SIAM J. NUMER. ANAL.
Vol. 30, No. 5, pp. 1291–1314, October 1993

SIZING AND LEAST-CHANGE SECANT METHODS*

J. E. DENNIS, JR.† AND H. WOLKOWICZ‡

Abstract. Oren and Luenberger introduced in 1974 a strategy for replacing Hessian approximations by their scalar multiples and then performing quasi-Newton updates, generally least-change secant updates such as the BFGS or DFP updates [Oren and Luenberger, *Management Sci.*, 20 (1974), pp. 845–862]. In this paper, the function

$$\omega(A) = \left(\frac{\operatorname{trace}(A)/n}{\det(A)^{1/n}}\right)$$

is shown to be a measure of change with a direct connection to the Oren–Luenberger strategy. This measure is interesting because it is related to the ℓ_2 condition number, but it takes all the eigenvalues of A into account rather than just the extremes. If the class of possible updates is restricted to the Broyden class, i.e., scalar premultiples are not allowed, then the optimal update depends on the dimension of the problem. It may, or may not, be in the convex class, but it becomes the BFGS update as the dimension increases. This seems to be yet another explanation for why the optimally conditioned updates are not significantly better than the BFGS update. The theory results in several new interesting updates including a self-scaling, hereditarily positive definite, update in the Broyden class which is not necessarily in the convex class. This update, in conjunction with the Oren–Luenberger scaling strategy at the first iteration only, was consistently the best in numerical tests.

Key words. conditioning, least-change secant methods, quasi-Newton methods, unconstrained optimization, sizing, scaling, Broyden class

AMS subject classifications. 49M37, 65K05, 65K10, 90C30, 49S40

1. Introduction.

We consider quasi-Newton methods for the unconstrained optimization problem

$$\min_{x\in\Re^n} f(x),$$

where f is twice continuously differentiable. The methods use a local quadratic model of the form

$$f(x_c+s) \approx f(x_c) + g_c^t s + \tfrac{1}{2} s^t B_c s,$$

where x_c is the current approximation to a minimizer x^*, B_c is the current approximation to the true Hessian G at x_c, and g_c is the gradient at x_c. We will use the notation that $B^{-1} = H$. Our particular interest is in the secant-type methods based upon approximating Newton's method by accumulating Hessian approximations using gradient differences. These methods have the property that the next Hessian approximation $B = B_+$ satisfies the secant condition

$$Bs = y \equiv g_+ - g_c \quad \text{or} \quad Hy = s \equiv x_+ - x_c.$$

* Received by the editors January 10, 1990; accepted for publication (in revised form) July 9, 1992. This report was issued simultaneously at the Department of Combinatorics and Optimization, University of Waterloo, Ontario, Canada and the Department of Mathematical Sciences, Rice University, Houston, Texas.

† Department of Mathematical Sciences, Rice University, Houston, Texas, and Department of Combinatorics and Optimization, University of Waterloo, Ontario, Canada. The research of this author was partially supported by Air Force grant AFOSR-89-0363 and National Science Foundation grant DMS-8903751.

‡ Department of Combinatorics and Optimization, University of Waterloo, Ontario, Canada. The research of this author was partially supported by the National Science and Engineering Research Council of Canada.

Many of these methods accumulate Hessian information based on some measure of least change to the matrix containing the old information. The best known class of such approximations is the Broyden family of updates. The best known members of this family are the BFGS and DFP methods. These two updates arise using two different scaled Frobenius norms as the measure of least change. The BFGS update is more popular and has proven to be more efficient; see, e.g., [1]. This emphasizes the fact that scaling is important in least-change secant methods. We will not restrict ourselves solely to the Broyden family.

Using the standard ℓ_2 measure of conditioning κ, optimal updates in the Broyden family of rank-two updates have been found by Davidon [2]. Specifically, Davidon chooses a member $B_+(\phi)$ of the Broyden class that minimizes $\kappa(H_c B_+(\phi))$. Although it is common to call these methods *optimally conditioned*, we think it is more salient, as well as more in the mainstream of the subject, to view $\kappa(H_c B_+(\phi))$ as a measure of the change made in the Hessian approximation by the update. This is because we like to view κ as a measure of deviation of a matrix from a multiple of the identity.

In this paper we will use a different measure

$$\omega(A) = \frac{\operatorname{trace}(A)}{n \det(A)^{1/n}} \tag{1.1}$$

of the deviation of a matrix from a multiple of the identity. To us, this measure seems more relevant to the updating context than κ. Furthermore, it is related to the measures and functions used in the proofs in [3], [4]. (For some interesting results on the specific measure

$$\psi(A) = \operatorname{trace}(A) - \log(\det(A)) \tag{1.2}$$

used in [4]; see [5]. In fact, it is shown there that the measure ψ gives rise to the BFGS and DFP updates.) We give some properties of ω and relate it to κ in §2. (For further relations between the measures ω and κ, see [6]. In fact, it is shown there that the ω optimal updates are κ optimal.) In particular we show that the particular optimal column scaling that arises from our measure ω is used in practice, because it is successful and easy to implement. However it is not the optimal scaling that arises from the measure κ. In §3, we find least-change secant updates from the Broyden class using $\omega(H_c B_+)$ and $\omega(B_c H_+)$. These results are interesting, but they mainly serve as lemmas for our main results, which are given in §5. The main results include very interesting connections between updates that minimize the measures $\omega(H_c B_+)$ and $\omega(B_c H_+)$ and the so-called Oren–Luenberger [7] *scaling*.

In order to interpret the results of §3, and to prepare for the results of §5, we give some results on Oren–Luenberger scaling in §4. Also, since Oren–Luenberger scaling, which we prefer to call *sizing*, is generally regarded as useful only in the first step of an iteration [8], we look for an alternative for subsequent iterations. The alternative we look for should accomplish the same task as sizing does, i.e., it should guarantee that the weak secant condition given below is satisfied so that the curvature information along the step is correct. However, it should avoid large changes in the spectrum. This leads us to some interesting weighted Frobenius norm problems for weak forms of the secant condition

$$s^T B s = s^T y \quad \text{or} \quad y^T H y = y^T s.$$

These problems are solved with some surprises in §4 by rank-one updates, which we call *weak secant updates*. Weak updating followed by weighted Frobenius updating leads to a pair of *Fletcher-dual* members of the Broyden class which we have not seen identified before, but which resemble the Hoshino update [9]. These updates are self-scaling, hereditarily positive definite, but not necessarily in the convex class. In fact, one of these updates, in conjunction with inverse sizing at the first iteration only, proved to be the best update among the various updates we tested. Moreover, it is shown that weak secant updating followed by the BFGS update is equivalent to the BFGS update. This provides an explanation for why the Oren–Luenberger scaling is successful at only the first iteration.

