Practical Update Criteria for Reduced Hessian SQP, Part I: Global Analysis

Yuanfu Xie
Richard H. Byrd

CU-CS-753-94

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

Practical Update Criteria for Reduced Hessian SQP, Part I: Global Analysis

Yuanfu Xie[1] and Richard H. Byrd[2]

December 7, 1994
Revised November 1995

Short title: Reduced Hessian Algorithms

[1]Department of Computer Science, Campus Box 430, University of Colorado, Boulder, Colorado 80309-0430 U.S.A. This author was supported by Air Force Office of Scientific Research grant AFOSR-90-0109 and National Science Foundation grant CCR-8920519

[2]Department of Computer Science, Campus Box 430, University of Colorado, Boulder, Colorado, 80309 U.S.A. This author was supported, in part, by Air Force Office of Scientific Research grant AFOSR-90-0109, Army Research Office grant DAAL 03-91-G-0151, and National Science Foundation grant CCR-9101795.

PRACTICAL UPDATE CRITERIA FOR REDUCED HESSIAN SQP, PART I: GLOBAL ANALYSIS

Y. F. XIE* AND R. H. BYRD†

Abstract. In this paper, a new update criterion is proposed to improve the Nocedal and Overton update criterion for the reduced Hessian successive quadratic programming. Global and R-linear convergence is proved for the new criterion and the Nocedal and Overton criterion using non-orthogonal basis matrices, which allow efficient implementations of the reduced Hessian successive quadratic programming for solving large scale equality constrained problems.

Key words. constrained optimization, nonlinear programming, quasi-Newton methods, reduced Hessian algorithms, successive quadratic programming

AMS subject classifications. 49, 65

1. Introduction. In this paper, we consider some critical issues in efficiently solving nonlinearly constrained optimization problems by reduced Hessian successive quadratic programming.

Successive quadratic programming (SQP) algorithms have proven to be very efficient for solving small and medium size equality constrained optimization problems,

$$\begin{array}{rl} min & f(x) \\ s.t. & c(x) = 0 \end{array} \tag{1.1}$$

where $f : R^n \rightarrow R^1$ and $c : R^n \rightarrow R^t$ for positive integers n and t, with $n > t$ (see Han [8], Powell [14]). The reduced Hessian approach allows us to use SQP for a significant class of very large problems, especially when implemented with generalized basis matrices.

Given an approximate solution x_k, SQP algorithms compute a search direction d_k from the quadratic programming problem:

$$\left.\begin{array}{rl} min & g_k^T d + \frac{1}{2} d^T M_k d \\ s.t. & c(x_k) + A_k^T d = 0 \end{array}\right\}$$

where $g_k = \nabla f(x_k)$, $A_k = \nabla c(x_k) \equiv (\nabla c_1(x_k), \cdots, \nabla c_t(x_k))$ and the matrix M_k approximates to the Hessian $G_k = \nabla^2_{xx} L(x_k, \lambda_k)$ of the Lagrangian function of (1.1), $L(x, \lambda)$, which has the form:

$$L(x, \lambda) = f(x) + \lambda^T c(x), \tag{1.2}$$

where λ is a Lagrangian multiplier. The Lagrangian multiplier is given by

$$\lambda_k = -(A_k)_L^{-1} g_k, \tag{1.3}$$

* NOAA, ERL/FSL, The Forecasting Systems Laboratory, 325 Broadway, Boulder, CO 80303.
This author was supported by the Air Force Office of Scientific Research Grant AFOSR-90-0109 and National Science Foundation grant CCR-8920519.

† Department of Computer Science, Campus Box 430, University of Colorado, Boulder, CO 80309.
This author was supported, in part, by Air Force Office of Scientific Research grant AFOSR-90-0109, Army Research Office grant DAAL 03-91-G-0151, and National Science Foundation grant CCR-9101795.

where $(A_k)_L^{-1}$ is a left inverse of A_k, giving an approximate solution to $g_k + A_k\lambda = 0$. Then a new approximation to the solution x^* is given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.4}$$

where some line search strategy is used to determine step length α_k and ensure convergence. (Alternatively, a trust region can be used instead of a line search, but this paper will focus on the more widely used line search approach.) Although SQP methods are of the best approaches for small and medium size problems, the applicability of this approach for very large problems is limited because of the need to store and manipulate a $n \times n$ matrix M_k, which cannot be expected to be sparse if a quasi-Newton update is used.

However, reduced Hessian SQP (RHSQP) algorithms, which use a matrix B_k to approximate $Z_k^T G_k Z_k$ where Z_k is a null space basis, can potentially be very efficient for solving large scale constrained optimization problems (i.e. n is very large), especially for the problems with small $n - t$.

RHSQP algorithms have two advantages compared with the other SQP algorithms.

- It is reasonable to use the quasi-Newton secant methods to approximate the reduced Hessian matrix $Z_k^T G_k Z_k$ because this matrix is positive definite if x_k is close to the solution of (1.1) and the second-order sufficient optimality condition holds at the solution.
- It is more efficient to store a $(n-t) \times (n-t)$ matrix B_k than to store a $n \times n$ matrix M_k. Thus for a given n, a larger t requires less space for storing B_k. This is an important advantage for solving large scale problems.

To update the matrix B_k by means of the quasi-Newton methods has been an interesting issue and many update strategies have been proposed, for examples, [2], [3], [6], [7], [10] and [11]. Among these update strategies, there are two typical strategies and others are slight variations of these two. One uses the exact null space information [3] and one uses a full step information [11]. We call them the null space secant update strategy and the step secant update strategy, respectively.

To ensure the accuracy of the step secant update strategy, Nocedal and Overton suggest an update criterion [11], under which B_k is updated. In order to improve the numerical performance of the step secant update strategy using the Nocedal and Overton criterion, a new update criterion is proposed in this paper.

For the methods using these two update strategies, several convergence results have been established. For the RHSQP algorithms using the null space secant update strategy, Coleman and Conn [3] have proved 2-step Q-superlinear convergence assuming x_1 and B_1 are sufficiently close to x^* and $Z_*^T \nabla_{xx}^2 L(x^*, \lambda^*) Z_*$, respectively, and Byrd and Nocedal [2] have shown its global convergence, R-linear and 2-step Q-superlinear convergence with the l_1 and Fletcher merit functions. For the step secant update strategy with the Nocedal and Overton update criterion (2.12), Nocedal and Overton [11] established local 2-step Q-superlinear convergence for x_1 and B_1 sufficiently close to x^* and $Z_*^T \nabla_{xx}^2 L(x^*, \lambda^*) Z_*$, respectively, however no global and R-linear convergence is proved. All of these analyses assume Z_k is an orthonormal basis of $null(A_k^T)$.

A general basis Z_k of $null(A_k)$ has been used by Fletcher [5], Gabay [6], and Gilbert [7]. Fletcher has discussed his Successive Linear Programming algorithm using any basis of $null(A_k)$. Gabay and Gilbert use general bases to discuss RHSQP. Gabay's update strategy is equivalent to the step secant update strategy but he used Powell's damped technique to inherit the positive definiteness. It is difficult to prove superlinear convergence without assuming $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded for the Powell damped technique. Although Gilbert [7], however, has discussed general Z_k in his global analysis, his longitudinal path strategy may cost more gradient evaluations.

Because of the complexity of the analysis, superlinear convergence will be discussed in a second paper subsequent to this one. This paper is devoted to propose a new update criterion and to prove the global and R-linear convergence for the step secant update strategy with two commonly used merit functions. All of these results are proved without requiring orthogonality of the basis matrix Z_k and without assuming $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded. In the next section, the new update criterion used in the step secant update strategy is introduced, and the general RHSQP algorithms, and the merit functions used to force global convergence are described. The global convergence analysis will be in section 3. The R-linear convergence will be established in Section 4. The numerical experiments are presented in Section 5.

In the rest of the paper, the following notations are used for simplicity

$$\begin{aligned} S_1 &= \{\ j \mid B_{j+1} = B^{BFGS}(B_j, s_j, y_j)\ \} \\ S_2 &= \{\ j \mid B_{j+1} = B_j\ \} \\ S_1^k &= [1, 2, \cdots, k] \cap S_1 \\ S_2^k &= [1, 2, \cdots, k] \cap S_2 \end{aligned}$$

where

$$B^{BFGS}(B_j, s_j, y_j) = B_j - \frac{B_j s_j s_j^T B_j}{s_j^T B_j s_j} + \frac{y_j y_j^T}{s_j^T y_j}, \tag{1.5}$$

the BFGS update. Furthermore, $\|\cdot\|$ stands for the l_2 norm, $\|\cdot\|_1$ for the l_1 norm and $\|\cdot\|_\infty$ for the infinity norm.

2. A New Update Criterion and General RHSQP with Merit Functions. The reduced Hessian technique can be derived using general basis matrices and their pseudo-inverses from the SQP methods. Suppose Z_k is any basis matrix of the null space of A_k^T (i.e. $A_k^T Z_k = 0$ and Z_k is full rank), $(Z_k)_L^{-1}$ and $(A_k)_L^{-1}$ are left inverse matrices of Z_k and A_k, respectively and satisfy

$$(A_k)_L (Z_k)_L^{-T} = (Z_k)_L^{-1} (A_k)_L^{-T} = 0. \tag{2.1}$$

Then, we have

$$\begin{pmatrix} (Z_k)_L^{-1} \\ A_k^T \end{pmatrix} (Z_k \quad (A_k)_L^{-T}) = (Z_k \quad (A_k)_L^{-T}) \begin{pmatrix} (Z_k)_L^{-1} \\ A_k^T \end{pmatrix} = I, \tag{2.2}$$

and G_k can be written as the following,

$$\begin{aligned} G_k &= ((Z_k)_L^{-T} \quad A_k) \begin{pmatrix} Z_k^T \\ (A_k)_L^{-1} \end{pmatrix} G_k (Z_k \quad (A_k)_L^{-T}) \begin{pmatrix} (Z_k)_L^{-1} \\ A_k^T \end{pmatrix} \\ &= ((Z_k)_L^{-T} \quad A_k) \begin{pmatrix} Z_k^T G_k Z_k & Z_k^T G_k (A_k)_L^{-T} \\ (A_k)_L^{-1} G_k Z_k & (A_k)_L^{-1} G_k (A_k)_L^{-T} \end{pmatrix} \begin{pmatrix} (Z_k)_L^{-1} \\ A_k^T \end{pmatrix}. \end{aligned}$$

As is well-known, the reduced Hessian approach is to neglect the cross terms i.e.,

$$\begin{aligned} G_k &\simeq ((Z_k)_L^{-T} \quad A_k) \begin{pmatrix} Z_k^T G_k Z_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} (Z_k)_L^{-1} \\ A_k^T \end{pmatrix} \\ &= (Z_k)_L^{-T} (Z_k^T G_k Z_k)(Z_k)_L^{-1} = M_k. \end{aligned}$$

It chooses such matrix M_k in the SQP method and uses B_k, a $(n-t)\times(n-t)$ matrix, to approximate $(Z_k^T G_k Z_k)$. Thus, RHSQP algorithms generate a search direction d_k at x_k by solving:

$$\left.\begin{array}{ll} min & g_k^T d + \frac{1}{2} d^T (Z_k)_L^{-T} B_k (Z_k)_L^{-1} d \\ s.t. & c(x_k) + A_k^T d = 0 \end{array}\right\}$$

Note that (2.1) implies that (1.3) is also a Lagrangian multiplier of the above quadratic programming. The solution d_k, may be expressed as

$$d_k = h_k + v_k, \tag{2.3}$$

where

$$h_k = -Z_k B_k^{-1} Z_k^T g_k \tag{2.4}$$

and

$$v_k = -(A_k)_L^{-T} c(x_k). \tag{2.5}$$

There are two widely used instantiations of the generalized inverses. One is based on a QR factorization and widely used in discussions of RHSQP methods, for example, [11]. In this instantiation, the inverse matrices are given using Nocedal and Overton's notation,

$$(A_k)_L^{-1} = (R_k^{-1} \quad 0) \begin{pmatrix} Y_k^T \\ Z_k^T \end{pmatrix} \qquad (Z_k)_L^{-1} = Z_k^T \tag{2.6}$$

where Y_k and Z_k are orthonormal matrices derived from a QR factorization of A_k,

$$A_k = (Y_k \quad Z_k) \begin{pmatrix} R_k \\ 0 \end{pmatrix}. \tag{2.7}$$

The other instantiation, which is rather interesting for large scale problems, is based on a LU decomposition of A_k. Suppose $A_k^T = (A_B \quad A_N)$, where A_B is non-singular. This instantiation chooses

$$\begin{aligned} (A_k)_L^{-1} &= (A_B^{-T} \quad 0) \\ Z_k &= \begin{pmatrix} -A_B^{-1} A_N \\ I \end{pmatrix} \\ (Z_k)_L^{-1} &= (0 \quad I) \end{aligned}$$

where a LU decomposition of A_B is necessary for its inverse, which may take an advantage of the sparsity of A_k for a large scale problem. For some very large scale problems where LU is not applicable, iterative methods could be used to invert A_B and its transpose. The generalized basis matrices give the RHSQP methods a great flexibility in dealing with large scale problems.

