_M \leq [1 + \alpha_1\sigma] \|B_S^{-1} - \nabla_x^2 L(x_*, \lambda_*)^{-1}\|_M + \alpha'_2\sigma,$$

where $\sigma = \sigma(x, x_+)$ and α_1 and α'_2 are positive constants.

The local and q -superlinear convergence of the SQP Lagrangian augmented scale BFGS secant method now follows from the bounded deterioration inequality (7.34) using an argument similar to the one used in the case of the DFP update. \square

THEOREM 7.3. *Suppose that the standard assumptions given in Section 1 for problem (1.1) hold and $C \geq 0$ has been chosen so that $\nabla_x^2 L(x_*, \lambda_*) = \nabla_x^2 l(x_*, \lambda_*) + C \nabla g(x_*) \nabla g(x_*)^T$ is positive definite. Then there are positive constants ε, δ such that for $x_0 \in R^n$ and symmetric $B_L^0 \in R^{n \times n}$ satisfying $\|x_0 - x_*\| < \varepsilon$ and $\|B_L^0 - \nabla_x^2 L(x_*, \lambda_*)\| < \delta$ the sequence $\{x_k\}$ generated by the SQP structured augmented Lagrangian BFGS, DFP or PSB secant method is q -superlinearly convergent to x_* .*

Proof. The proof is basically the same as the proof of Theorem 7.2. However, in this case the derivation of the bounded deterioration inequalities is slightly less complicated. \square

8. Summary and Concluding Remarks. In this paper we have taken a very close look at the structure present in the Hessian of the augmented Lagrangian. The understanding gained from this look motivated two new classes of SQP secant methods for the equality-constrained optimization problem.

The SQP augmented scale secant methods were derived by considering the SQP augmented Lagrangian secant method, utilizing structure as completely as possible, and then observing that the penalty constant cancels out in all parts of the algorithm except in the scale of the secant update. This leaves an SQP Lagrangian secant method with a modified scale. It follows from the fact that $y_i^T s$ is not necessarily positive that the augmented scale BFGS and DFP updates do not possess the hereditary positive definiteness property. However, they do have the hereditary positive definiteness property on a proper subspace (see Proposition 4.4). The change of scale compensates for the lack of positive definiteness in the Hessian of the Lagrangian and allows us to conclude (from Lemma 9.2.1 of Dennis and Schnabel [8] and Theorem 7.3 of Dennis and Moré [7]) that the augmented scale BFGS and DFP updates are least-change secant updates. This change of scale also allowed us in Section 7 to establish local q -superlinear convergence of the SQP augmented scale BFGS and DFP secant methods. We believe that this convergence result is the first that establishes either local or q -superlinear convergence of an SQP Lagrangian secant method which uses either the BFGS or the DFP updating philosophy and assumes no more than the standard assumptions. Recall that Powell [18] assumed convergence and established an r -superlinear convergence rate. Boggs and Tolle [3] assumed convergence and used additional assumptions to establish the q -superlinear convergence rate.

The SQP structured augmented Lagrangian secant methods are derived by considering the SQP augmented Lagrangian secant methods and utilizing structure only in the y vector used in the secant update (see (1.6)). The choice for the y vector is y_L^S given by (4.6) and the constant C is chosen so that $s^T y_L^S > 0$. It follows that the BFGS and DFP updates possess the hereditary positive definiteness property. These algorithms have the flavor of the modified SQP Lagrangian BFGS secant method proposed by Powell [18] (see (1.11)). In both cases, y is modified so that the modified y will have a positive inner product with the step s and hereditary positive definiteness is achieved. While Powell's modification gives hereditary positive definiteness at the expense of the Lagrangian secant equation, our modification gives hereditary positive definiteness and satisfies a structured augmented Lagrangian secant equation. That our choices of y and secant equation

are compatible was demonstrated by the fact that we were able to establish local q -superlinear convergence of the SQP structured augmented Lagrangian BFGS, DFP and PSB secant methods. Local convergence has not been established for the Powell algorithm.

In Section 6 we showed that for any secant update from the Broyden class, e.g., BFGS or DFP, the augmented scale update has the property that as the penalty constant approaches infinity, the scale approaches a scale which involves only the gradients of the constraints and does not involve the gradient of the Lagrangian. We call the secant update with this limiting scale the limit update.