In §5, we bring together sizing, and weak and strong least-change secant updating in ω and in the traditional weighted Frobenius norms associated with the DFP and BFGS methods. This leads to even stronger connections between sizing, ω-least-change secant updates, and the DFP and BFGS methods. In §6, we consider the special two-dimensional example of Powell in [1] used there to illustrate that the BFGS behaves better than the DFP. We show that the ω-least-change secant method for this special case is a sized DFP, or equivalently an inverse-sized BFGS, and we give the corresponding numerical results for the sized DFP. These results are better than for the BFGS. We also include numerical tests for the standard set of test problems given in [10]. The numerical tests show that the optimal ϕ updates improve on the BFGS update and that replacing sizing, after the first sizing step, with the weak secant update improves convergence. This strategy leads to the self-scaling updates from §4, which proved to be the best updates in our numerical testing. In fact, several of the updates presented in this paper are shown to consistently improve on the BFGS update. Our tests also showed that sizing is a fix for the DFP update and that sizing without any line search outperformed many updates with line search. The sizing methods were particularly successful when the initial estimates were far from the optimal solution.

2. Preliminaries.

Let A be an $n \times n$ symmetric positive definite matrix (denoted s.p.d.), with eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_n > 0,$$

and corresponding eigenvectors $u_1, \ldots, u_n$. Then the usual ℓ_2 condition number of A is given by

$$\kappa(A) = \frac{\lambda_1}{\lambda_n}.$$

The condition number $\kappa(A)$ is used in perturbation analysis for matrix inversion; e.g., for the systems of linear equations $Ax = b$ and $A\bar{x} = \bar{b}$, we obtain bounds on the relative differences

$$\frac{1}{\kappa(A)} \frac{||b - \bar{b}||}{||b||} \leq \frac{||x - \bar{x}||}{||x||} \leq \kappa(A) \frac{||b - \bar{b}||}{||b||}. \tag{2.1}$$

See, e.g., [11]. This condition number acts as an upper bound on the amplification factor of the relative change in the right-hand side in bounding the relative change in the solution. Scaling A to reduce the condition number is a standard practice. See, e.g., [11].

It has often been noted that κ depends only on the largest and smallest eigenvalues. We propose using the following measure which depends more uniformly on all the eigenvalues:

$$\omega(A) \equiv \frac{\operatorname{trace}(A)/n}{\det(A)^{1/n}}.$$

(We allow the argument A to be s.p.d. or the product of two s.p.d. matrices.) Just as the usual condition number $\kappa(A)$ is a measure of how close A is to the pencil αI, where I is the identity matrix and $\alpha > 0$, the function $\omega(A)$ similarly measures the "distance" of A from αI. This measure takes all the eigenvalues of A into account and so is a more uniform or average indicator of the distance of A. Thus we can differentiate between two matrices with different spectrums but which have identical largest and smallest eigenvalues. Moreover, $\omega(A)$ can be calculated in terms of the actual data of A rather than its spectrum and can be more easily differentiated and manipulated. This is evident in the optimal scaling derived in Proposition 2.1 (v) where ω yields the optimal column scaling used in practice. The optimal scaling derived from κ is not easy to implement or derive.

We now give some useful properties of ω and address some related issues. Here we restrict ourselves to the s.p.d. matrices and use the Loewner order, i.e., $A \geq B$ means $A - B$ is positive semidefinite. See [12, p. 475]. Remember that a function being pseudoconvex means:

$$(y-x)^t \nabla f(x) \geq 0 \Longrightarrow f(y) \geq f(x).$$

See [13].

PROPOSITION 2.1. *The measure* $\omega(A)$ *satisfies*

(i) $1 \leq \omega(A) \leq \kappa(A) < ((\kappa(A)+1)^2/\kappa(A)) \leq 4\omega^n(A)$, *with equality in the first and second inequality if and only if A is a multiple of the identity, and equality in the last inequality if and only if*

$$\lambda_2 = \cdots = \lambda_{n-1} = \frac{\lambda_1 + \lambda_n}{2}. \tag{2.2}$$

(ii) $\omega(\alpha A) = \omega(A)$, *for all* $\alpha > 0$.

(iii) *If* $n = 2$, $\omega(A)$ *is isotonic with* $\kappa(A)$.

(iv) *The measure* ω *is pseudoconvex on the set of* s.p.d. *matrices, and thus any stationary point is a global minimizer of* ω.

(v) *Let* V *be a full rank* $m \times n$ *matrix,* $n \leq m$. *Then the optimal column scaling that minimizes the measure* ω, *i.e.,*

$$\min\ \omega((VD)^t(VD)),$$

over D *positive, diagonal, is given by*

$$D_{ii} = \frac{1}{\|V_{\cdot i}\|}, \qquad i = 1, \ldots, n,$$

where $V_{\cdot i}$ *is the ith column of* V.

Proof. That $1 \leq \omega(A)$ follows from the arithmetic-geometric mean inequality, while $\omega(A) \leq \kappa(A) < ((\kappa(A)+1)^2/\kappa(A))$ follows from the definitions. The equality conditions also follow directly from the definitions. To prove the last inequality in (i), we fix λ_1 and λ_n and thus also $\kappa(A)$. We now minimize $\omega(A)$ by differentiating.

(Note (iv) in the proposition.) This yields the equality conditions (2.2). Substitution shows that

$$\min\ \omega^n(A) = \frac{(\kappa(A)+1)^2}{4\kappa(A)}.$$

If $n = 2$,

$$2\omega(A) = \frac{\kappa(A)+1}{\kappa(A)^{1/2}} = \kappa(A)^{1/2} + \kappa(A)^{-1/2}. \tag{2.3}$$

The derivative of $\omega(A)$ with respect to $\kappa(A)$ can now be seen to be positive since $\kappa(A) \geq 1$.

The function $\det(A)$ is log concave and strictly increasing and the function $\det(A)^{1/n}$ is concave and increasing. See, e.g., [12]. The trace function is linear and so convex. Thus, ω is the quotient of a convex function by a concave function and so is pseudoconvex. Pseudoconvex functions have the property that every stationary point is a global minimizer. See [13].