The matrix B_k is to be updated using gradient difference information. Two ways of obtaining this information have been proposed. Consider the gradient difference of the Lagrangian function $\nabla_x L(x, \lambda_k) - \nabla_x L(x_k, \lambda_k)$. Its projection on the null space using a given basis matrix Z_k can be approximated by

$$\begin{aligned} (2.8) \qquad & Z_k^T(\nabla_x L(x, \lambda_k) - \nabla_x L(x_k, \lambda_k)) \\ \simeq\ & Z_k^T G_k (Z_k \ \ (A_k)_L^{-T}) \begin{pmatrix} (Z_k)_L^{-1} \\ A_k^T \end{pmatrix} (x - x_k) \\ =\ & Z_k^T G_k Z_k (Z_k)_L^{-1}(x - x_k) + Z_k^T G_k (A_k)_L^{-T} A_k^T (x - x_k). \end{aligned}$$

To apply a quasi-Newton method such that $B_k \simeq Z_k^T G_k Z_k$, it is ideal to choose x such that the second term of the last equation in (2.8) disappears. By using x_{k+1}, one would choose the component of $x_{k+1} - x_k$ along the null space of A_k from x_k i.e., $x = x_k + \alpha_k h_k$ and

$$(2.9) \qquad y_k = Z_k^T \nabla_x L(x_k + \alpha_k h_k, \lambda_k) - \nabla_x L(x_k, \lambda_k)$$

$$(2.10) \qquad s_k = (Z_k)_L^{-1} \alpha_k h_k = (Z_k)_L^{-1}(x_{k+1} - x_k)$$

and then the quasi-Newton equation $y_k = B_{k+1} s_k \simeq Z_k^T G_k Z_k s_k$ is satisfied. We call this first update strategy the null space secant update strategy because it uses the exact reduced Hessian information along the null space of A_k. The drawback of this strategy is that it imposes a significant extra cost to evaluate y_k when gradient evaluations of f and c are expensive.

The step secant update strategy of the second update category, uses

$$(2.11) \qquad y_k = Z_k^T(\nabla_x L(x_{k+1}, \lambda_k) - \nabla_x L(x_k, \lambda_k))$$

to update B_{k+1} and saves the extra computation of the gradients of the Lagrangian. However such y_k may not provide accurate information on the derivatives of $L(x, \lambda)$ along the null space of the constraints because of presence of the second term in (2.8). Thus, updates of B_{k+1} must be skipped at some iterations where the second terms are large. If we replace x by x_{k+1} in (2.8), the second term becomes $Z_k^T G_k v_k$. Because B_{k+1} is expected to approximate $Z_k^T \nabla_{xx}^2 L(x_k, \lambda_k) Z_k$, the update could in fact result in great loss in accuracy of B_{k+1} if the vertical component v_k is not small when B_{k+1} is updated. Nocedal and Overton [11] proposed a criterion which is simply referred as the Nocedal and Overton update criterion and under which B_{k+1} is updated if and only if

$$(2.12) \qquad \|v_k\|_2 \leq \frac{\eta}{(k+1)^{1+\epsilon}} \|h_k\|_2,$$

where η and ϵ are positive constants; otherwise $B_{k+1} = B_k$. Actually, they use $\|s_k\|$ instead of $\|h_k\|$ in their criterion and under their orthogonality assumption of Z_k, $\|s_k\| = \|h_k\|$. It can be seen that the larger the k is, the more accurate information of the reduced Hessian (2.12) can be provided to B_{k+1} when B_{k+1} is updated. In Section 4, we show a set of similar criteria with milder conditions on update steps. To globalize the algorithm, $s_k^T y_k > 0$ has to be tested in order to inherit the positive definiteness from B_k.

In this section, a new update criterion is introduced, which is designed in a such way to improve the numerical performance of the Nocedal and Overton update criterion, and general RHSQP algorithms are described with two merit functions.

Numerical experiments of the step secant update strategy show the Nocedal and Overton update criterion often skips the updates in a large proportion of the cases. It appears to be that the criterion (2.12), which depends on the iteration number, is too strong, forcing updates to be skipped and sometimes slowing down the convergence. The criterion (2.12) may be relaxed by allowing updates whenever the horizontal component $||h_k||$ is not small comparing to the vertical component $||v_k||$. A new update criterion is thus proposed, which not only allows more updates but also automatically guarantees the positive definiteness of $\{B_k\}$.

Positive Curvature Criterion: For constants, $\zeta_1 \geq \zeta_2 > 0$, the update criterion requires

$$s_k^T y_k > \zeta_1 ||\alpha_k v_k||^2, \quad \forall k \in S_1 \tag{2.13}$$

$$s_k^T y_k \leq \zeta_2 ||\alpha_k v_k||^2, \quad \forall k \in S_2. \tag{2.14}$$

If $\zeta_2 < \zeta_1$, these conditions leave an intermediate case where neither equation is satisfied, giving the algorithm flexibility in deciding whether to update. This new criterion is referred to as the "positive curvature update criterion" for simplicity because (2.13) implies that the Lagrangian has a significantly positive curvature, which makes B_{k+1} automatically inherit the positive definiteness from B_k. Lemma 4.7 proved later shows that this criterion satisfies $||v_k|| \leq \gamma_8 ||h_k||$ whenever B_{k+1} is updated, and $||h_k|| \leq \gamma_8 ||v_k||$ whenever an update is skipped, where $\gamma_8 > 0$ is a constant. Intuitively, it allows more updates than the Nocedal and Overton update criterion and our numerical experiments in Section 5 support this. Its numerical performance is so good that the step secant update strategy with the positive curvature criterion is very competitive to the null space secant update strategy.

Because the global and R-linear convergence of the null space secant update strategy has been proved by Byrd and Nocedal [2], we consider the step secant update strategy only in this paper. In the following description of the algorithm, $\varphi(x)$ stands for the merit function and $D\varphi(x; d)$ denotes the directional derivative of φ along d at x.

Algorithm 2.1.

The constants $\eta \in (0, \frac{1}{2})$ and τ, τ' with $0 < \tau < \tau' < 1$ are given. Let x_1 and B_1 be an initial point and initial positive definite matrix.

1. *Compute $d_k = h_k + v_k$ by solving (2.4) and (2.5).*
2. *Adjust the merit function φ according to x_k if it is necessary.*
3. *Set $\alpha_k = 1$ and check the line search condition,*

$$\varphi(x_k + \alpha_k d_k) \leq \varphi(x_k) + \alpha_k \eta D\varphi(x_k; d_k).$$

If it is violated, choose a new $\alpha_k \in [\tau\alpha_k, \tau'\alpha_k]$ and check it again.

4. *Set $x_{k+1} = x_k + \alpha_k d_k$.*
5. *Compute s_k by (2.10) and y_k by (2.11). Update B_{k+1} by*

$$B_{k+1} = \begin{cases} B^{BFGS}(B_k, s_k, y_k), & \text{if a criterion holds} \\ B_k, & \text{otherwise} \end{cases}$$

6. *If a stopping condition is not satisfied, set $k = k + 1$ and go to step 1; otherwise, stop.* □

In analyzing this algorithm, we do not use a stopping condition, in order to study the entire sequence. In our numerical tests, we use a stopping condition, $\|Z_k^T g_k\| + \|c_k\| < \varepsilon$ where $\varepsilon > 0$ is a given constant.

There are two widely used merit functions, the l_1 and Fletcher merit functions. Han [8] first introduced the l_1 merit function as

$$\phi_\mu(x) = f(x) + \mu\|c(x)\|_1,$$

where μ is called a penalty parameter. The l_1 merit function is very successful in global analysis. However its penalty term is nondifferentiable and this non-differentiability may affect the speed of the convergence (the Maratos effect). However, a directional derivative exists and as shown in [2],

$$D\phi_\mu(x_k; d_k) \le g_k^T h_k - (\mu - \|\lambda_k\|)\|c_k\|_1 \tag{2.15}$$

using the fact that $g_k^T v_k = -\lambda_k^T c_k$. The Fletcher merit function [4] is a differentiable merit function:

$$\Phi_\nu(x) = f(x) + \lambda(x)^T c(x) + \frac{1}{2}\nu\|c(x)\|_2^2,$$

where $\lambda(x)$ is a Lagrange multiplier estimate at x and has a form of (1.3) and ν the penalty parameter. Note that at x_k, the directional derivative is

$$\nabla\Phi_\nu(x_k)^T d_k = g_k^T h_k + g_k^T v_k + \lambda_k^T A_k^T d_k + c_k^T \nabla\lambda(x_k) d_k + \nu c_k^T A_k^T d_k.$$

From (1.3) and (2.5), $A_k^T d_k = -c_k$ and $g_k^T v_k = -g_k^T (A_k)_L^{-T} c_k = -\lambda_k^T c_k$. Thus

$$\nabla\Phi_\nu(x_k)^T d_k = g_k^T h_k + c_k^T \nabla\lambda(x_k) d_k - \nu c_k^T c_k. \tag{2.16}$$

With these merit functions, we can explicitly define the step 2 of Algorithm 2., i.e., how to choose the penalty parameters, μ and ν. In our global and R-linear convergence analysis, it is assumed that the penalty parameters, μ_k and ν_k are monotonically increased. The following adjusting procedure of these penalty parameters is simply called **step** $2'$ of Algorithm 2.1: for the l_1 merit function, the penalty parameter μ_k is chosen by

$$\mu_{k+1} = \begin{cases} \|\lambda_k\|_\infty + 2\rho & \text{if } \mu_k < \|\lambda_k\|_\infty + \rho \\ \mu_k & \text{otherwise,} \end{cases} \tag{2.17}$$

and for the Fletcher merit function, the penalty parameter ν_k is chosen by

$$\nu_{k+1} = \begin{cases} \bar\nu_k + 2\rho & \text{if } \nu_k < \bar\nu_k + \rho \\ \nu_k & \text{otherwise} \end{cases} \tag{2.18}$$

where $\bar\nu_k$ is defined by

$$\bar\nu_k = \frac{d_k^T \nabla\lambda(x_k) c_k + \frac{1}{2} g_k^T h_k}{\|c_k\|^2} \tag{2.19}$$

where $\rho > 0$ is a constant. It is shown that in [2] that $\bar\nu_k$ is bounded above by $O\left(\frac{s_k^T s_k}{s_k^T B_k s_k}\right)$ and thus $\{\bar\nu_k\}$ is bounded if $\{\|B_k^{-1}\|\}$ is. Although we cannot bound

$\{||B_k^{-1}||\}$ prior to our global convergence analysis, this fact indicates that boundedness of $\{\bar{\nu}_k\}$ is at least a reasonable assumption. If (2.18) is imposed, it follows easily from (2.16) that, as shown in [2],

$$\nabla\Phi_{\nu_k}(x_k)^T d_k \leq -\frac{1}{2} g_k^T h_k - \rho||c_k||^2. \tag{2.20}$$

Since by (2.15) and (2.20), d_k is a descent direction for either merit functions, it follows that step 3 of Algorithm 2.1 will terminate in a finite number of iterations. There are also some non-monotonically increasing strategies which are widely used, e.g., $\mu_{k+1} = ||\lambda_k|| + 2\rho$. Although non-monotonically increasing procedures numerically perform better than the monotonically increasing strategies in many numerical tests, there is no global and R-linear analysis established.

3. Global Convergence. The global convergence of the RHSQP algorithms using the step secant update strategy is based upon the following assumptions.