In the original version of this paper, the author implied that an SQP Lagrangian secant method which used the limit update might be a viable algorithm. Both referees expressed serious doubts about such a possibility and documented these doubts in a convincing manner. These concerns motivated the author, with the aid of graduate student Hector Martinez at Rice University, to conduct a numerical investigation of the SQP Lagrangian secant method using the limit update. We observed that if an iterate was far from being feasible, then the algorithm produced a good step and the subsequent iterate was usually closer to being feasible. However, if an iterate was nearly feasible, then the behavior of the algorithm deteriorated and the subsequent iterates converged slowly. In retrospect, we see that for many reasons, including the form of the scale in the limit update, this behavior should have been expected. Moreover, as one referee pointed out, the limit update is actually a discontinuous function of s . Specifically, if $\nabla g(x_+)^T s = 0$, then the augmented scale is the standard scale and the limit update is the update corresponding to this scale. However, if in the limit update we let $s \rightarrow s^*$, where $\nabla g(x_+)s^* = 0$, then the scale converges to zero and the update with a zero scale is undefined. This will not be the case for any augmented scale with a finite value of C . It is fair to conjecture that the limit update does not lead to an algorithm which is locally q -superlinearly convergent. The analysis given in Section 7 fails for the limit update. However, from Lemma 9.2.1 of Dennis and Schnabel [8] and Theorem 7.3 of Dennis and Moré [7] the limit update is a least-change secant update.

It follows, then, that the existence of the limit update should be interpreted as a robustness property of the SQP Lagrangian augmented scale BFGS and DFP secant methods which is not shared by the SQP augmented Lagrangian BFGS and DFP secant method described in Section 2 or the SQP structured augmented Lagrangian BFGS or DFP secant method described in Section 5. Namely, if it happens in a particular iteration that C is chosen to be excessively large, then we should expect that the algorithm may not give much improvement; but we should not expect a complete disaster (as is the case for the usual SQP augmented Lagrangian BFGS and DFP secant methods).

Acknowledgments. The author acknowledges comments made by P. T. Boggs, R. B. Schnabel and H. Martinez. He also thanks H. Martinez for programming assistance. The author is particularly grateful to R. H. Byrd and J. E. Dennis for numerous discussions and comments made throughout the course of this research. He is also indebted to two anonymous referees for their thorough reviews. It was their concerns that motivated the author to produce the convergence analysis given

in Section 7 and absent from the original paper. Moreover, it was Byrd's comment that bounded deterioration for the new algorithms should follow from the known bounded deterioration of the standard augmented Lagrangian method that motivated the author to establish the approach to convergence presented in Section 7.

Appendix A. The Various SQP Secant Methods. In this appendix we collect and catalog the various SQP secant methods discussed in the main body of this paper. Recall that these algorithms approximate a solution of the equality-constrained optimization problem (1.1). Also recall that the Lagrangian $l(x, \lambda)$ associated with problem (1.1) is given by (1.4).

ALGORITHM A1: SQP Lagrangian Secant Method. By a successive quadratic programming (SQP) Lagrangian secant method for problem (1.1) we mean the iterative process

$$\begin{aligned} \text{(A.1a)} \quad & x_+ = x + s, \\ \text{(A.1b)} \quad & \lambda_+ = \lambda + \Delta\lambda, \\ \text{(A.1c)} \quad & B_l^+ = B_l + \text{SECANT}(s, y_l, B_l, v(s, y_l, B_l)), \end{aligned}$$

with $\text{SECANT}(s, y, B, v(s, y, B))$ given by (1.7b),

$$\text{(A.2)} \quad y_l = \nabla_x l(x_+, \lambda_+) - \nabla_x l(x, \lambda_+),$$

and s and $\Delta\lambda$ are respectively the solution and the multiplier associated with the solution of the quadratic program

$$\begin{aligned} \text{(A3)} \quad & \text{minimize } \nabla_x l(x, \lambda)^T s + \frac{1}{2} s^T B_l s \\ & \text{subject to } \nabla g(x)^T s + g(x) = 0. \end{aligned}$$

Notice that B_l^+ will satisfy the Lagrangian secant equation

$$\text{(A.4)} \quad B_l^+ s = y_l.$$

ALGORITHM A2: SQP Augmented Lagrangian Secant Method. By a successive quadratic programming (SQP) augmented Lagrangian secant method for problem (1.1) we mean the iterative process

$$\begin{aligned} \text{(A.5a)} \quad & x_+ = x + s, \\ \text{(A.5b)} \quad & \lambda_+ = \lambda + \Delta\lambda, \\ \text{(A.5c)} \quad & B_L^+ = B_L + \text{SECANT}(s, y_L, B_L, v(s, y_L, B_L)), \end{aligned}$$

with $\text{SECANT}(s, y, B, v(s, y, B))$ given by (1.7b),

$$\text{(A.6)} \quad y_L = \nabla_x L(x_+, \lambda_+) - \nabla_x L(x, \lambda_+),$$

s and $\Delta\lambda$ are respectively the solution and the multiplier associated with the solutions of the quadratic program

$$\begin{aligned} \text{(A.7)} \quad & \text{minimize } \nabla_x L(x, \lambda)^T s + \frac{1}{2} s^T B_L s \\ & \text{subject to } \nabla g(x)^T s + g(x) = 0. \end{aligned}$$