To prove (v), let V be given. Then the arithmetic-geometric mean inequality yields

$$\begin{aligned}
\omega((VD)^t(VD)) &= \frac{trD^tV^tVD/n}{(\det D^tV^tVD)^{1/n}} \\
&= \frac{trV^tVD^2/n}{(\det D^tD)^{1/n}}(\det V^tV)^{1/n} \\
&= \frac{\Sigma\|V_{\cdot i}\|^2 D_{ii}^2/n}{(\Pi\|V_{\cdot i}\|^2 D_{ii}^2)^{1/n}}\left(\frac{\Pi\|V_{\cdot i}\|^2}{\det V^tV}\right)^{1/n} \\
&\geq \left(\frac{\Pi\|V_{\cdot i}\|^2}{\det V^tV}\right)^{1/n}
\end{aligned}$$

with equality if and only if $\|V_{\cdot i}\|^2 D_{ii}^2 =$ constant, $i = 1, \dots, n$. □

Property (i) above shows that $\omega(A)$ is a valid condition number; e.g., we can replace $\kappa(A)$ by $4\omega^n(A)$ in (2.1). As mentioned above, property (v) shows that the measure ω predicts the best column scaling which is used in practice (see, e.g., [11]), and that it is not the one found by minimizing the measure κ. Moreover, the proof of (v) is particularly simple. (It is interesting to note that the measure ω arose in an attempt to find an optimal scaling for A. In addition, ω is equivalent to the potential function used by Karmarkar [14] for linear programming in the case where the problem is scaled so that the objective cost vector is a vector of ones.)

We use ω as a measure of "best" in determining some quasi-Newton updates. This leads to minimizing this measure subject to constraints. The following lemma shows, under very mild assumptions, that we do not have to worry about maintaining positive definiteness in our updates since they will solve problems like the one posed here.

LEMMA 2.1. *Given the* s.p.d. *matrix C, consider the quantity*

$$\begin{aligned}
\mu^* &= \inf_B \omega(BC) \\
\text{subject to}\quad & v^tBv = \gamma \\
& B \text{ s.p.d.},\ \ B \in \Omega,
\end{aligned}$$

where $v \neq 0$ and $\gamma > 0$ are given, and Ω is a closed set of symmetric matrices. Assume that a feasible B exists. Then the finite value μ^* is attained at some B^* s.p.d.

Proof. First note that

$$1 \leq \mu^* \leq \alpha, \tag{2.4}$$

where $\alpha = \omega(B_0C) < \infty$, and B_0 is a feasible matrix. Choose $\{B_k\}$ such that each B_k is feasible and

$$\lim_{k\to\infty} \omega(B_kC) = \mu^*. \tag{2.5}$$

If either $\lambda_1(B_k)$ is unbounded above, or $\lambda_n(B_k)$ is not bounded away from zero, then $\sup \kappa(B_k) = \sup \omega(B_k) = \infty$, by Proposition 2.1 (i). However, this implies $\sup \omega(B_kC) = \infty$, a contradiction. Thus we can assume that $B_k \to \tilde{B}$ for some $\tilde{B} \in \Omega$ which is s.p.d. □

In the sequel we consider the space of symmetric matrices as a subspace of the $n \times n$ matrices with the inner product

$$\langle A, B\rangle = \text{trace}(AB).$$

The gradients of functionals restricted to this subspace are symmetric matrices.

3. ω-optimal rank-two updates. The results in this section are interesting, but the main reason we include them is for their use in §5. We now consider the Broyden family of updates

$$B_\phi = B_c - \frac{1}{s^tB_cs}B_css^t\,B_c + \frac{1}{y^ts}yy^t + (1-\phi)s^tB_cs\;ww^t, \tag{3.1}$$

where $s = x_+ - x_c$ is the step taken, $y = g_+ - g_c$, the current approximation B_c to the Hessian is s.p.d., $y^ts > 0$, and

$$w = \frac{1}{y^ts}y - \frac{1}{s^tB_cs}B_cs.$$

If $\phi = 1$ we get the BFGS update and $\phi = 0$ yields the DFP update. The updates for $\phi \in [0,1]$ are called the convex class. This is not the most common parameterization, but it is well known and it allows us to use results directly from [15] without reproving them here and uselessly lengthening the paper.

If we form the Fletcher dual updates, i.e., we exchange the roles of y and s and let $H_c = B_c^{-1}$, then we get the inverse updates

$$H_{\hat{\phi}} = H_c - \frac{1}{y^tH_cy}H_cyy^tH_c + \frac{1}{y^ts}ss^t + (1-\hat{\phi})y^tH_cyvv^t, \tag{3.2}$$

where

$$v = \frac{1}{y^ts}s - \frac{1}{y^tH_cy}H_cy. \tag{3.3}$$

We now have that $\hat{\phi} = 1$ and $\hat{\phi} = 0$ yield the DFP and BFGS updates, respectively.

Every member of the Broyden family of updates satisfies the secant condition, i.e., for every ϕ,

$$B_\phi s = y.$$

Furthermore, let

$$a = y^t H_c y, \quad b = y^t s, \quad c = s^t B_c s.$$

Note that $b^2 \leq ac$ with equality if and only if $B_c^{-1/2} y$ and $B_c^{1/2} s$ are collinear, which is true if and only if y and $B_c s$ are collinear, which is true if and only if $H_c y$ and s are collinear. From [15], B_ϕ is s.p.d. if $b > 0$ and $\phi < (ac/ac - b^2)$.

A κ-optimal rank-two update is found in [2] by minimizing the measure $\kappa(H_c B_\phi)$ over the Broyden family of rank-two updates. Note that the spectrum of a matrix product $C_1 C_2$ is equal to the spectrum of $C_2 C_1$ and $\kappa(C) = \kappa(C^{-1})$. Consequently, we can replace $H_c B_\phi$ by any of $B_\phi H_c$, $B_\phi^{-1} B_c$, $B_c B_\phi^{-1}$, $B_\phi^{-1/2} B_c B_\phi^{-1/2}$, etc. (See also [15, Chap. VII].) Related work is found in [16].