ASSUMPTION 3.1. *It is assumed that*

1. *$f : R^n \to R^1$ and $c : R^n \to R^t$ and their first- and second-order derivatives are uniformly bounded in a closed set $D \subset R^n$, which contains $\{x_k\}$.*
2. *The matrix $A(x)$ is full rank for all $x \in D$ and there are constants $\gamma_A > 0$ and $\gamma_Z > 0$ such that*

$$\begin{array}{rl} ||A(x)|| \leq \gamma_A & \quad ||A(x)_L^{-1}|| \leq \gamma_A \\ ||Z(x)|| \leq \gamma_Z & \quad ||Z(x)_L^{-1}|| \leq \gamma_Z \\ A(x)^T Z(x) = 0 & \quad Z(x)_L^{-1} A(x)_L^{-T} = 0. \end{array}$$

3. *For a given μ or ν, $\phi_\mu(x)$ and $\Phi_\nu(x)$ are bounded below.*

We also assume that there are $m > 0$ and $M > 0$ such that

$$\frac{s_k^T y_k}{s_k^T s_k} \geq m \tag{3.1}$$

$$\frac{y_k^T y_k}{s_k^T y_k} \leq M \tag{3.2}$$

for the global analysis. For unconstrained problems, a proper line search strategy and the convexity of the objective function imply (3.1) and (3.2) [13]. However, it requires strong conditions for constrained problems. The following lemma shows that the uniformly positive definiteness of the reduced Hessian along the null space of $A(x)^T$, which is true locally around the solution x^* where the second order optimality condition holds, implies (3.1) and (3.2) for constrained problem if the step secant update strategy is used with an update criterion which satisfies (2.13) at all update steps or in such a way that $||v_k||/||h_k||$ is sufficiently small at all update steps. Obviously, the positive curvature criterion is one of such criteria and so is the Nocedal and Overton update criterion.

LEMMA 3.1. *Suppose an RHSQP algorithm uses the step secant update strategy in a such way that (2.13) is satisfied or $||v_k||/||h_k||$ is sufficiently small for $k \in S_1$ large enough. Let D be a closed convex set containing $\{x_k\}_{k=K_0}^{\infty}$ for some K_0. Assume:*

1. *the second order sufficient conditions hold on D,*

$$m_0||u||^2 \leq u^T \nabla_{xx}^2 L(x, \lambda_k) u \qquad \forall u \in R^n : \quad A_k^T u = 0$$

for all $x \in D$, any integer $k \geq K_0$, and some constant $m_0 > 0$;

2. for some constant $M_0 > 0$ and any $x \in D$ and $k \geq K_0$,

$$\|\nabla^2_{xx} L(x, \lambda_k)\| \leq M_0.$$

Then there are constants $m > 0$ and $M > 0$ such that (3.1) and (3.2) hold whenever B_{k+1} is updated with $k > K_0$.

Proof. First consider a criterion satisfying (2.13). Consider two cases.

Case 1. $2\gamma_Z M_0 \|\alpha_k v_k\| \leq m_0 \|s_k\|$:

By (2.10) and the Taylor expansion of y_k and the inequalities $\|Z_k s_k\| \leq \gamma_Z \|s_k\|$ and $\|s_k\| = \|(Z_k)_L^{-1} Z_k s_k\| \leq \gamma_Z \|Z_k s_k\|$, the hypothesis of this lemma implies

$$\begin{aligned} s_k^T y_k &= s_k^T Z_k^T \nabla^2_{xx} L(x_k + \xi d_k, \lambda_k)(\alpha_k h_k + \alpha_k v_k) \\ &= s_k^T Z_k^T \nabla^2_{xx} L(x_k + \xi d_k, \lambda_k) Z_k s_k + s_k^T Z_k^T \nabla^2_{xx} L(x_k + \xi d_k, \lambda_k)(\alpha_k v_k) \\ &\geq m_0 \|Z_k s_k\|^2 - M_0 \|Z_k s_k\| \|\alpha_k v_k\| \geq \frac{1}{2} \frac{m_0}{\gamma_Z^2} \|s_k\|^2. \end{aligned}$$

Since $\alpha_k h_k = Z_k s_k$, then we have

$$\begin{aligned} \frac{y_k^T y_k}{s_k^T y_k} &= \|Z_k^T \nabla^2_{xx} L(x_k + \xi d_k, \lambda_k)(Z_k s_k + \alpha_k v_k)\|^2 / (s_k^T y_k) \\ &\leq (\gamma_Z M_0)^2 \frac{(\gamma_Z \|s_k\| + \|\alpha_k v_k\|)^2}{\frac{m_0}{2\gamma_Z^2} \|s_k\|^2} \leq 2 \frac{\gamma_Z^4 M_0^2}{m_0} \left(\gamma_Z + \frac{m_0}{2\gamma_Z M_0}\right)^2. \end{aligned}$$

Case 2. $2\gamma_Z M_0 \|\alpha_k v_k\| > m_0 \|s_k\|$:

As the criterion satisfies (2.13),

$$s_k^T y_k \geq \zeta_1 \|\alpha_k v_k\|^2 \geq \zeta_1 \left[\frac{m_0}{2\gamma_Z M_0}\right]^2 \|s_k\|^2,$$

and

$$\begin{aligned} \frac{y_k^T y_k}{s_k^T y_k} &\leq (\gamma_Z M_0)^2 \frac{(\gamma_Z \|s_k\| + \|\alpha_k v_k\|)^2}{\zeta_1 \|\alpha_k v_k\|^2} \\ &\leq \frac{(\gamma_Z M_0)^2}{\zeta_1} \left(2 \frac{\gamma_Z M_0}{m_0} + 1\right)^2. \end{aligned}$$

Therefore, there exist $m > 0$ and $M > 0$ such that (3.1) and (3.2) hold. For the criterion that $\|v_k\| / \|h_k\|$ is sufficiently small, the analysis is identical to Case 1 above. □

By imposing (3.1) and (3.2), the global convergence is proved in the following. Let us define two quantities for simplicity,

$$\cos \hat{\theta}_k = \frac{s_k^T B_k s_k}{\|s_k\| \, \|B_k s_k\|} \quad \text{and} \quad \hat{q}_k = \frac{s_k^T B_k s_k}{s_k^T s_k}.$$

For these quantities, the following theorem holds if Assumption 3.1 is satisfied.

THEOREM 3.2. *Let $\{B_k\}_{k\in S_1}$ be generated by the BFGS method. Suppose (3.1) and (3.2) hold for any $s_k \neq 0$. Then for any $p \in (0,1]$, there exist constants $\beta_1 > 0$, $\beta_2 > 0$, and $\beta_3 > 0$ such that for any k, the relations*

$$\begin{aligned} \cos\hat{\theta}_i \geq \beta_1 &> 0 \\ 0 < \beta_2 \leq \hat{q}_i &\leq \beta_3 \\ \beta_2 \leq \frac{||B_i s_i||}{||s_i||} &\leq \frac{\beta_3}{\beta_1} \end{aligned}$$

hold for at least $\lceil p|S_1^k|\rceil$ values of $i \in S_1^k$. In other words, the index set J_k in which for any i the above three inequalities hold has at least $\lceil p|S_1^k|\rceil$ elements, i.e. $|J_k| \geq \lceil p|S_1^k|\rceil$.

Theorem 3.2 can be proved by applying the analysis of Theorem 3.1 of Byrd and Nocedal [1] to S_1^k and the proof is omitted. The following two theorems are about the behaviors of the two merit functions, the l_1 and Fletcher merit functions.

THEOREM 3.3. *Suppose $\{x_k\}$ is generated by an RHSQP algorithm using the l_1 merit function with its penalty parameter chosen so that*

$$\mu_k \geq ||\lambda_k||_\infty + \rho \tag{3.3}$$

for all k, where ρ is a positive constant. Then for all k,

$$D\phi_{\mu_k}(x_k; d_k) \leq -\frac{1}{\gamma_Z}||Z_k^T g_k||||h_k|| \cos\hat{\theta}_k - \rho||c_k||_1. \tag{3.4}$$

In addition, for given constants, $\beta_1 > 0$, $\beta_2 > 0$, and $\beta_3 > 0$, there is a constant $\hat{\gamma} > 0$ such that if the conditions

$$\cos\hat{\theta}_k \geq \beta_1 > 0 \tag{3.5}$$

$$0 < \beta_2 \leq \hat{q}_k \leq \beta_3, \tag{3.6}$$

hold for some k, the direction derivative at x_k satisfies,

$$D\phi_{\mu_k}(x_k; d_k) \leq -\gamma_D[||Z_k^T g_k||^2 + ||c_k||_1]. \tag{3.7}$$

Moreover, for any value μ, there is a positive constant γ_μ such that if $\mu_k = \mu$ satisfies (3.3) and if (3.5) and (3.6) hold, then

$$\phi_{\mu_k}(x_k) - \phi_{\mu_k}(x_{k+1}) \geq \gamma_\mu[||Z_k^T g_k||^2 + ||c_k||_1]. \tag{3.8}$$

Proof. The proof of this theorem is similar to that for Lemma 3.3 in [2] but it handles more general basis $Z(x)$ satisfying Assumption 3.1. The main difference is in (3.4) and we prove it as follows.

By the definition of ϕ_{μ_k},

$$D\phi_{\mu_k}(x_k; d_k) = g_k^T d_k - \mu_k||c_k||_1$$

as shown in [2]. Since $v_k^T g_k == c_k^T \lambda_k$, (2.3) and (2.4) implies

$$\begin{aligned} D\phi_{\mu_k}(x_k; d_k) &\leq g_k^T h_k - (\mu_k - ||\lambda_k||_\infty)||c_k||_1 \leq -g_k^T Z_k B_k^{-1} Z_k^T g_k - \rho||c_k||_1 \\ &= -\cos\hat{\theta}_k ||B_k^{-1} Z_k^T g_k|| ||Z_k^T g_k|| - \rho||c_k||_1. \end{aligned}$$

By (2.4) and Assumption 3.1.2, $||h_k|| \le \gamma_Z ||B_k^{-1} Z_k^T g_k||$ and then (3.4) holds. The remaining of this theorem can be proved using the same analysis as [2] and considering the factor γ_Z. □

A corresponding result can be proved for the Fletcher merit function.

THEOREM 3.4. *Suppose $\{x_k\}$ are generated by an RHSQP algorithm using the Fletcher merit function with the penalty parameter chosen so that*

$$\nu_k \ge \frac{d_k^T \nabla\lambda(x_k)c_k + \frac{1}{2} g_k^T h_k}{||c_k||^2} + \rho = \bar{\nu}_k + \rho \tag{3.9}$$

for $k > 0$ and some positive constant $\rho > 0$. Then for all $k \ge 0$,

$$D\Phi_{\nu_k}(x_k; d_k) \le -\frac{1}{\gamma_Z} ||Z_k^T g_k|| ||h_k|| \cos\hat{\theta}_k - \rho ||c_k||^2. \tag{3.10}$$

In addition, for given constants, $\beta_1 > 0$, $\beta_2 > 0$, and $\beta_3 > 0$, there is a constant $\hat{\gamma} > 0$ such that if the conditions (3.5) and (3.6) hold for some k, the direction derivative at x_k satisfies,

$$D\Phi_{\nu_k}(x_k; d_k) \le -\gamma_D [||Z_k^T g_k||^2 + ||c_k||^2]. \tag{3.11}$$

Moreover, for any value ν, there is a positive constant γ_ν such that if $\nu_k = \nu$ satisfies (3.9) and if (3.5) and (3.6) hold, then

$$\Phi_{\nu_k}(x_k) - \Phi_{\nu_k}(x_{k+1}) \ge \gamma_\nu [||Z_k^T g_k||^2 + ||c_k||^2]. \tag{3.12}$$

Proof. The proof is analogous to the previous analysis by considering the general basis matrices and using the directional derivative

$$\nabla\Phi_{\nu_k}(x_k)^T d_k = g_k^T h_k + d_k^T \nabla\lambda_k c_k + \nu_k ||c_k||^2.$$

□

Based on the above two theorems about the two merit functions, the global convergence of RHSQP algorithms using the step secant update strategy is proved.

THEOREM 3.5. *Suppose $\{x_k\}$ is generated by an RHSQP algorithm using the step secant update strategy with any update criteria and using the l_1 and Fletcher merit functions with step 2 in Algorithm 2.1 replaced by step 2′. Suppose Assumption 3.1 and (3.1) and (3.2) are satisfied for all k sufficiently large. For the Fletcher merit function, $\bar{\nu}_k$ is assumed to be bounded above. Then*

$$\lim_{k\to\infty} \inf_{i \le k} \{||Z_i^T g_i|| + ||c_i||\} = 0.$$

Proof. If the l_1 merit function is used, it follows that $\mu_k \equiv \mu$ for some constant $\mu > 0$ and for sufficiently large k because μ_k is chosen by (2.17) and $||\lambda(x)||$ is bounded above. Similarly, if the Fletcher merit function is used, $\nu_k \equiv \nu$ for some constant ν and for k sufficiently large because $\bar{\nu}_k$ is assumed bounded above. Without loss of generality, we assume for any k, $\mu_k = \mu$ and $\nu_k = \nu$.