Notice that B_L^+ will satisfy the augmented Lagrangian secant equation

$$\text{(A.8)} \quad B_L^+ s = y_L.$$

ALGORITHM A3: SQP Lagrangian Augmented Scale Secant Method. By an SQP Lagrangian augmented scale secant method we mean an SQP Lagrangian secant method, i.e., Algorithm A1, where the choice for B_l^+ in (A.1c) is

$$(A.9) \quad B_l^+ = B_l + SECANT(s, y_l, B_l, v_L)$$

with $SECANT(s, y, B, v)$ given by (1.7b), and for a given choice of scale $v = v(s, y, B)$ the augmented scale v_L is defined to be

$$(A.10) \quad v_L = v(s, y_L^S, B_L^S),$$

where

$$(A.11) \quad y_L^S = y_l + C \nabla g(x_+) \nabla g(x_+)^T s$$

and

$$(A.12) \quad B_L^S = B_l + C \nabla g(x_+) \nabla g(x_+)^T$$

for some $C \geq 0$.

ALGORITHM A4: SQP Structured Augmented Lagrangian Secant Method. By an SQP structured augmented Lagrangian secant method we mean an algorithm of the form of Algorithm A2, where the choice for B_L^+ in (A.5c) is

$$(A.13) \quad B_L^+ = B_L + SECANT(s, y_L^S, B_L, v(s, y_L^S, B_L))$$

with $SECANT(s, y, B, v(s, y, B))$ given by (1.7b) and y_L^S given by (A.11), i.e.,

$$(A.14) \quad y_L^S = \nabla_x l(x_+, \lambda_+) - \nabla_x l(x, \lambda_+) + C \nabla g(x_+) \nabla g(x_+)^T s.$$

In this case, B_L^+ satisfies the structured augmented Lagrangian secant equation

$$(A.15) \quad B_L^+ s = y_L^S,$$

which is not the same as the augmented Lagrangian secant equation (A.8) because of the fact that y_L^S in (A.14) employs structure and differs from y_L in (A.8).

Appendix B. The Role of the Term $Cg(x) \cdot \nabla^2 g(x)$. We will attempt to justify the claim made in Section 3 that the role the term $Cg(x) \cdot \nabla^2 g(x)$ plays in the SQP augmented Lagrangian quasi-Newton methods (see Algorithm A2 of Appendix A) is a negative one and all attempts at utilizing structure should lead to approximations which ignore this second-order term.

Consider the SQP augmented Lagrangian quasi-Newton method, where B_L^S is given by

$$(B.1) \quad B_L^S = B_l + Cg(x) \cdot \nabla^2 g(x) + C \nabla g(x) \nabla g(x)^T$$

with B_l interpreted as an approximation to $\nabla_x^2 l(x, \lambda)$. From Section 2 we know that the defining relations for the SQP augmented Lagrangian quasi-Newton method are given equivalently by (2.2). As long as $C \neq 0$, we can divide the first equation in (2.2) by C without changing the solution set. Now, if we let C become arbitrarily large, we see that our defining relations approach the limiting defining relations

$$(B.2) \quad \begin{pmatrix} g(x) \cdot \nabla^2 g(x) & 0 \\ \nabla g(x)^T & 0 \end{pmatrix} \begin{pmatrix} s \\ \lambda_+ \end{pmatrix} = - \begin{pmatrix} 0 \\ g(x) \end{pmatrix}.$$

The defining relations (B.2) are clearly singular. Moreover, we cannot guarantee that they are consistent. Indeed, they will be inconsistent in any application where $g(x) \cdot \nabla^2 g(x)$ is nonsingular, e.g., $g(x) = x^T x - 1$ and x such that $g(x) \neq 0$.

The above observation may shed some light on the situation concerning the poor numerical performance of the SQP augmented Lagrangian secant methods for large C . Toward this end, let us attempt to identify those attributes of the SQP formulation and the augmented Lagrangian which have created this bad fit.