We now consider the problem of finding those updates in the Broyden family that minimize the measure ω. The updates depend on n but begin to look like the BFGS and DFP updates for large n. In the following, we assume $n \geq 2$, B_c is an $n \times n$ s.p.d. matrix, $s, y \in \Re^n$ such that $s^t y > 0$, and $y, B_c s$ are linearly independent. If y and $B_c s$ are linearly dependent, then the entire Broyden family of updates (as well as the sized updates) reduce to the symmetric rank-one (SR1) update, or simply to just B_c. Otherwise, $\phi_{SR1} = -(c/b - c)$ and $\hat{\phi}_{SR1} = -(a/b - a)$. Moreover, the inverse of the DFP update B_0 is H_1; the inverse of the BFGS update B_1 is H_0. In general,

$$\hat{\phi} = \iota(\phi) = \frac{1 - \phi}{1 + \phi[(b^2/ac) - 1]} \tag{3.4}$$

is a 1-1 and onto mapping (cf. [17]) that relates ϕ and $\hat{\phi}$ for which B_ϕ is s.p.d. and $B_\phi^{-1} = H_{\hat{\phi}}$. Note that the formula is undefined exactly when B_ϕ is positive semidefinite and singular. (This formula corrects a typographical error in [15] and [17].) The mapping satisfies $\iota(\iota(\phi)) = \phi$ and the convex class $\phi \in [0, 1] \mapsto \hat{\phi} \in [0, 1]$.

LEMMA 3.1 [15, p. 111]. *The matrix $B_c^{-1/2} B_\phi B_c^{-1/2}$ has $n - 2$ unit eigenvalues and the two remaining eigenvalues are*

$$\lambda_\pm(\phi) = f_1(\phi) \pm (f_1(\phi)^2 - f_2(\phi))^{1/2}, \tag{3.5}$$

where

$$f_1(\phi) = \frac{a(b + c) - \phi(ac - b^2)}{2b^2}, \qquad f_2(\phi) = \frac{a}{b} - \frac{\phi(ac - b^2)}{bc}. \tag{3.6}$$

We now present the ω-optimal update from the Broyden family. Notice that the update depends on n and, for large n, the ω-optimal update looks more and more like the BFGS update but is in the convex class only when $a \geq b$.

THEOREM 3.1. *The minimum over ϕ of $\omega(B_c^{-1/2} B_\phi B_c^{-1/2})$ is attained at*

$$\phi_* = 1 + \frac{(a - b)b}{(1 - n)(ac - b^2)}. \tag{3.7}$$

Furthermore, B_{ϕ_*} is s.p.d. *and*

$$\iota(\phi_*) = \left(\frac{ac}{ac - b^2}\right) \frac{a - b}{(n - 1)b + (a - b)}. \tag{3.8}$$

Proof. From Lemma 3.1,

$$\omega(B_c^{-1/2} B_\phi B_c^{-1/2}) = \frac{(2f_1(\phi) + n - 2)/n}{(f_2(\phi))^{1/n}}.$$

We can now substitute using (3.6), differentiate, and solve for ϕ_*, the critical point. This yields (3.7). Since ω is pseudoconvex and both f_1 and f_2 are linear in ϕ, ϕ_* is the global minimizer. The ω-optimal update is s.p.d. because it is easy to show that for $n \geq 2$, $\phi_* < (ac/ac - b^2)$. (This also follows from Lemma 2.1 with the appropriate choice of the closed set Ω.) □

COROLLARY 3.1. *The minimum over $\hat{\phi}$ of $\omega(B_c^{1/2} H_{\hat{\phi}} B_c^{1/2})$ is attained at*

$$\hat{\phi}_* = 1 + \frac{(c-b)b}{(1-n)(ac-b^2)}.$$

Furthermore, $H_{\hat{\phi}_*}$ is s.p.d. *and*

$$\iota(\phi_*) = \left(\frac{ac}{ac - b^2}\right) \frac{c - b}{(n-1)b + (c - b)}. \tag{3.9}$$

Proof. The update formulas (3.1) and (3.2) are obtained by exchanging the roles of y and s, i.e., the secant condition can be expressed as $B_\phi s = y$ or $H_{\hat{\phi}} y = s$, where ϕ and $\hat{\phi}$ are related by (3.4). This is equivalent to exchanging the roles of a and c in Theorem 3.1. □

Note that, although $\kappa(A) = \kappa(A^{-1})$, whenever A^{-1} is defined, this is not the case for the measure ω. Therefore, the optimal update obtained in Theorem 3.1 is not, in general, the same as the update obtained in Corollary 3.1, nor are the values of the measure ω equal. The following table summarizes the results.

ω-optimal rank-two updates.

Measure	ϕ	$\hat{\phi}$
$\omega(H_c^{1/2} B_\phi H_c^{1/2})$ (for optimal ϕ)	$\phi_* = 1 + \frac{(a-b)b}{(1-n)(ac-b^2)}$	$\iota(\phi_*) = \hat{\phi} = \frac{\frac{-(a-b)b}{(1-n)(ac-b^2)}}{1+\left(1+\frac{(a-b)b}{(1-n)(ac-b^2)}\right)(\frac{b^2}{ac}-1)}$
$\omega(B_c^{1/2} H_{\hat{\phi}} B_c^{1/2})$ (for optimal $\hat{\phi}$)	$\iota(\hat{\phi}_*) = \phi = \frac{\frac{-(c-b)b}{(1-n)(ac-b^2)}}{1+\left(1+\frac{(c-b)b}{(1-n)(ac-b^2)}\right)(\frac{b^2}{ac}-1)}$	$\hat{\phi}_* = 1 + \frac{(c-b)b}{(1-n)(ac-b^2)}$

In general, to find an optimal ϕ_*, we minimize the measure $\omega(H_c^{1/2} B_\phi H_c^{1/2})$ over B_ϕ, and then use Stoer's formula to find the corresponding $\hat{\phi} = \iota(\phi_*)$. Conversely, an optimal $\hat{\phi}$ refers to the measure $\omega(B_c^{1/2} H_{\hat{\phi}} B_c^{1/2})$. Notice that for large n, $\hat{\phi}_*$ is near 1, which corresponds to the DFP, and $\hat{\phi}_*$ is in the convex class only when $c \geq b$.