Suppose $|S_1| = \infty$. Since (3.1) and (3.2) hold for large k, by Theorem 3.2, there are constants β_1, β_2, β_3, and an index set J_k with $|J_k| \ge p|S_1^k|$ for a given constant

$p > 0$ and any k such that for any $j \in J_k$, (3.5) and (3.6) hold. Theorem 3.3 and 3.4 imply,

$$\begin{aligned}\phi_\mu(x_0) - \phi_\mu(x_k) &\geq \gamma_\mu \sum_{j \in J_k} [\|Z_j^T g_j\|^2 + \|c_j\|_1] \\ \Phi_\nu(x_0) - \Phi_\nu(x_k) &\geq \gamma_\nu \sum_{j \in J_k} [\|Z_j^T g_j\|^2 + \|c_j\|^2]\end{aligned}$$

as both $\{\phi_\mu(x_k)\}$ and $\{\Phi_\nu(x_k)\}$ are decreasing sequences. Then

$$\begin{aligned}\sum_{j \in J_k} [\|Z_j^T g_j\|^2 + \|c_j\|_1] &\leq \phi_\mu(x_0) - \min_x \phi_\mu(x) < \infty \\ \sum_{j \in J_k} [\|Z_j^T g_j\|^2 + \|c_j\|^2] &\leq \Phi_\nu(x_0) - \min_x \Phi_\nu(x) < \infty\end{aligned}$$

since the merit functions are bounded below for fixed penalty parameters by Assumption 3.1. Because $|J_k| \geq p|S_1^k| \to \infty$ as $k \to \infty$,

$$\begin{aligned}\lim_{j \in J_k \to \infty} \|Z_j^T g_j\|^2 + \|c_j\|_1 = 0 \\ \lim_{j \in J_k \to \infty} \|Z_j^T g_j\|^2 + \|c_j\|^2 = 0\end{aligned} \quad .$$

In general,

$$\lim_{j \in J_k \to \infty} [\|Z_j^T g_j\| + \|c_j\|] = 0$$

by the equivalence of the l_1 and l_2 norms.

If $|S_1|$ is finite, there is a K_1 large enough so that for any $k > K_1$, $B_k \equiv B_{K_1}$ and thus for all $k \geq K_1$, (3.5) and (3.6) hold for some constants $\beta_1 > 0$, $\beta_2 > 0$, and $\beta_3 > 0$. Similarly by Theorem 3.3 and 3.4, we know that there are constants $\gamma_\mu > 0$ and $\gamma_\nu > 0$ such that for any $k > K_1$

$$\begin{aligned}\phi_\mu(x_{K_1}) - \phi_\mu(x_k) &\geq \gamma_\mu \sum_{j=K_1}^{k} [\|Z_j^T g_j\|^2 + \|c_j\|_1] \\ \Phi_\nu(x_{K_1}) - \Phi_\nu(x_k) &\geq \gamma_\nu \sum_{j=K_1}^{k} [\|Z_j^T g_j\|^2 + \|c_j\|^2].\end{aligned}$$

These two inequalities imply that

$$\lim_{k \to \infty} [\|Z_k^T g_k\| + \|c_k\|] = 0$$

in the case S_1 is finite. □

Note that the convergence result for the Fletcher merit function is somewhat weaker than for the l_1 merit function because of the plausible but optimistic assumption on $\{\bar{\nu}_k\}$.

With global convergence now established, in the next section we discuss the R-linear convergence of the step secant update strategy.

4. Local and R-linear Convergence. In this section R-linear convergence is proved for the RHSQP algorithms using the step secant update strategy with the Nocedal and Overton update strategy and the positive curvature criterion and using either the l_1 merit function or the Fletcher merit function. Although the positive curvature criterion allows more updates than the Nocedal and Overton update criterion and we prove that they are both R-linear convergent, we cannot establish a unified analysis for them. In this section, we present the analysis of R-linear convergence of the two criteria separately because of their different update characters.

4.1. Properties of the local minimizer. Before the analysis of the R-linear convergence, some characteristics of the solution of (1.1) are shown under the following assumption.

ASSUMPTION 4.1. *Let x^* be a local minimizer of (1.1).*

1. *Assumption 3.1 holds on a set D containing x^* in its interior,*
2. *The Matrix $A(x^*)$ is full rank. This implies x^* is a Kuhn-Tucker point. That is, there is a $\lambda^* \in R^t$, the Lagrangian multiplier, such that*

$$\nabla_x L(x^*, \lambda^*) = g(x^*) + A(x^*)\lambda^* = 0. \tag{4.1}$$

3. *The matrix $Z(x^*)^T \nabla^2_{xx} L(x^*, \lambda^*) Z(x^*)$ is positive definite.*
4. *In a neighborhood of x^* the functions, $\lambda(x)$ and $Z(x)$, are Lipschitz continuous, i.e.,*

$$\begin{aligned} ||\lambda(x) - \lambda(z)|| &\leq \gamma_\lambda ||x - z|| && (4.2)\\ ||Z(x) - Z(z)|| &\leq \gamma_z ||x - z|| && (4.3) \end{aligned}$$

where γ_λ and γ_z are constants.

Assumption 4.1.1 and 4.1.3 imply that for any (x, λ) sufficiently close to (x^*, λ^*) and $\delta > 0$ sufficiently small,

$$m_0 ||u||^2 \leq u^T Z(x)^T \nabla^2_{xx} L(x + \Delta x, \lambda) Z(x) u \leq M_0 ||u||^2 \tag{4.4}$$

for some constants $m_0 > 0$ and $M_0 > 0$ with $||\Delta x|| \leq \delta$. That is, the assumptions of Lemma 3.1 are satisfied, and thus (3.1) and (3.2) hold near (x^*, λ^*).

Under Assumption 4.1, the following lemma similar to Lemma 4.1 and 4.2 given by Byrd and Nocedal [2] can be proved with a mild condition, twice differentiability of f and c and the general basis matrix and inverse matrices.

LEMMA 4.1. *If Assumption 4.1 holds, then for x sufficiently close to x^*,*

$$\gamma_1 ||x - x^*|| \leq ||c(x)|| + ||Z(x)^T g(x)|| \leq \gamma_2 ||x - x^*|| \tag{4.5}$$

for some constants $\gamma_1 > 0$ and $\gamma_2 > 0$. In addition, for any $\mu > ||\lambda^*||_\infty$ and for any ν sufficiently large, there are constants $\gamma_3 > 0$, $\gamma_4 > 0$, and $\gamma_5 > 0$, $\gamma_6 > 0$ such that

$$\begin{aligned} \gamma_3 ||x - x^*||^2 &\leq \phi_\mu(x) - \phi_\mu(x^*) \leq \gamma_4 [||Z(x)^T g(x)||^2 + ||c(x)||_1] && (4.6)\\ \gamma_5 ||x - x^*||^2 &\leq \Phi_\nu(x) - \Phi_\nu(x^*) \leq \gamma_6 [||Z(x)^T g(x)||^2 + ||c(x)||^2]. && (4.7) \end{aligned}$$

Proof. By using (2.10) and (3.1) for the general matrix functions, $Z(x)$, $(Z(x))_L^{-1}$, and $A(x)_L^{-1}$, the inequality (4.5) follows the analysis of Lemma 4.1 in [2] because there is no higher than second order derivatives involved. If (4.7) holds, (4.6) follows by using the same technique of Lemma 4.2 in [2]. The analysis in [2] involves the third order derivatives only in the proof of (4.7) itself.

Let us consider (4.7). Since (2.2) also holds on x^*, we can express $x - x^* = h + v$, where $h = Z^*(Z^*)_L^{-1}(x - x^*)$ and $v = {A^*}_L^{-T} {A^*}^T (x - x^*)$. Because $\Phi_\nu(x^*) = L(x^*, \lambda^*)$ and $\nabla_x L(x^*, \lambda^*) = 0$, it follows from Taylor's theorem applied to L and from (4.4), (4.2) and (4.3) that

$$\begin{aligned}
\Phi_\nu(x) - \Phi_\nu(x^*) &= L(x, \lambda(x)) - L(x^*, \lambda^*) + \frac{\nu}{2}||c(x)||^2 \\
&\geq \frac{1}{2}(x - x^*)^T \nabla^2_{xx} L(x^*, \lambda^*)(x - x^*) + (\lambda(x) - \lambda^*)^T c(x) + \\
&\quad + o(||x - x^*||^2) + \frac{\nu}{2}||c(x)||^2 \\
&= \frac{1}{2}(h^T \nabla^2_{xx} L(x^*, \lambda^*) h + 2h^T \nabla^2_{xx} L(x^*, \lambda^*) v + v^T \nabla^2_{xx} L(x^*, \lambda^*) v) + \\
&\quad + (\lambda(x) - \lambda^*)^T c(x) + o(||x - x^*||^2) + \frac{\nu}{2}||c(x)||^2 \\
&\geq \frac{1}{2} m_0 ||h||^2 - M_0 ||h|| ||v|| - \frac{1}{2} M_0 ||v||^2 - \\
&\quad - \gamma_\lambda ||x - x^*|| ||c(x)|| + \frac{\nu}{2}||c(x)||^2 + o(||x - x^*||^2).
\end{aligned}$$

Since $c(x) - c(x^*) = {A^*}^T(x - x^*) + O(||x - x^*||^2)$ and ${A^*}_L^{-1}$ is bounded, it follows that $||v|| \leq \gamma_A ||c(x)|| + O(||x - x^*||^2)$. Thus,

$$\begin{aligned}
\Phi_\nu(x) - \Phi_\nu(x^*) &\geq -\gamma_\lambda(||h|| + \gamma_A ||c(x)||)||c(x)|| + \frac{1}{2} m_0 ||h||^2 - M_0 \gamma_A ||h|| ||c(x)|| \\
&\quad - \frac{1}{2} M_0 \gamma_A ||c(x)||^2 + \frac{\nu}{2}||c(x)||^2 + o(||x - x^*||^2) \\
&= \frac{1}{2} m_0 ||h||^2 + \left(-\gamma_\lambda \gamma_A - \frac{1}{2} M_0 \gamma_A + \frac{\nu}{2}\right) ||c||^2 + \\
&\quad + (-\gamma_\lambda - M_0 \gamma_A) ||h|| ||c|| + o(||x - x^*||^2).
\end{aligned}$$

Consider the above equation as a quadratic polynomial in $||h||$ and $||c||$. There are positive constants $\bar{\nu}$, γ' and γ_5 such that if $\nu > \bar{\nu}$,

$$\Phi_\nu(x) - \Phi_\nu(x^*) \geq \gamma'(||h||^2 + ||v||^2) + o(||x - x^*||^2) \geq \gamma_5 ||x - x^*||^2.$$

Similarly, using the Lipschitz continuity of $\lambda(x)$, $\nabla_x L(x^*, \lambda^*) = 0$ and (4.5),

$$\begin{aligned}
\Phi_\nu(x) - \Phi_\nu(x^*) &= L(x, \lambda(x)) - L(x^*, \lambda^*) + \frac{\nu}{2}||c(x)||^2 \\
&\leq \gamma_\lambda ||x - x^*|| ||c(x)|| + \frac{M_0}{2}||x - x^*||^2 + \\
&\quad + o(||x - x^*||^2) + \frac{\nu}{2}||c(x)||^2 \\
&\leq O(||x - x^*||^2) + \frac{\nu}{2}||c(x)||^2 \\
&\leq O(||Z(x)^T g(x)|| + ||c(x)||)^2 + \frac{\nu}{2}||c(x)||^2 \\
&\leq \gamma_6(||Z(x)^T g(x)||^2 + ||c(x)||^2).
\end{aligned}$$

□

In order to guarantee that $\{x_k\}_{k=1}^{\infty}$ converges to x^*, another assumption is made for the constrained problem,

ASSUMPTION 4.2. *The line search procedure has the property that if x_k is sufficiently close to x^*, then $\forall\theta \in [0,1]$,*

$$\varphi((1-\theta)x_k + \theta x_{k+1}) \leq \varphi(x_k),$$

where φ is the merit function used in RHSQP algorithms. Actually, there is no practical line search strategy that can absolutely guarantee Assumption 4.2 to be satisfied, but it seems unlikely that it is violated when x_k is close to x^*. It is clearly satisfied when φ is quasi-convex. The following theorem shows that Assumption 4.2 implies $\{x_k\} \rightarrow x^*$.

THEOREM 4.2. *Let $\{x_k\}$ be generated by an RHSQP algorithm using the l_1 merit function with μ_k chosen by (2.17) and using either the Nocedal and Overton criterion or the positive curvature criterion. Suppose Assumptions 4.1 and 4.2 hold and $\{\lambda_k\}$ is bounded above. Then for sufficiently large K, μ_k is fixed for $k > K$ and there is a neighborhood of x^* such that if an iterate x_{k_0} with $k_0 > K$ falls in the neighborhood, then $x_k \rightarrow x^*$ and (3.1) and (3.2) hold for all k sufficiently large. If the Fletcher merit function with ν_k chosen by (2.18) is used, the same conclusion holds under the additional assumption that $\bar{\nu}_k$ is bounded and ν_k is large enough.*

Proof. By Assumptions 4.1, there exists $\delta_1 > 0$ such that, for all x in the neighborhood $N_1 = \{x : ||x - x^*|| < \delta_1\}$ of x^*,

$$||\lambda(x)||_\infty + \rho > ||\lambda^*||_\infty, \tag{4.8}$$

and the conditions of Assumption 3.1 hold for $D = N_1$.