From (2.2) and (B.1) we see that one SQP augmented Lagrangian quasi-Newton step consists of constructing (x_+, λ_+) , where $x_+ = x + s$ and (s, λ_+) is the solution of the linear system

$$(B.3) \quad [B_l + \nabla_x l(x, \lambda_+)] + C[(g(x) \cdot \nabla^2 g(x) + \nabla g(x) \nabla g(x)^T s + \nabla g(x) g(x))] = 0,$$

$$(B.4) \quad \nabla g(x)^T + g(x) = 0.$$

Equation (B.3) can be viewed as a weighted average of the two equations

$$(B.5) \quad B_l s = -\nabla_x l(x, \lambda_+)$$

and

$$(B.6) \quad (g(x) \cdot \nabla^2 g(x) + \nabla g(x) \nabla g(x)^T) s = -\nabla g(x) g(x).$$

We recognize (B.5) as a quasi-Newton step for the nonlinear equation problem

$$(B.7) \quad \nabla_x l(x, \lambda_+) = 0$$

and (B.6) as a Newton step for the unconstrained optimization problem

$$(B.8) \quad \text{minimize } g(x)^T g(x).$$

The second equation in the SQP formulation, namely (B.4), can be viewed as a (right-inverse) Newton step for the nonlinear equation

$$(B.9) \quad g(x) = 0.$$

It follows that increasing C in this SQP formulation corresponds to requiring greater agreement between the Newton iterate for problem (B.8) and a particular right-inverse Newton iterate for problem (B.9). Moreover, not only is there no reason that these two iterates should agree, but as we have seen, there are cases in which they cannot agree.

In summary, then, we have seen that in the SQP formulation the augmented Lagrangian itself takes the constraints into account, and adding the requirement that the iterate satisfy linearized constraints is in effect restricting the algorithm, i.e., overdetermining the iterate. It follows that the term $g(x) \cdot \nabla^2 g(x)$ and the satisfaction of linearized constraints are incompatible. If we remove the term $g(x) \cdot \nabla^2 g(x)$ from the SQP augmented Lagrangian formulation, then from (B.3) and (B.4) (equivalently Proposition 3.1) we see that, except possibly for the updating of the Hessian approximation B_l , the role of the penalty constant has been eliminated and the augmented Lagrangian subproblem has been reduced to that of the Lagrangian. Clearly, this is the situation described in Section 4, where the SQP Lagrangian augmented scale secant methods are presented.

1. M. AVRIEL, *Nonlinear Programming: Analysis and Methods*, Prentice-Hall, Englewood Cliffs, N. J., 1976.
2. M. C. BARTHOLOMEW-BIGGS, *Matrix Updating in Nonlinear Programming Calculation*, TR No. 156, The Hatfield Polytechnic Numerical Optimization Center, Hatfield, Hertfordshire, England, 1985.
3. P. T. BOGGS & J. W. TOLLE, *An Efficient Strategy for Utilizing a Merit Function in Nonlinear Programming Algorithms*, T-R 85-5, Curriculum in Operations Research and Systems Analysis, University of North Carolina, Chapel Hill.
4. 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.
5. C. G. BROYDEN, J. E. DENNIS, JR. & J. J. MORÉ, "On the local and superlinear convergence of quasi-Newton methods," *J. Inst. Math. Appl.*, v. 12, 1973, pp. 223-245.
6. 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.
7. J. E. DENNIS, JR. & J. J. MORÉ, "Quasi-Newton methods, motivation and theory," *SIAM Rev.*, v. 19, 1974, pp. 46-89.
8. J. E. DENNIS, JR. & R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, N. J., 1983.
9. J. E. DENNIS, JR. & H. F. WALKER, "Convergence theorems for least-change secant update methods," *SIAM J. Numer. Anal.*, v. 18, 1981, pp. 949-987.
10. R. FLETCHER, *Practical Methods of Optimization*, vol. 1, Wiley, New York, 1980.
11. R. FONTECILLA, *The Lack of Positive Definiteness in the Hessian in Constrained Optimization*, TR 1334, Department of Computer Science, University of Maryland, College Park, Maryland 20742, 1983.
12. R. FONTECILLA, T. STEIHAUG & R. A. TAPIA, "A convergence theory for a class of quasi-Newton methods for constrained optimization," *SIAM J. Numer. Anal.*, v. 24, 1987, pp. 1133-1152.
13. P. GILL, W. MURRAY & M. WRIGHT, *Practical Optimization*, Academic Press, New York, 1981.
14. S. T. GLAD, "Properties of updating methods for the multipliers in augmented Lagrangians," *J. Optim. Theory Appl.*, v. 28, 1979, pp. 135-156.
15. S.-P. HAN, "Superlinearly convergent variable metric algorithms for general nonlinear programming problems," *Math. Programming*, v. 11, 1976, pp. 263-283.
16. S.-P. HAN, "Dual variable metric algorithms for constrained optimization," *SIAM J. Control Optim.*, v. 15, 1977, pp. 546-565.
17. J. NOCEDAL & M. OVERTON, "Projected Hessian updating algorithms for nonlinearly constrained optimization," *SIAM J. Numer. Anal.*, v. 22, 1985, pp. 821-850.
18. 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.
19. R. A. TAPIA, "Diagonalized multiplier methods and quasi-Newton methods for constrained optimization," *J. Optim. Theory Appl.*, v. 22, 1977, pp. 135-194.
20. 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.