4. Sizing and weak secant updating. This section provides some preliminary results on the effect of sizing. It serves as a preliminary introduction to §5. We also introduce *shifting* (weak secant updating) to replace sizing, in an attempt to better preserve accumulated Hessian information. In fact, the Broyden class update in Theorem 4.4 (ii) proved to be the best update in our numerical tests. The Broyden family, B_ϕ, of rank-two updates satisfies the secant condition and preserves positive definiteness whenever $y^t s > 0$ for $\phi < (ac/ac - b^2)$. If B_+ is any symmetric matrix that satisfies the secant condition $B_+s = y$, then we are guaranteed that the curvature information in B_+ along the step s is approximately correct, i.e.,

$$\nabla f(x+s) = g_+ = g + Gs + \bigcirc(\|s\|^2) \tag{4.1}$$

and

$$s^tGs \approx y^ts = s^tB_+s, \tag{4.2}$$

$$y^tG^{-1}y \approx y^ts = y^tH_+y. \tag{4.3}$$

We refer to the above two equations as the direct and inverse *weak secant conditions.* (Our numerical results at the end of this paper show that the weak secant conditions improve on sizing. Since the writing of this paper, related numerical tests have been done in [18] which also show that these conditions, in conjunction with improved spectral information, can be used to obtain very successful nonsecant method updates.) By choosing updates that minimize a measure such as $\omega(B_+), \kappa(B_+)$, or $||B_+ - B_c||_F$, we attempt to guarantee that the new directional information does not destroy too much information already built up in B_c. Note that the measure ω does this uniformly over all the eigenvalues while κ only deals with the two extreme eigenvalues. However, the low-rank updates correct the curvature information of B_+ in one direction at a time. Therefore when B_c has large eigenvalues, the new update B_+ can become ill conditioned (cf. [19, p. 275]). This can be corrected by sizing B_c. More precisely, B_c is to be replaced by $(y^ts/s^tB_cs)B_c$ before updating (cf. [20]). Conversely, if H_c is replaced by $(y^ts/y^tH_cy)H_c$ before updating, then we are sizing H_c. We now find the ω-optimal ϕ and $\hat{\phi}$ to determine a member of the Broyden family that is obtained after sizing. For each sizing, we obtain two "optimal" matrices and their inverses. Again we see that the BFGS and DFP updates play a role.

THEOREM 4.1. *If $B_c \leftarrow (b/c)B_c$, then the optimal ϕ_* and corresponding $\hat{\phi} = \iota(\phi_*)$ are given by*

$$\phi_* = 1 + \frac{1}{1-n}, \tag{4.4}$$

and

$$\hat{\phi} = \iota(\phi_*) = \frac{1}{(n-1) + (n-2)((b^2/ac) - 1)}, \tag{4.5}$$

respectively. Similarly, if $H_c \leftarrow (c/b)H_c$, then the optimal $\hat{\phi}$ and corresponding $\phi = \iota(\hat{\phi}_*)$ are

$$\hat{\phi}_* = 1, \tag{4.6}$$

and

$$\iota(\hat{\phi}_*) = \phi = 0. \tag{4.7}$$

All values are in the convex class. The optimal ϕ_* gives the DFP *update if $n = 2$ and approaches the* BFGS *update as n grows. For every n, the optimal $\hat{\phi}_*$ gives the* DFP *update. If $B_c \leftarrow (a/b)B_c$, then the optimal $\phi_* = 1$ and the* BFGS *update is optimal.*

Similarly, if $H_c \leftarrow (b/a)H_c$*, then the optimal* $\hat{\phi}_*$ *and corresponding* $\phi = \iota(\hat{\phi}_*)$ *are given by*

$$\hat{\phi}_* = 1 + \frac{1}{1-n}, \tag{4.8}$$

and

$$\iota(\hat{\phi}_*) = \phi = \frac{1}{(n-1) + (n-2)((b^2/ac) - 1)}, \tag{4.9}$$

respectively. This is always in the convex class. The optimal ϕ_* *is the* BFGS *update. The optimal* $\hat{\phi}_*$ *is also the* BFGS *update, if* $n = 2$*, but it approaches the* DFP *update as* n *grows.*

Before giving the proof, we summarize the results in some tables. We prematurely include the result of Theorem 4.4 that the ϕ values are the same after weak updating. (Note that the BFGS update is the ω-optimal rank-two update both after inverse sizing or after weak inverse sizing, while the DFP update is the ω-optimal update after sizing or weak sizing.)

ω-optimal rank-two updates after sizing $B_c \leftarrow \frac{b}{c}B_c$ *(or direct shift).*

Measure	ϕ	$\hat{\phi}$
$\omega(H_c^{1/2}B_\phi H_c^{1/2})$ (for optimal ϕ)	$\phi_* = 1 + \frac{1}{1-n}$	$\iota(\phi_*) = \hat{\phi} = \frac{1}{(n-1)+(n-2)((b^2/ac)-1)}$
$\omega(B_c^{1/2}H_{\hat{\phi}}B_c^{1/2})$ (for optimal $\hat{\phi}$)	$\iota(\hat{\phi}_*) = \phi = 0$	$\hat{\phi}_* = 1$

ω-optimal rank-two updates after inverse sizing $H_c \leftarrow \frac{b}{a}H_c$ *(or weak inverse update).*

Measure	ϕ	$\hat{\phi}$
$\omega(H_c^{1/2}B_\phi H_c^{1/2})$ (for optimal ϕ)	$\phi_* = 1$	$\iota(\phi_*) = \hat{\phi} = 0$
$\omega(B_c^{1/2}H_{\hat{\phi}}B_c^{1/2})$ (for optimal $\hat{\phi}$)	$\iota(\hat{\phi}_*) = \phi = \frac{1}{(n-1)+(n-2)((b^2/ac)-1)}$	$\hat{\phi}_* = 1 + \frac{1}{1-n}$

Proof. Since $\tilde{B} \leftarrow (b/c)B$, we see that

$$\tilde{b} \leftarrow b, \quad \tilde{c} \leftarrow \frac{b}{c}c = b, \quad \tilde{a} \leftarrow \frac{ca}{b}.$$

Thus Theorem 3.1 yields

$$\phi_* = 1 + \frac{((ca/b) - b)b}{(1-n)(ac - b^2)} = 1 + \frac{1}{1-n}.$$

Applying (3.8) then yields (4.5). If $\tilde{B}^{-1} \leftarrow (b/a)\tilde{B}^{-1}$ or $\tilde{B} \leftarrow (a/b)B$, we have $\tilde{b} \leftarrow b, \tilde{a} \leftarrow b,\ \tilde{c} \leftarrow (a/b)c$, and (3.7) yields

$$\tilde{\phi} = 1 + \frac{(b-b)b}{(1-n)(ac - b^2)} = 1.$$

Using Corollary (3.1) instead of Theorem 3.1 completes the proof. □

Sizing B_c (or H_c) can cause a drastic change in all the eigenvalues of B_c so that, although G and B_c may have had a relatively good overlapping of their spectra, they no longer do. After the first sizing step, which should be done to avoid possible ill conditioning, a better strategy might be to *shift* the spectrum by a rank-one update and thus avoid a drastic change in the whole spectrum. If we want to size both B_c and H_c simultaneously, we can use a rank-two update. Consequently, we will consider finding the "closest" matrix to B_c that satisfies the weak secant conditions.