Now, since $\{||\lambda(x_k)||_\infty\}$ and $\{\bar{\nu}_k\}$ are bounded, the procedure (2.17) or (2.18) implies that for all k greater than some value $\bar{k}$, μ_k or ν_k are fixed at some values μ and ν. Suppose ν sufficiently large so that (4.7) holds. By (2.17), (2.18) and (4.8), if an iterate x_k, with $k > \bar{k}$, occurs in N_1 then it must be that $\mu > ||\lambda^*||_\infty$. In other words, Lemma 4.1 holds on N_1 and ϕ_μ and Φ_ν have a strict local minimizer x^*. Suppose K is an integer such that $\mu_k = \mu$ or $\nu_k = \nu$ for any $k > K$. For such μ and ν, it follows from Lemma 4.1 that there exists $\delta_2 \in (0, \delta_1]$ such that if $||x_{k_0} - x^*|| < \delta_2$ for $k_0 > K$, the connected component of the level set $\{z : \phi_\mu(z) < \phi_\mu(x_{k_0})\}$ or $\{z : \Phi_\nu(z) < \Phi_\nu(x_{k_0})\}$ containing x^* is a subset N_2 of N_1. Since N_2 is connected, by Assumption 4.2, all iterates x_k for $k > k_0$ remain in N_2. If δ_2 is chosen sufficiently small, then by Assumption 4.1 the hypotheses of Lemma 3.1 hold for $D = N_2$ and therefore (3.1) and (3.2) hold at all update steps for $k > k_0$. Then the assumptions of Theorem 3.5 are satisfied for $k > \hat{k}$ and thus there is a subsequence $\{x_{k_i}\}$ of $\{x_k\}_{k=k_0}^{\infty}$ such that

$$\lim_{i\rightarrow\infty} [||Z_{k_i}^T g_{k_i}|| + ||c_{k_i}||] = 0.$$

By (4.5), (4.6) and (4.7), we have

$$\begin{aligned} \lim_{i\rightarrow\infty} \phi_\mu(x_{k_i}) - \phi_\mu(x^*) &= 0 \\ \lim_{i\rightarrow\infty} \Phi_\nu(x_{k_i}) - \Phi_\nu(x^*) &= 0, \end{aligned}$$

and the decreasing property of $\{\phi_\mu\}$ or $\{\Phi_\nu\}$,

$$\begin{aligned}\lim_{i\to\infty} \phi_\mu(x_k) - \phi_\mu(x^*) &= 0\\ \lim_{i\to\infty} \Phi_\nu(x_k) - \Phi_\nu(x^*) &= 0.\end{aligned}$$

Lemma 4.1 implies $x_k \to x^*$. The neighborhood satisfying this theorem is N_2. □

Based on this theorem, we can show the R-linear convergence under the hypothesis $\{x_k\} \to x^*$ for the Nocedal and Overton criterion and the positive curvature criterion, respectively. First, we show that both criteria have an R-linear convergent subsequence and the remaining subsequence is discussed separately for these two criteria.

4.2. A subsequential R-linear convergence. Let us define a subsequential R-linear convergence. It means that there is a subsequence $S \subset [1, \cdots, \infty)$ such that for any k and $S^k = S \bigcap [1, \cdots, k]$, $\|x_k - x^*\| \le r^{|S^k|}$. We call $\{x_k\}_{k=1}^{\infty}$ S R-linear convergent. We show that both criteria generate a S_1 R-linear convergent sequence.

LEMMA 4.3. *Suppose $\{x_k\}$ is generated by an RHSQP algorithm using the step secant update strategy with either the Nocedal and Overton criterion or the positive curvature criterion and using the l_1 merit function or the Fletcher merit function. Then $\{x_k\}$ converges S_1 R-linearly if the hypotheses of Theorem 4.2 are satisfied.*

Proof. Since $x_k \to x^*$ by Theorem 4.2, (3.1) and (3.2) hold and Lemma 4.1 also holds for $x = x_k$, if k is large enough. Without loss of generality, assume that these lemmas hold for any k. Choose $p = \frac{1}{2}$ and apply Theorems 3.2, 3.3, and 3.4 to the RHSQP algorithm. Then for the index set J_k defined in Theorem 3.2

$$\begin{aligned}\phi_\mu(x_i) - \phi_\mu(x_{i+1}) &\ge \gamma_\mu[\|Z_i^T g_i\|^2 + \|c_i\|_1], \quad \forall i \in J_k\\ \Phi_\nu(x_i) - \Phi_\nu(x_{i+1}) &\ge \gamma_\nu[\|Z_i^T g_i\|^2 + \|c_i\|^2].\end{aligned}$$

By (4.6) or (4.7), the above inequalities imply

$$\begin{aligned}\phi_\mu(x_i) - \phi_\mu(x_{i+1}) &\ge \frac{\gamma_\mu}{\gamma_4}(\phi_\mu(x_i) - \phi_\mu(x^*)) \quad \forall i \in J_k\\ \Phi_\nu(x_i) - \Phi_\nu(x_{i+1}) &\ge \frac{\gamma_\nu}{\gamma_6}(\Phi_\nu(x_i) - \Phi_\nu(x^*)) \quad \forall i \in J_k.\end{aligned}$$

Then

$$\begin{aligned}\phi_\mu(x_{i+1}) - \phi_\mu(x^*) &\le (1 - \frac{\gamma_\mu}{\gamma_4})(\phi_\mu(x_i) - \phi_\mu(x^*))\\ \Phi_\nu(x_{i+1}) - \Phi_\nu(x^*) &\le (1 - \frac{\gamma_\nu}{\gamma_6})(\Phi_\nu(x_i) - \Phi_\nu(x^*)).\end{aligned}$$

Let $r' = (1 - \frac{\gamma_\mu}{\gamma_4})^{\frac{1}{4}} < 1$ for the l_1 or $r' = (1 - \frac{\gamma_\nu}{\gamma_6})^{\frac{1}{4}} < 1$ for the Fletcher merit functions and choose $r'' = \frac{1}{\gamma_3}(\phi_\mu(x_0) - \phi_\mu(x^*))^{\frac{1}{2}} > 0$ for the l_1 and $r'' = \frac{1}{\gamma_5}(\Phi_\nu(x_0) - \Phi_\nu(x^*))^{\frac{1}{2}} > 0$ for the Fletcher merit function. Then for any $i \in J_k$,

$$\begin{aligned}\phi_\mu(x_{i+1}) - \phi_\mu(x^*) &\le r'^4(\phi_\mu(x_i) - \phi_\mu(x^*)), \quad \text{and}\\ \Phi_\nu(x_{i+1}) - \Phi_\nu(x^*) &\le r'^4(\Phi_\nu(x_i) - \Phi_\nu(x^*)).\end{aligned}$$

and then by the decreasing properties of $\{\phi_\mu(x_i)\}$ and by (4.6),

$$\begin{aligned} \|x_k - x^*\| &\leq \frac{1}{\gamma_3}(\phi_\mu(x_k) - \phi_\mu(x^*))^{\frac{1}{2}} \\ &\leq \frac{1}{\gamma_3}(r'^{4|J_k|}(\phi_\mu(x_0) - \phi_\mu(x^*)))^{\frac{1}{2}} \\ &\leq \frac{1}{\gamma_3}(r'^{2|S_1^k|}(\phi_\mu(x_0) - \phi_\mu(x^*)))^{\frac{1}{2}} \\ &= r'' r'^{|S_1^k|} \end{aligned}$$

because $p = \frac{1}{2}$ and $|J_k| \geq p|S_1^k|$. For the Fletcher merit function, it holds similarly. This implies that there is a constant $r \in (0,1)$ such that

$$\|x_k - x^*\| \leq r^{|S_1^k|}$$

for both merit functions. □

Note that this analysis can be applied to any update criterion with (3.1) and (3.2) satisfied. By the S_1 R-linear convergence, the sequence is R-linearly convergent if:

- there is a constant $p > 0$ such that $|S_1^k| \geq pk$ for any k; or
- $|S_1|$ is finite because B_k will be a fixed matrix for large k and the proof of Lemma 4.3 can be applied to S_2.

The extreme case is that neither of these holds and the S_2 R-linear convergence has to be proved. Because the differences of S_2 between the Nocedal and Overton update criterion and the positive curvature criterion, we prove their R-linear convergence separately.

4.3. The Nocedal and Overton criterion. For the Nocedal and Overton update criterion, one can prove the matrices $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded and then it is not difficult to prove its R-linear convergence. We need to show only the boundness of $\{B_k\}$ and $\{B_k^{-1}\}$ for the Nocedal and Overton update criterion.

Consider the scaled version of the matrix function $\psi(\cdot)$ developed by Byrd and Nocedal [1] for the quasi-Newton methods. The ψ function is defined as

$$\psi(B) = \mathrm{Tr}(H^{*-\frac{1}{2}} B H^{*-\frac{1}{2}}) - \ln\det(H^{*-\frac{1}{2}} B H^{*-\frac{1}{2}}) \tag{4.9}$$

where $H^* = Z^{*T} \nabla_{xx}^2 L(x^*, \lambda^*) Z^* > 0$. In order to discuss the boundness of B_k and B_k^{-1} using ψ, we define the quantities $\cos\theta_k$ and q_k, which are scaled versions of the quantities $\cos\hat{\theta}_k$ and $\hat{q}_k$, used for the global convergence analysis.

$$\cos\theta_k = \frac{s_k^T B_k s_k}{\|H^{*\frac{1}{2}} s_k\| \|H^{*-\frac{1}{2}} B_k s_k\|} \qquad q_k = \frac{s_k^T B_k s_k}{s_k^T H^* s_k}. \tag{4.10}$$

Now we estimate $\psi(B_{k+1})$ by the following lemma.

LEMMA 4.4. *When x_k and x_{k+1} are close to x^* and $k \in S_1$,*

$$\psi(B_{k+1}) \leq \psi(B_k) - \frac{q_k}{\cos^2\theta_k} + \ln q_k + 1 + \tilde{\gamma} \quad \text{and} \tag{4.11}$$

$$\psi(B_{k+1}) \leq \psi(B_k) - \frac{q_k}{\cos^2\theta_k} + \ln q_k + 1 + L_0\sigma_k + \tilde{\gamma}\omega_k, \tag{4.12}$$

where L_0 *and* $\tilde{\gamma}$ *are constants,* $\omega_k = ||\alpha_k c_k||/||s_k||$ *and* $\sigma_k = \max\{||e_{k+1}||, ||e_k||\}$ *with* $e_k = x_k - x^*$.

Proof. By a result of Pearson [12] for the BFGS update

$$\det(H^{*-\frac{1}{2}} B_{k+1} H^{*-\frac{1}{2}}) = \det(H^{*-\frac{1}{2}} B_k H^{*-\frac{1}{2}}) \frac{s_k^T y_k}{s_k^T B_k s_k}. \tag{4.13}$$

By the definition of ψ, (4.13) and (4.10),

$$\begin{aligned} \psi(B_{k+1}) &= \operatorname{Tr}(B_k) - \operatorname{Tr}\left(H^{*-\frac{1}{2}} \left(\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \right) H^{*-\frac{1}{2}} \right) - \\ &\quad - \ln\det(H^{*-\frac{1}{2}} B_k H^{*-\frac{1}{2}}) + \ln \frac{s_k^T B_k s_k}{s_k^T y_k} \\ &= \psi(B_k) - \frac{||H^{*-\frac{1}{2}} B_k s_k||^2}{s_k^T B_k s_k} + \frac{y_k^T H^{*-1} y_k}{s_k^T y_k} + \ln \frac{s_k^T B_k s_k}{s_k^T y_k} \\ &= \psi(B_k) - \frac{q_k}{\cos^2\theta_k} + \frac{y_k^T H^{*-1} y_k}{s_k^T y_k} - \ln \frac{s_k^T y_k}{s_k^T H^* s_k} + \ln q_k. \end{aligned} \tag{4.14}$$

Then by Assumption 4.1, the conditions (3.1) and (3.2) hold on S_1. Thus

$$\begin{aligned} \psi(B_{k+1}) &\le \psi(B_k) - \frac{q_k}{\cos^2\theta_k} + \ln q_k + 1 + \\ &\quad + \left(||H^{*-1}||M - 1 - \ln \frac{m}{||H^*||} \right). \end{aligned}$$

That is, (4.11) holds. To prove (4.12), we need to estimate the third and fourth terms in the last equation of (4.14). Let

$$H_k = Z_k^T G_k Z_k \quad \tilde{H}_k = Z_k^T G_k (A_k)_L^{-T},$$

and then by Taylor's theorem and (2.2),

$$\begin{aligned} y_k &= Z_k^T [\nabla_x L(x_{k+1}, \lambda_k) - \nabla_x L(x_k, \lambda_k)] \\ &= Z_k^T [G_k \alpha_k d_k] + O(||\alpha_k d_k||^2) \\ &= Z_k^T \left[G_k (Z_k (Z_k)_L^{-1} + (A_k)_L^{-T} A_k^T) \alpha_k d_k \right] + O(||\alpha_k d_k||^2) \\ &= H_k s_k - \tilde{H}_k \alpha_k c_k + O(||\alpha_k d_k||^2) \\ &= H^* s_k - \tilde{H}_k (\alpha_k c_k) + O(||e_k|| ||s_k|| + ||\alpha_k d_k||^2). \end{aligned} \tag{4.15}$$

To estimate the third term of (4.12), we multiply both sides of (4.15) by $y_k^T H^{*-1}$ and yield

$$\begin{aligned} y_k^T H^{*-1} y_k &= s_k^T y_k - y_k^T H^{*-1} \tilde{H}_k (\alpha_k c_k) + O(||e_k|| ||s_k|| + ||\alpha_k d_k||^2) ||y_k|| \\ &\le s_k^T y_k + O(||\alpha_k c_k||) ||y_k|| + O(||e_k|| ||s_k|| + ||\alpha_k d_k||^2) ||y_k||. \end{aligned}$$

By (3.1) and (3.2), we have $s_k^T y_k \ge mM||s_k|| ||y_k||$ and then

$$\frac{y_k^T H^{*-1} y_k}{s_k^T y_k} \le 1 + O\left(\frac{||\alpha_k c_k||}{||s_k||} \right) + O(||e_k||) + O\left(\frac{||\alpha_k d_k||^2}{||s_k||} \right).$$

Since $\alpha_k d_k \le O(\sigma_k)$ and

$$\begin{aligned}\frac{||\alpha_k d_k||^2}{||s_k||} &= O(\sigma_k)\frac{||\alpha_k d_k||}{||s_k||} \le O(\sigma_k)\frac{||\alpha_k h_k|| + ||\alpha_k v_k||}{\sqrt{s_k^T y_k}} \\ &= O(\sigma_k) + O\left(\sigma_k \frac{||\alpha_k c_k||}{||s_k||}\right),\end{aligned}$$

therefore

$$\frac{y_k^T H^{*-1} y_k}{s_k^T y_k} \le 1 + O\left(\frac{||\alpha_k c_k||}{||s_k||}\right) + O(\sigma_k). \tag{4.16}$$

Similarly, to estimate the fourth term, we multiply both sides of (4.15) by s_k^T ,

$$\begin{aligned}s_k^T y_k &= s_k^T H^* s_k - s_k^T \tilde{H}_k(\alpha_k c_k) + O(||e_k|| ||s_k|| + ||\alpha_k d_k||^2)||s_k||, \\ \frac{s_k^T y_k}{s_k^T H^* s_k} &= 1 + \frac{1}{s_k^T H^* s_k}[-s_k^T \tilde{H}_k(\alpha_k c_k) + O(||e_k|| ||s_k|| + ||\alpha_k d_k||^2)||s_k||] \\ &= 1 + O\left(\frac{||\alpha_k c_k||}{||s_k||}\right) + O(\sigma_k).\end{aligned}$$

Since $\frac{s_k^T H^* s_k}{s_k^T y_k} \le \frac{||H^*||}{m}$, we have

$$\begin{aligned}-\ln \frac{s_k^T y_k}{s_k^T H^* s_k} &= \ln \frac{s_k^T H^* s_k}{s_k^T y_k} \\ &\le \frac{s_k^T H^* s_k}{s_k^T y_k} - 1 \\ &= \frac{s_k^T H^* s_k}{s_k^T y_k}\left(1 - \frac{s_k^T y_k}{s_k^T H^* s_k}\right) \\ &\le \frac{||H^*||}{m}\left|\frac{s_k^T y_k}{s_k^T H^* s_k} - 1\right| \\ &\le O\left(\frac{||\alpha_k c_k||}{||s_k||}\right) + O(\sigma_k)\end{aligned} \tag{4.17}$$

Using (4.16) and (4.17) and the definition of ω_k, we know there exist constants L_0 and $\tilde{\gamma}$ satisfying this lemma. □

From Lemmas 4.3 and 4.4, it follows that for any update criterion satisfying

$$\sum_{j \in S_1} \frac{||v_j||}{||h_j||} < \infty, \tag{4.18}$$

the quasi-Newton matrices and their inverses are bounded below and above. This is because $||\alpha_k c_k||/||s_k|| \le (\gamma_A/\gamma_Z)(||v_k||/||h_k||)$ and

$$\begin{aligned}\psi(B_k) &\le \psi(B_0) + L_0 \sum_{j \in S_1^{k-1}} \sigma_j + \tilde{\gamma} \sum_{j \in S_1^{k-1}} \left(\frac{\gamma_A}{\gamma_Z}\frac{||v_j||}{||h_j||}\right) \\ &\le \psi(B_0) + L_0 \sum_{j \in S_1^{k-1}} r^{|S_1^j|} + \tilde{\gamma} \sum_{j \in S_1^{k-1}} \left(\frac{\gamma_A}{\gamma_Z}\frac{||v_j||}{||h_j||}\right) < \infty\end{aligned}$$

since $-q_k/\cos^2\theta_k + \ln q_k + 1 \le 0$. That is, there is a $\bar{\psi} > 0$ such that $\psi(B_k) \le \bar{\psi} < \infty$ which implies that $||B_k||$ and $||B_k^{-1}||$ are bounded and we have the following theorem.

THEOREM 4.5. *Suppose $\{x_k\}$ is generated by an RHSQP algorithm with an update criterion satisfying (4.18) using the l_1 or the Fletcher merit function with $\bar{\nu}_k$ bounded. If Assumptions 4.1 and 4.2 hold, then $\{||B_k||\}$ and $\{||B_k^{-1}||\}$ are bounded above and $\{x_k\}$ converges to x^* R-linearly.* □

The Nocedal and Overton update criterion satisfies (4.18). Actually, there are other criteria satisfying (4.18), for example, (4.18) will hold if

$$||v_k|| \le \frac{\zeta}{|S_1^k|^{1+\epsilon}}||h_k|| \tag{4.19}$$

whenever B_{k+1} is updated, where ζ and ϵ are positive constants.

COROLLARY 4.6. *The sequence $\{x_k\}$ is generated by an RHSQP algorithm using the l_1 or the Fletcher merit function and the Nocedal and Overton update criterion or the criterion given by (4.19). If Assumptions 4.1 and 4.2 hold, then $\{x_k\}_{k=1}^{\infty}$ converges to x^* R-linearly.* □

4.4. R-linear convergence using the positive curvature criterion. Unlike the Nocedal and Overton update criterion, the positive curvature criterion allows updates even $||v_k||/||h_k||$ is not small. The analysis of Theorem 4.4 cannot be applied. Without assuming that $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded, the sufficient reductions of the merit functions by (3.4) and (3.10) cannot rely on the terms involving $\cos\theta_k$ because based on the current analysis, $\cos\theta_k$ cannot be proved to be bounded away from zero even though numerically it rarely happens that $\cos\theta_k$ tends to be unbounded. Fortunately, the positive curvature criterion guarantees that $||c_k||$ is relatively large corresponding to $||h_k||$ for any $k \in S_2$.

LEMMA 4.7. *If Algorithm 2.1 is used with the positive curvature criterion and the conditions of Theorem 4.2 are satisfied, then there are constants $\gamma_8 > 0$ and $\gamma_9 > 0$ such that for sufficiently large k,*

$$\begin{aligned} ||v_k|| &\le \gamma_8||h_k|| \qquad k \in S_1 \\ ||h_k|| &\le \gamma_9||v_k|| \qquad k \in S_2. \end{aligned}$$

Proof. By (2.2) and (4.15),

$$s_k^T y_k \le O(||s_k||^2 + ||s_k||||\alpha_k v_k||) + ||s_k||O(||s_k|| + ||\alpha_k v_k||)^2.$$

Thus, based on (2.13), for $k \in S_1$,

$$\zeta_1||\alpha_k v_k||^2 \le ||\alpha_k h_k||O(||s_k|| + ||\alpha_k v_k||) + ||\alpha_k h_k||O(||s_k|| + ||\alpha_k v_k||)^2.$$

Either $||h_k|| \le ||v_k||$, which implies,

$$||v_k|| \le ||h_k||O(||v_k||) \le ||h_k||O(\gamma_A \sup_D ||c(x)||) = O(||h_k||),$$

or $||v_k|| \le ||h_k||$ shows the existence of the constant $\gamma_8 > 0$.

For the second part of this lemma, the existence of $\gamma_9 > 0$ can be proved as follows. Because for any $k \in S_2$, Lemma 3.1 and (2.14) imply

$$m||s_k||^2 \le s_k^T y_k \le \zeta_2||\alpha_k v_k||^2$$

if Assumption 4.1 holds,

$$\begin{aligned} s_k^T y_k &\geq s_k^T H_k s_k + s_k^T \tilde{H}_k(\alpha_k v_k) - O(\|x_{k+1} - x_k\|^3) \\ &\geq m_0\|s_k\|^2 - M_0\gamma_Z\gamma_A \sup_{x\in D} \|A(x)\|\|\alpha_k v_k\|\|s_k\| - O(\|\alpha_k d_k\|^3). \end{aligned}$$

Then for $k \in S_2$,

$$\zeta_2\|\alpha_k v_k\|^2 \geq s_k^T y_k \geq m_0\|s_k\|^2 - M_0\gamma_Z\gamma_A \sup_{x\in D} \|A(x)\|\|\alpha_k v_k\|\|s_k\| - O(\|d_k\|^3).$$

Since $\{x_k\} \to x^*$ and $v_k \to 0$, $\|s_k\|$ is small for large k. Either

$$M_0\gamma_Z\gamma_A \sup_{x\in D} \|A(x)\|\frac{\|\alpha_k v_k\|}{\|s_k\|} \leq \frac{m_0}{3},$$

which shows

$$\begin{aligned} \zeta_2\|\alpha_k v_k\|^2 &\geq \frac{2m_0}{3}\|s_k\|^2 - O(\|s_k\|^3) \geq \frac{m_0}{2}\|s_k\|^2 \\ \sqrt{\zeta_2}\|\alpha_k v_k\| &\geq \sqrt{\frac{m_0}{2}}\|s_k\| \geq \sqrt{\frac{m_0}{2}}\frac{\|\alpha_k h_k\|}{\gamma_Z} \end{aligned}$$

or $M_0\gamma_Z\gamma_A \sup_{x\in D} \|A(x)\|\frac{\|\alpha_k v_k\|}{\|s_k\|} > \frac{m_0}{3}$, which implies

$$(M_0\gamma_Z\gamma_A \sup_{x\in D} \|A(x)\|)\|v_k\| \geq \frac{m_0}{3}\|B_k^{-1}Z_k^T g_k\| \geq \frac{m_0}{3\gamma_Z}\|h_k\|.$$

Then γ_9 exists. □

To prove the R-linear convergence for the positive curvature criterion, we mainly concentrate the reductions in the vertical direction. We show that $\alpha_k = 1$ for $k \in S_2$ sufficiently large if an RHSQP algorithm uses the positive curvature criterion and either the l_1 or Fletcher merit function in the following lemmas.

LEMMA 4.8. *If the conditions of Theorem 4.2 are satisfied and the l_1 merit function is used in Algorithm 2.1 with the positive curvature criterion to generate a sequence of $\{x_k\}$, then for any $k \in S_2$ large enough, $\alpha_k = 1$.*

Proof. For $\alpha_k < 1$ for $k \in S_2$ in the backtracking line search, we show the reduction of the l_1 merit function is greater than a positive constant and then this implies there are a finite number $k \in S_k$ such that $\alpha_k < 1$ based on the boundness assumption of the l_1 merit function.