The next few results hold some surprises for experts in the field.

THEOREM 4.2. *Let u, v be nonzero vectors in $\mathbb{R}^n$, and let A, M be symmetric matrices with M* s.p.d. *Then,*

$$\tilde{A} = A + \frac{(u^T v - v^T A v) M v v^T M}{(v^T M v)^2}$$

uniquely solves

$$\min \qquad \|W^T(\tilde{A} - A)W\|_F$$

$$\textit{subject to} \qquad v^T \tilde{A} v = u^T v$$

independent of W such that $M^{-1} = (WW^T)$. Moreover, if A is s.p.d. *and $M = A$, then $\tilde{A}$ is* s.p.d. *if and only if $u^t v > 0$.*

Proof. First note that $\tilde{A}$ is feasible, and let C be any other feasible matrix. Set $x = W^{-1}v$. Then $W^T(\tilde{A} - A)W = (v^T(C - A)v/x^T x) \cdot (xx^T/x^T x)$. Thus, $\|W^T(\tilde{A} - A)W\|_F = \left|(x^T(W^T(C - A)W)x/x^T x)\right| \leq \|W^T(C - A)W\|_2$. Now if $A = M$ is s.p.d. then $\tilde{A}$ is s.p.d. if and only if the rank-one update $W^T \tilde{A} W = I + (u^T v - v^T A v/(v^T M v)^2)W^{-1}vv^T W^{-T}$ is s.p.d. The latter is s.p.d. if and only if

$$0 < 1 + \operatorname{trace}\frac{(u^T v - v^T A v)}{(v^T A v)^2} W^{-1} v v^T W^{-T} ,$$

which is equivalent to $(v^T A v - u^T v/(v^T A v)^2)v^T A v < 1$, which is true if and only if $u^T v > 0$. □

Now we apply this theorem to direct shifting and then to weak inverse updating, i.e., to shifting B and then to shifting H. We use the terms Greenstadt and DFP to refer to the choice of W. In fact, Greenstadt never considered least changes to B_c, only to H_c.

COROLLARY 4.1. *The direct weak Greenstadt update*

$$\tilde{B} = B_c + \frac{1}{(s^T B_c s)^2}(y^T s - s^T B_c s) B_c s s^T B_c \tag{4.10}$$

uniquely solves

$$\min \qquad \|W^T(\tilde{B} - B_c)W\|_F$$

$$\textit{subject to} \qquad s^T \tilde{B} s = y^T s$$

for any square W such that $H_c = WW^T$. Moreover, $\tilde{B}$ is s.p.d. *if and only if $y^T s > 0$. Also, the direct weak* DFP *update*

$$\tilde{B} = B_c + \frac{1}{(y^T s)^2}(y^T s - s^T B_c s) y y^T \tag{4.11}$$

uniquely solves

$$\begin{array}{ll} \min & \|W^T(\tilde{B} - B_c)W\|_F \\ \textit{subject to} & s^t\tilde{B}s = y^Ts \end{array}$$

for any square W *such that* $(WW^T)^{-1}s = y$.

It is surprising that (4.10) is a hereditarily positive definite update, but (4.11) is not. This is directly opposite the case for strong secant methods with the same weighted Frobenius norms, since (4.11) corresponds to the DFP secant update. Now we will see in the next corollary that the same twist holds for weak inverse updating; the Greenstadt inverse update is hereditarily positive definite, but the BFGS is not, as we found with the first two randomly generated examples in MATLAB that we tried.

COROLLARY 4.2. *The inverse weak Greenstadt update*

$$\tilde{H} = H_c + \frac{1}{(y^TH_cy)^2}(y^Ts - y^TH_cy)H_cyy^TH_c$$

uniquely solves

$$\begin{array}{ll} \min & \|W^T(\tilde{H} - H_c)W\|_F \\ \textit{subject to} & y^T\tilde{H}y = y^Ts \end{array}$$

for any square W *such that* $B_c = WW^T$. *Moreover,* $\tilde{H}$ *is* s.p.d. *if and only if* $y^Ts > 0$. *Also the weak* BFGS *update*

$$\tilde{H} = H_c + \frac{1}{(y^Ts)^2}(y^Ts - y^TH_cy)ss^T$$

uniquely solves

$$\begin{array}{ll} \min & \|W^T(\tilde{H} - H_c)W\|_F \\ \textit{subject to} & y^T\tilde{H}y = y^Ts \end{array}$$

for any square W *such that* $(WW^T)^{-1}y = s$.

If we try to find the "best" update with respect to our measure ω that satisfies the spectral conditions (4.2) and (4.3), then we just scale B_c or H_c since the value of $\omega(\tilde{B}H_c)$ or $\omega(\tilde{B}^{-1}B_c)$ will be one. It is interesting that if we let $\sigma > 0$ and apply the following theorem to σB_c, we get $(\sigma B_c)_+ = \sigma B_+$.

THEOREM 4.3. *The update* $\tilde{B}$ *that solves*

$$\begin{array}{ll} \min & \omega(H_c^{-1/2}\tilde{B}H_c^{-1/2}) \ \ (\textit{or } \omega(B_c^{1/2}\tilde{B}^{-1}B_c^{1/2})) \\ \textit{subject to} & s^t\tilde{B}s = y^ts \end{array}$$

is

$$\tilde{B} \leftarrow \frac{y^ts}{s^tB_cs}B_c. \tag{4.12}$$

Proof. Since $H_c^{-1/2}\tilde{B}H_c^{-1/2} = (y^ts/s^tBs)I$, we have obtained the minimum of κ and ω, i.e., $\omega = 1 = \kappa$. $\square$

COROLLARY 4.3. *The update $\tilde{B}$ that solves*

$$\begin{array}{ll} \min & \omega(B_c^{1/2}\tilde{B}^{-1}B_c^{1/2})\,(or\ \omega(B_c^{1/2}\tilde{B}^{-1}B_c^{1/2})\ ) \\ subject\ to & y^t\tilde{B}^{-1}y = y^ts \end{array}$$

is obtained from

$$\tilde{B} \leftarrow \frac{y^tH_cy}{y^ts}B_c. \tag{4.13}$$

Optimal updates for weak secant equations.

Measure	Optimal	Optimal inverse
$\lVert H_c^{1/2}(\tilde{B}-B)H_c^{1/2}\rVert_F$ (constraint $s^tBs = y^ts$)	$\tilde{B} = B_c + \frac{1}{c^2}(b-c)B_css^tB_c$	$\tilde{B}^{-1} = H_c + \frac{c-b}{bc}ss^t$
$\lVert B_c^{1/2}(\tilde{B}^{-1}-H_c)B_c^{1/2}\rVert_F$ (constraint $y^tHy = y^ts$)	$\tilde{B} = B_c + \frac{a-b}{ba}yy^t$	$\tilde{B}^{-1} = H_c + \frac{1}{a^2}(b-a)H_cyy^tH_c$
$\omega(H_c^{1/2}\tilde{B}H_c^{1/2})$ (*or* $\omega(B_c^{1/2}\tilde{B}^{-1}B_c^{1/2})$) (constraint $s^tBs = y^ts$)	$\tilde{B} = \frac{b}{c}B_c$	$\tilde{B}^{-1} = \frac{c}{b}B_c^{-1}$
$\omega(B_c^{1/2}\tilde{B}^{-1}B_c^{1/2}$ (or $\omega(H_c^{1/2}\tilde{B}H_c^{1/2})$) (constraint $y^tHy = y^ts$)	$\tilde{B} = \frac{a}{b}B_c$	$\tilde{B}^{-1} = \frac{b}{a}B_c^{-1}$

In Theorem 4.1 we presented the optimal rank-two updates in the Broyden family obtained after sizing and using the measure ω. We now show that we obtain the same optimal ϕ_* (and $\hat{\phi}_*$) to strongly update the weakly updated matrix. This does not mean that the corresponding B_+ matrices will be the same, it just means they are obtained from the sized or weakly updated B_c (or H_c) using the same formula from the Broyden class.

THEOREM 4.4. *The optimal ϕ_* and $\hat{\phi}_*$ expressions in Theorem 4.1 are unchanged if we replace sizing B_c ($B_c \leftarrow (b/c)B$ and $H_c \leftarrow (c/b)H_c$) with the direct weak Greenstadt update* (4.1) *of Corollary* 4.1 *and we replace inverse sizing $H_c(H_c \leftarrow (b/a)H_c$ and $B_c \leftarrow (a/b)B_c$) with the inverse weak Greenstadt update* (4.2) *of Corollary* 4.2.

Proof. Suppose that we apply the direct weak update. Then the Sherman–Morrison formula yields

$$\tilde{H} = H_c + \frac{c-b}{bc}ss^t,$$

which implies

$$\tilde{a} \leftarrow a + \frac{b}{c}(c-b).$$

We also have $\tilde{b} \leftarrow b$ and $\tilde{c} \leftarrow b$. Therefore, Theorem 3.1 yields

$$\begin{aligned} \phi_* &= 1 + \frac{(\tilde{a}-b)b}{(1-n)(\tilde{a}\tilde{c}-b^2)} \\ &= 1 + \frac{(a+b-\frac{b^2}{c}-b)b}{(1-n)((a+\frac{b(c-b)}{c})b-b^2)} \\ &= 1 + \frac{1}{1-n}. \end{aligned}$$

Similarly, the optimal $\hat{\phi}_* = 1$, since $\tilde{b} = \tilde{c}$. By exchanging the roles of H_c and B_c, we see that the weak inverse update yields

$$\hat{\phi}_* = 1 + \frac{1}{1-n}, \qquad \phi_* = 1\,. \qquad \square$$

We now apply the weighted Frobenius norm measures after weak updating. Although we do not restrict the updates to the Broyden class, we get a new Fletcher-dual pair of updates in the Broyden class. These symmetric updates are hereditarily positive definite but not always in the convex class. The new updates are (iii) and (iv) in the following theorem. Note that (iii) reduces to the DFP update after sizing, while (iv) reduces to BFGS after inverse sizing.

THEOREM 4.5. *The following are equivalent updating sequences. The first four are hereditarily positive definite.*

(i) *The result of a direct weak Greenstadt update or a weak* BFGS *update followed in either case by a* BFGS *update is a* BFGS *update.*

(ii) *The result of an inverse weak Greenstadt update or a weak* DFP *update followed in either case by a* DFP *update is a* DFP *update.*

(iii) *The result of a direct weak Greenstadt update followed by a* DFP *update is the* $\phi = 1 - (b/c), \hat{\phi} = 1/((b/a) - ((b^2/ac) - 1))$ *update from the Broyden class.*

(iv) *The result of an inverse weak Greenstadt update followed by a* BFGS *update is the* $\hat{\phi} = 1 - (b/a),\ \phi = 1/((b/c) - ((b^2/ac) - 1))$ *update from the Broyden class.*

The following sequences may not be hereditarily positive definite.

(v) *The result of a weak* DFP *update followed by a* BFGS *update is the* $\phi = (c/b)$, $\hat{\phi} = (ab - ac/ab - ac + b^2)$ *member of the Broyden class.*

(vi) *The result of a weak* BFGS *update followed by a* DFP *update is the* $\hat{\phi} = (a/b)$, $\phi = (cb - ac/cb - ac + b^2)$ *member of the Broyden class.*

Proof. The proofs are much the same, so we will do only (i) and (iv) since they seem to be the most interesting updates.

The direct weak Greenstadt update is

$$\bar{B} = B_c + \frac{b-c}{c^2} B_c s s^T B_c \quad \text{and} \quad \bar{B}s = \left(1 + \frac{b-c}{c}\right) B_c s \quad \text{with } b = s^t \bar{B} s.$$

The weak BFGS is

$$\tilde{H} = H_c + \frac{b-c}{b^2} s s^T \quad \text{and} \quad s - \tilde{H}y = s - H_c y - \left(\frac{b-a}{b}\right) s = \left(1 - \frac{b-a}{b}\right) s - H_c y.$$

In the first case,

$$\begin{aligned} B_+ &= \bar{B} - \frac{\bar{B} s s^T \bar{B}}{b} + \frac{y y^T}{b} \\ &= B_c + \frac{b-c}{c^2} B_c s s^T B_c - \left(1 + \frac{b-c}{c}\right)^2 \frac{B_c s s^T B_c}{b} + \frac{y y^T}{b} \\ &= B_c - \frac{B_c s s^T B_c}{c} + \frac{y y^T}{b}. \end{aligned}$$

In the second,

$$\begin{aligned} H_+ &= \tilde{H} + \frac{(s-\tilde{H}y)s^T + s(s-\tilde{H}y)^T}{b} - \frac{y^T(s-\tilde{H}y)ss^T}{b^2} \\ &= H_c + \frac{b-a}{b}\frac{ss^T}{b} + \frac{((1-\frac{b-a}{b})s - H_c t)s^T + s((1-\frac{b-a}{b})s - H_c y)^T}{b} - 0 \\ &= H_c + \frac{(s-H_c y)s^T + s(s-H_c y)^T}{b} + \left(\frac{b-a}{b^2} - 2\frac{b-a}{b^2}\right) ss^T, \end{aligned}$$

which is the BFGS update of H_c. The proof of (iv) is as direct.