Suppose $\alpha_k < 1$ for $k \in S_2$. That means the line search fails for step length $\tilde{\alpha}$ and $\alpha_k \geq \tau\tilde{\alpha}$. This means

$$\phi_\mu(x_k + \tilde{\alpha}d_k) - \phi_\mu(x_k) \quad > \quad \eta\tilde{\alpha}D\phi_\mu(x_k; d_k).$$

On the other hand by the Taylor expansion,

$$\phi_\mu(x_k + \tilde{\alpha}d_k) - \phi_\mu(x_k) \quad \leq \quad \tilde{\alpha}D\phi_\mu(x_k; d_k) + O(\tilde{\alpha}^2\|d_k\|^2).$$

Thus

$$-(1-\eta)D\phi_\mu(x_k; d_k) \quad < \quad \tilde{\alpha}O(\|d_k\|^2) \leq \tilde{\alpha}\gamma_{10}\|d_k\|^2, \tag{4.20}$$

and furthermore, we have an estimation of α_k

$$\alpha_k \geq \tau\tilde{\alpha} > -\tau(1-\eta)D\phi_\mu(x_k;d_k)/(\gamma_{10}\|d_k\|^2).$$

Using (3.4),

$$\begin{aligned}
\alpha_k &\geq \tau(1-\eta)(\frac{1}{\gamma_Z}\|Z_k^T g_k\|\|h_k\|\cos\hat{\theta}_k + \rho\|c_k\|_1)/(\gamma_{10}\|d_k\|^2) \\
&\geq \tau(1-\eta)\rho\|c_k\|_1/(\gamma_{10}\|d_k\|^2) \\
&\geq \tau(1-\eta)\rho\|c_k\|_1/(\gamma_{10}(1+\gamma_9)^2\|v_k\|^2 \\
&\geq \frac{\gamma_{11}}{\|c_k\|}.
\end{aligned}$$

Since $c_k \to 0$, this contradicts the assumption that $\alpha_k < 1$. Thus, for large k, $\alpha_k = 1$. □

For the Fletcher merit function, a stronger condition on the penalty parameter must be added to force $\alpha_k = 1$.

LEMMA 4.9. *Suppose the conditions of Theorem 4.2 are satisfied and the Fletcher merit function is used in Algorithm 2.1 with the positive curvature criterion, then there is a constant $\tilde{\nu} > 0$ such that if the penalty parameter ν_k is greater than $\tilde{\nu}$, $\alpha_k = 1$ for any $k \in S_2$ large enough.*

Proof. Since the Fletcher merit function is differentiable,

$$\nabla\Phi_\nu(x) = g(x) + \nabla\lambda(x)c(x) + A(x)\lambda(x) + \nu A(x)c(x)$$

and by using the relation $\lambda(x_k)^T c(x_k) = g_k^T v_k$,

$$\nabla\Phi_\nu(x_k)^T d_k = g_k^T h_k + d_k^T \nabla\lambda(x_k)c_k - \nu c_k^T c_k.$$

By Lemma 4.7, $d_k \to 0$ for $k \in S_2$ as $x_k \to 0$. Thus, by noticing $\lambda_{k+1} - \lambda_k \to 0$ and $c_{k+1} = O(\|d_k\|^2)$, the Taylor expansion of the Lagrangian function gives

$$\begin{aligned}
&\Phi_\nu(x_{k+1}) - \Phi_\nu(x_k) - \eta\nabla\Phi_\nu(x_k)^T d_k \\
= \; & f(x_{k+1}) + \lambda_k^T c_{k+1} + \frac{\nu}{2}c_{k+1}^T c_{k+1} - (f(x_k) + \lambda_k^T c_k + \frac{\nu}{2}c_k^T c_k) + \\
&+(\lambda_{k+1} - \lambda_k)^T c_{k+1} - \eta(g_k^T h_k + d_k^T\nabla\lambda(x_k)c_k - \nu c_k^T c_k) \\
\leq \; & g_k^T d_k + \frac{1}{2}d_k^T\nabla^2 f_k d_k + \lambda_k^T A_k^T d_k + \frac{1}{2}d_k^T\sum_i(\lambda_k)_i\nabla^2 c_i(x_k)d_k + o(\|d_k\|^2) - \\
&-\frac{\nu}{2}c_k^T c_k - \eta(g_k^T h_k + d_k^T\nabla\lambda(x_k)c_k - \nu c_k^T c_k) \\
= \; & (1-\eta)g_k^T h_k - \eta d_k^T\nabla\lambda(x_k)c_k + \\
&+d_k^T\nabla_{xx}^2 L(x_k, \lambda(x_k))d_k - (\frac{1}{2} - \eta)\nu c_k^T c_k + o(\|d_k\|^2) \\
\leq \; & -\eta\bar{\nu}_k\|c_k\|^2 + M_0\|d_k\|^2 - (\frac{1}{2} - \eta)\nu c_k^T c_k + o(\|d_k\|^2).
\end{aligned}$$

Because Lemma 4.7 implies $\|h_k\| \leq \gamma_9\gamma_A\|c_k\|$ for any $k \in S_2$,

$$\Phi_\nu(x_{k+1}) - \Phi_\nu(x_k) - \eta\nabla\Phi_\nu(x_k)^T d_k$$

$$\leq -\eta\bar{\nu}_k||c_k||^2 + M_0(1+\gamma_9)^2\gamma_A^2||c_k||^2 - (\frac{1}{2}-\eta)\nu c_k^T c_k + o(||c_k||^2)$$
$$= -(\nu(\frac{1}{2}-\eta) - \eta\bar{\nu}_k - M_0(1+\gamma_9)^2\gamma_A^2 - \frac{1}{2})||c_k||^2 - \frac{1}{2}||c_k||^2 + o(||c_k||^2)$$
$$\leq 0$$

for k large enough and $\nu \geq \tilde{\nu} > 0$ where $\tilde{\nu}$ is a constant satisfying for any k,

$$\tilde{\nu} \geq \frac{-\eta\bar{\nu}_k - M_0(1+\gamma_9)^2\gamma_A^2 - \frac{1}{2}}{\frac{1}{2}-\eta} \tag{4.21}$$

since $|\bar{\nu_k}| \leq sup||\nabla\lambda(x)||||d_k||||c_k||/||c_k||^2 \leq sup||\nabla\lambda(x)||(1+\gamma_9\gamma_A)$ for $k \in S_2$. That is, for $k \in S_2$ large enough, $\alpha_k = 1$ is accepted by the line search for the Fletcher merit function. □

Given these results, we can show the R-linear convergence of the RHSQP algorithms using the positive curvature critérion. First of all, we have an estimate of the $||e_k||$ as follows.

LEMMA 4.10. *If the conditions of Theorem 4.2 hold and Algorithm 2.1 is used with the positive curvature criterion and either the l_1 merit function with (2.17) or the Fletcher merit function with (2.18) and ν_k eventually sufficiently large, then for any index set $\mathcal{S} \subset [1, 2, \cdots, k-1]$,*

$$||e_k|| \leq \gamma_{12}^{|\mathcal{S}|}\prod_{j\in\mathcal{S}}\frac{||e_{j+1}||}{||e_j||}.$$

Proof. Without loss of generality, assume $||e_1|| = 1$. Because Lemma 4.1, then for k' the largest index in $\mathcal{S}$

$$||e_k||^2 \leq \frac{1}{\gamma_5}(\Phi_\nu(x_k) - \Phi_\nu(x^*)) \leq \frac{1}{\gamma_5}(\Phi_\nu(x_{k'+1}) - \Phi_\nu(x^*))$$

as the sequence of $\{\Phi_\nu(x_j)\}$ is decreasing. By (4.5) and (4.7),

$$||e_k||^2 \leq \frac{\gamma_2\gamma_6}{\gamma_5}||e_{k'+1}||^2$$
$$||e_k|| \leq \left(\frac{\gamma_2\gamma_6}{\gamma_5}\right)^{\frac{1}{2}}\frac{||e_{k'+1}||}{||e_{k'}||}||e_{k'}||.$$

Therefore applying the same procedure to the second largest index k'' in $\mathcal{S}$ and so on, we have

$$||e_k|| \leq \left(\frac{\gamma_2\gamma_6}{\gamma_5}\right)^{\frac{1}{2}|\mathcal{S}|}\prod_{j\in\mathcal{S}}\frac{||e_{j+1}||}{||e_j||},$$

$\gamma_{12} = \left(\frac{\gamma_2\gamma_6}{\gamma_5}\right)^{\frac{1}{2}}$ is a constant satisfying this lemma. □

To show R-linear convergence, we need only considering the situation with $|S_2^k| \geq \frac{3}{4}k$. If $|S_2^k| < \frac{3}{4}k$, the S_1 R-linear convergence implies that the sequence $\{x_k\}$ is R-linear convergent. As $|S_2^k| \geq \frac{3}{4}k$, the index set

$$\bar{S}^k = \{ \ j \ | \ j, j+1 \in S_2^k \ \}$$

contains at least $\frac{1}{4}k$ elements; otherwise $|S_2^k| < 2 \times \frac{1}{4}k + |S_1^k| \leq \frac{1}{2}k + \frac{1}{4}k = \frac{3}{4}k$. Moreover for any $j \in \bar{S}^k$ the following estimation is true.

LEMMA 4.11. *Under the conditions of Lemma 4.10, for any positive constant* $\epsilon > 0$ *and any sufficiently large* k *and* $j \in \bar{S}^k$ *such that* $\alpha_k = 1$,

$$\frac{||e_{j+1}||}{||e_j||} \leq \epsilon.$$

Proof. Using the fact that

$$-\frac{q_k}{\cos^2\theta_k} + \ln q_k + 1 \leq 0$$

in the inequality (4.11), we can obtain a growth bound of $||B_k||$ and $||B_k^{-1}||$, which is

$$||B_k|| \leq \gamma |S_1^k|, \qquad ||B_k^{-1}|| \leq \gamma |S_1^k|$$

for some constant γ. For any constant τ with $1 > \tau > 0$, Theorem 4.3 implies

$$||e_k||^\tau ||B_k^{-1}||^{1-\tau} \leq r^{\tau|S_1^k|}\gamma|S_1^k|^{1-\tau} \leq 1$$

for k sufficiently large. Since $\alpha_j = 1$ and $j \in S_2$ large, $c_{j+1} = O(||d_j||^2)$ and by Lemma 4.1,

$$\begin{aligned}
||e_{j+1}|| &\leq \frac{1}{\gamma_1}(||Z_{j+1}^T g_{j+1}|| + ||c_{j+1}||) \\
&\leq \frac{1}{\gamma_1}(||Z_{j+1}^T g_{j+1}||^\tau ||Z_{j+1}^T g_{j+1}||^{1-\tau} + ||c_{j+1}||) \\
&\leq \frac{1}{\gamma_1}(||Z_{j+1}^T g_{j+1}||^\tau ||B_{j+1}^{-1}||^{1-\tau} ||h_{j+1}||^{1-\tau} + ||c_{j+1}||) \\
&\leq \frac{1}{\gamma_1}(||h_{j+1}||^{1-\tau} + ||c_{j+1}||) \\
&\leq \frac{1}{\gamma_1}(\gamma_8^{1-\tau}||c_{j+1}||^{1-\tau} + ||c_{j+1}||) \\
&\leq O(||c_{j+1}||^{1-\tau}) \leq O(||d_j||^{2(1-\tau)}) \\
&\leq O(||h_j||^{2(1-\tau)} + ||c_j||^{2(1-\tau)}) \\
&\leq O(||c_j||^{2(1-\tau)}) \leq O(||e_j||^{2(1-\tau)}).
\end{aligned}$$

Since the assumptions of Theorem 4.2 are satisfied, $||e_k|| \to 0$. Therefore as long as $\tau < \frac{1}{2}$ is chosen, for any given constant $\epsilon > 0$,

$$\frac{||e_{j+1}||}{||e_j||} \leq \epsilon$$

for k large enough. □

Using Lemma 4.10 and 4.11,

$$||e_k|| \leq \gamma_{12}^{|\bar{S}^k|} \prod_{j \in \bar{S}^k} \frac{||e_{j+1}||}{||e_j||}$$

$$\begin{aligned}
&\leq \gamma_{12}^{|\bar{S}^k|} \prod_{j\in\bar{S}^k} \epsilon \\
&\leq \gamma_{12}^{|\bar{S}^k|}\epsilon^{|\bar{S}^k|} \\
&\leq (\gamma_{12}\epsilon)^{|\bar{S}^k|} \\
&\leq (\gamma_{12}\epsilon)^{\frac{1}{4}k} \\
&\leq ([\gamma_{12}\epsilon]^{\frac{1}{4}})^k
\end{aligned}$$

i.e. $||e_k|| \leq r^k$ for some $r \in (0,1)$. Then $\{x_k\}_{k=1}^{\infty}$ converges R-linearly.

THEOREM 4.12. *If Assumption 4.1 and 4.2 hold and an RHSQP algorithm uses the step secant update strategy with the positive curvature criterion and either the l_1 or the Fletcher merit function with the penalty parameter with ν_k sufficiently large and $\bar{\nu}_k$ bounded, then the sequence $\{x_k\}_{k=1}^{\infty}$ produced by the algorithm converges to solution x^* R-linearly.* □

Now the global and R-linear convergence for the step secant is established by using the l_1 and Fletcher merit functions. In the next section, we present some results of the numerical experiments to compare the step secant update strategy using positive curvature criterion and the Nocedal and Overton criterion with the null space secant update strategy.

5. Numerical Experiments. Although the null space secant strategy and the step secant update strategies with the Nocedal and Overton update criterion and the positive curvature criterion are proved R-linearly convergent, they have different numerical performances. We present here numerical experiments with the step secant update strategy using these two update criteria and compare them with the null space secant update strategy, although it is known the null space secant update strategy is expensive because of the extra gradient evaluations. We used a single Fortran code that allowed us to vary update strategies. In these numerical experiments, the simpler l_1 merit function is used. We used the QR factorization (2.7) to compute the null space basis matrix Z_k as well as the inverse matrices as we described in (2.6). For the null space secant update, we used the BFGS update with y_k and s_k given by (2.9) and (2.10). We skipped the updates if $s_k^T y_k \leq 0$.

The algorithm parameters used are

- the general parameters, $\rho = 1$, $B_0 = I$, $\eta = 10^{-4}$ and $\tau = \tau_1 = 0.5$;
- parameters for the Nocedal and Overton criterion: $\zeta = 1.0$ and $\epsilon = 0.01$ (the same values used in [11]);
- parameters for the positive curvature criterion: $\zeta_1 = \zeta_2 = 0.01$.

The problems tested are chosen from Hock and Schittkowski's test problems [9]. For example, "hs10" stands for the problem 10 from [9]. The following notation is used in the tables showing the our numerical results.

- "upd" is the number of updates;
- "ite" is the number of iterations;
- "rsd" is the residual, $||Z_k^T g_k|| + ||c_k||$; and
- "**F**" indicates the algorithm's failure on that problem.

From the tables, we can see that there are two kinds of failure cases whose iteration numbers equal to or less than the maximum iteration allowance which was set to 100, respectively. The cases with a number of iterations less than the maximum iteration allowance is caused by the failure of line search. That is, at current approximation, the line search cannot find a positive step length greater than $\alpha_{\min}$ (which was chosen as $\alpha_{\min} = 10^{-30}$) such that (2.15) holds. At the ends of the two tables, there are rows to show the total numbers of updates/iterations and the update ratios. Note that the these numbers are obtained by counting the cases where all of the positive curvature criterion and the Nocedal and Overton criterion for the step secant update and the null space secant update have successively reached the solutions. The stopping criterion was rsd $\leq \varepsilon = 10^{-8}$ and the starting points were the standard points given in [9] and these problems were tested on a Sun Sparc-2 workstation.

First of all we present the results by using the monotonically increasing strategy given in **step** $2'$, for which the global and R-linear convergence is established. The results is presented in Table 1. It shows that the positive curvature criterion improves the Nocedal and Overton criterion as it not only has fewer failed cases but also uses fewer function evaluations and iterations. We believe that this improvement is due to the higher update rate of the positive criterion.

Even though there is no convergence analysis established for the nonmonotonically increasing strategy, numerical experiments in Table 2 show it works much better than the monotonically increasing strategy. For this numerical experiments in this paper, the following nonmonotonically increasing strategy is used,

$$\mu_k = ||\lambda_k|| + \rho$$

for some constant $\rho > 0$. For this nonmonotonically increasing strategy, the step secant update strategy with the positive curvature criterion works as well as the null space update strategy and in addition it saves the extra gradient evaluations.

6. Conclusions. The purpose of this paper was to present a more realistic analysis of the reduced Hessian SQP and to present a new practical update criterion. It presents the first analysis of the step secant update for reduced Hessian SQP in the context of a line search, and without assumptions on the accuracy of the initial Hessian approximation. We have done this for both the well-known Nocedal and Overton criterion and for the positive curvature criterion proposed here.

The positive curvature update criterion was proposed to allow more updates than the Nocedal and Overton update criterion, and based on numerical experiments, this seems to occur. It seems plausible that the superior performance using the positive curvature criterion is due to this greater update frequency.

From the numerical experiments made in this paper and the result of the global and R-linear convergence of the step secant update strategies, the positive curvature criterion may be a competitive candidate for solving very large scale constrained optimization problems, especially when it is combined with the nonmonotonically increasing strategy because it save the extra gradient evaluation required by the null space secant update strategy.

REFERENCES

[1] R. H. Byrd and J. Nocedal, *A tool for analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal., 26 (1989), pp. 727–739.

[2] ———, *An analysis of reduced Hessian methods for constrained optimization*, Math. Programming, 49 (1990), pp. 285–323.

[3] T. F. Coleman and A. R. Conn, *On the local convergence of a quasi-Newton method for the nonlinear programming problem*, SIAM J. Numer. Anal., 21 (1984), pp. 755–769.

[4] R. Fletcher, *A class of methods for nonlinear programming with termination and convergence propoerties*, Integer and Nonlinear Programming, North-Holland, Amesterdam, 1970.

[5] ———, *A first derivative method for nonlinear programming based on successive l_1 LP*, numerical analysis report NA/114, Department of Mathematics and Computer Science, University of Dundee, 1988.

[6] D. Gabay, *Reduced quasi-Newton methods with feasibility improvement for nonlinear constrained optimization*, Math. Programming Study, 16 (1982), pp. 18–44.

[7] J. C. Gilbert, *Maintaining the positive definiteness of the matrices in reduced Hessian methods for equality constrained optimization*, Mathematical Programming, 50 (1991), pp. 1–28.

[8] S. P. Han, *Superlinearly convergent variable metric algorithms for general nonlinear programming problems*, Math. Programming, 11 (1976), pp. 263–282.

[9] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 1981.

[10] W. Murray and M. H. Wright, *Projected Lagrangian methods based on the trajectories of penalty and barrier functions*, systems optimization laboratory report 78-23, Stanford University, 1978.

[11] J. Nocedal and M. L. Overton, *Projected Hessian updating algorithms for nonlinearly constrained optimization*, SIAM J. Numer. Anal., 22 (1985), pp. 821–850.

[12] J. D. Pearson, *Variable metric methods of minimization*, Computer J., 12 (1969), pp. 171–178.

[13] M. J. D. Powell, *Some global convergence properties of a variable metric algorithm for minimization without exact line searches*, Nonlinear Programming, SIAM-AMS Proceedings, Vol. IX, American Mathematical Society, Providence, 1976.

[14] ———, *Variable metric methods for constrained optimization*, Mathematical Programming: The State of the Art, Springer-Verlag, 1983, pp. 288–311.

TABLE 1
Numerical Tests with monotonically increasing parameter

	Positive Curvature		Nocedal-Overton		null space secant	
Pro #	upd/ite	rsd	upd/ite	rsd	upd/ite	rsd
hs6	8/11	0.8d-08	28/36	0.7d-08	8/10	0.2d-08
hs7	3/8	0.2d-11	3/10	0.5d-09	5/7	0.1d-11
hs10	12/18	0.2d-10	3/15	0.9d-08	13/14	0.3d-10
hs11	5/9	0.6d-10	5/9	0.6d-10	7/8	0.9d-11
hs12	10/11	0.3d-08	6/13	0.4d-08	7/8	0.2d-08
hs26	12/100	0.2d-02**F**	23/26	0.3d-02**F**	33/35	0.7d-08
hs27	15/100	0.2d-01**F**	25/32	0.1d-01**F**	47/49	0.1d-08
hs29	10/12	0.1d-08	13/17	0.1d-08	7/9	0.5d-08
hs39	13/15	0.2d-09	9/28	0.8d-09	21/22	0.1d-09
hs40	4/6	0.7d-08	3/6	0.1d-09	4/5	0.1d-09
hs43	12/13	0.1d-08	34/41	0.5d-01**F**	9/10	0.8d-08
hs46	17/100	0.8d-02**F**	21/22	0.2d-01**F**	99/100	0.2d-05**F**
hs47	20/100	0.1d+01**F**	22/32	0.1d-01**F**	31/32	0.1d-08
hs56	16/17	0.6d-12	14/18	0.1d-09	10/11	0.6d-09
hs60	11/12	0.1d-09	9/12	0.6d-10	10/11	0.4d-08
hs61	36/80	0.3d-10	42/45	0.1d+02**F**	7/9	0.2d-08
hs63	5/7	0.5d-09	4/8	0.4d-10	5/6	0.6d-08
hs65	6/100	0.7d+01**F**	19/36	0.8d+01**F**	10/13	0.5d-09
hs66	10/11	0.6d-13	34/36	0.1d-02**F**	11/12	0.1d-13
hs71	10/11	0.4d-08	6/9	0.1d-09	6/7	0.3d-08
hs72	13/23	0.2d-08	13/27	0.6d-08	19/20	0.1d-09
hs77	14/15	0.2d-08	9/14	0.5d-09	18/19	0.2d-08
hs78	5/7	0.5d-08	4/7	0.9d-08	6/7	0.1d-08
hs79	26/28	0.1d-08	11/14	0.9d-09	11/12	0.1d-08
hs80	6/9	0.1d-10	6/10	0.2d-09	4/5	0.1d-09
hs81	14/22	0.5d-05**F**	8/19	0.2d+02**F**	4/5	0.1d-09
hs93	39/41	0.2d-09	26/32	0.1d+03**F**	22/26	0.1d-08
hs100	27/28	0.1d-08	36/43	0.1d-08	99/100	0.2d-06**F**
Total[1]	171/219	6**F**	146/253	10**F**	161/181	2**F**
upd. ratio	0.781		0.577		0.890	

TABLE 2
Numerical Tests with non-monotonically increasing parameter

	New Criterion		N-O Criterion		null space secant	
Pro #	upd/ite	rsd	upd/ite	rsd	upd/ite	rsd
hs6	11/13	0.7d-09	12/20	0.7d-09	8/10	0.2d-08
hs7	4/8	0.4d-15	3/10	0.3d-09	5/7	0.1d-11
hs10	26/34	0.5d-11	6/30	0.3d-08	21/22	0.4d-09
hs11	10/17	0.3d-09	3/16	0.4d-08	13/14	0.1d-12
hs12	9/10	0.1d-08	6/14	0.1d-10	7/8	0.2d-08
hs26	28/29	0.7d-08	27/31	0.7d-08	33/35	0.7d-08
hs27	34/36	0.3d-08	30/56	0.4d-08	27/28	0.1d-10
hs29	13/15	0.6d-09	8/30	0.6d-09	13/15	0.9d-08
hs39	16/18	0.1d-08	6/21	0.2d-15	16/17	0.5d-08
hs40	4/6	0.7d-08	3/6	0.1d-09	4/5	0.1d-09
hs43	14/15	0.1d-09	9/18	0.1d-10	10/11	0.1d-08
hs46	33/34	0.2d-08	33/34	0.2d-08	40/41	0.6d-08
hs47	19/23	0.2d-08	17/26	0.5d-08	29/30	0.2d-08
hs56	15/16	0.1d-09	0/100	**F**	16/17	0.2d-11
hs60	11/12	0.5d-10	10/13	0.3d-08	11/12	0.3d-08
hs61	10/13	0.6d-12	16/26	**F**	7/9	0.1d-11
hs63	8/9	0.9d-10	5/11	0.1d-11	7/8	0.1d-09
hs65	20/22	0.4d-10	8/15	0.4d-09	9/11	0.2d-09
hs66	8/9	0.1d-08	7/9	0.2d-08	8/9	0.2d-12
hs71	10/11	0.4d-08	6/9	0.1d-09	6/7	0.3d-08
hs72	13/23	0.2d-08	13/27	0.6d-08	19/20	0.1d-09
hs77	13/14	0.8d-10	9/14	0.3d-09	18/19	0.2d-08
hs78	5/7	0.5d-08	4/7	0.9d-08	6/7	0.1d-08
hs79	21/22	0.1d-09	8/12	0.7d-09	11/12	0.1d-08
hs80	4/21	0.3d-10	2/20	0.6d-08	17/18	0.1d-11
hs81	25/39	0.4d-08	8/100	**F**	4/5	0.1d-09
hs93	72/75	0.8d-08	2/100	**F**	21/24	0.1d-09
hs100	28/29	0.4d-08	32/39	0.1d-09	35/36	0.5d-13
Total[2]	362/437	0**F**	267/488	4**F**	373/402	0**F**
upd. ratio	0.830		0.550		0.928	

[1,2] Note that the totals are for the problem solved by all three strategies.