The inverse weak Greenstadt update of H_c is

$$\tilde{H} = H_c + \frac{(b-a)}{a^2} H_c y y^T H_c$$

and $\tilde{H}y = (1 + (b - a/a))H_c y$, $s - \tilde{H}y = s - H_c y - (b - a/a)H_c y$. So, following with a BFGS update,

$$\begin{aligned} H_+ &= \tilde{H} + \frac{(s-\tilde{H}y)s^T + s(s-\tilde{H}y)^T}{b} - \frac{y^T(s-\tilde{H}y)ss^T}{b^2} \\ &= H_c + \frac{b-a}{a^2} H_c y y^T H_c + \frac{(s-H_c y)s^T + s(s-H_c y)^T}{b} \\ &\quad - (\tfrac{b-a}{ab})(H_c y s^T + s y^T H_c) \pm \tfrac{b-a}{b^2} ss^T \\ &= H_c + \frac{(s-H_c y)s^T + s(s-H_c y)^T}{b} - \frac{b-a}{b^2} ss^T \\ &\quad + (b-a)(\tfrac{H_c y}{a}\tfrac{y^T H_c}{a} - \tfrac{H_c y}{a}\tfrac{s^T}{b} - \tfrac{s}{b}\tfrac{y^T H_c}{a} + \tfrac{s}{b}\tfrac{s^T}{b}). \end{aligned}$$

The first three terms are the BFGS and the last is

$$(b-a)\left(\frac{s}{b} - \frac{H_c y}{a}\right)\left(\frac{s}{b} - \frac{H_c y}{a}\right)^T = \left(\frac{b}{a} - 1\right) a\, vv^T,$$

where v is given by (3.3). Now use (3.2) with $\hat{\phi} = 0$ for the BFGS portion and we get

$$H_+ = H_c - \frac{1}{a} H_c y y^T H_c + \frac{1}{b} ss^T + avv^T + \left(\frac{b}{a} - 1\right) avv^T,$$

which is (3.2) with $1 - \hat{\phi} = 1 + (b/a) - 1 = (b/a)$, and (iv) is proven. Note that hereditary positive definiteness follows from the corresponding property for the BFGS and DFP updates and from the above two corollaries. □

Other combinations of weak updating and updating yield Broyden class updates, e.g., direct weak Greenstadt update followed by optimal ϕ yields the Broyden class update with $\phi = 1 - (b/(n-1)c)$ (compare with (iii) above).

5. ω-optimal updates. In §3 we found the "best" rank-two updates in the Broyden class, i.e., the rank-two updates parametrized by ϕ and $\hat{\phi}$ in (3.1) and (3.2) that minimize the measure ω. However, if we do not restrict ourselves to rank-two updates but only to maintaining positive definiteness and the secant equation, then we obtain a different result. We see that the best ω update of B_c in the case $ac \neq b^2$ is obtained by inverse sizing B_c and then applying the BFGS update. Similarly, the best possible inverse update is obtained by sizing and operating on H_c with DFP. The two updates are different although the optimal values of ω are ultimately equal. The proofs use the results from §3. Note that the measure ω is scale invariant. These results are parallel to the results in [5] where it is shown that the measure (1.2) in [4] (not scale invariant) gives rise to the BFGS and DFP updates.

We continue to assume that $y^t s > 0$ and that y and Bs are linearly independent.

THEOREM 5.1. *Assume B_c is* s.p.d. *and $ac \neq b^2$. Then for*

$$\alpha = \frac{b}{a},$$

the BFGS *update of* $(1/\alpha)B_c$,

$$H_+ = \alpha H_c + us^T + su^T, \qquad u = \frac{s - \alpha H_c y}{b},$$

is the unique solution of

$$\begin{array}{ll} \min & \omega(H_c B_+) \\ \textit{subject to} & B_+ s = y, \quad B_+ \text{ s.p.d.} \end{array}$$

In addition, the Lagrange multiplier for the secant equation is uniquely

$$\frac{2(s - \alpha H_c y)}{\alpha b n (\det(H_c B_+))^{1/n}},$$

while

$$\alpha = \frac{n}{\operatorname{trace}(H_c B_+)}, \qquad \omega(H_c B_+) = \left(\frac{ac}{b^2}\right)^{1/n}.$$

Proof. First, we note that an optimal B_+ s.p.d. exists from Lemma 2.1. The Lagrangian for our problem is

$$L(\lambda, B_+) = g(B_+) + \lambda^t (B_+ s - y),$$

where

$$g(B_+) = \frac{\operatorname{trace}(H_c B_+)}{n(\det(H_c))^{1/n}(\det(B_+))^{1/n}}$$

and $\lambda \in \Re^n$ is the Lagrange multiplier. The stationary point is optimal by Proposition 2.1 (v). The Lagrange multiplier exists since the constraints are linear. We now apply the necessary conditions of optimality. Uniqueness follows since we get a unique solution to the necessary conditions. For simplicity of notation, we multiply the Lagrangian by the constant $n(\det(H_c))^{1/n}$ and remove this constant from λ at the end. Now

$$\begin{aligned}
\lambda^t B_+ s &= \text{trace}(\lambda^t B_+ s) \\
&= \text{trace}(B_+ s \lambda^t) \\
&= \text{trace}(\lambda s^t B_+) \\
&= \text{trace}(s \lambda^t B_+) \\
&= \text{trace}\left(\frac{s\lambda^t + \lambda s^t}{2} B_+\right)
\end{aligned}$$

by adding the previous two equivalences and dividing by 2. Therefore the gradient of the linear functional $\lambda^t B_+ s = \langle (s\lambda^t + \lambda s^t/2), B_+ \rangle$ on the subspace of symmetric matrices is

$$\frac{s\lambda^t + \lambda s^t}{2}.$$

Using the cofactor expansion of the determinant along a row of the matrix, i.e., $\det(A) = \sum_j a_{k_j}(-1)^{k+j} \det(A(k,j))$ where $A(k,j)$ denotes the submatrix of A obtained by deleting row k and column j, we see that the gradient of $\det(B_+)$ is $adj B_+$, the (symmetric) matrix of signed cofactors. We will also need Cramer's rule, i.e., $B_+^{-1} = (1/\det(B_+)) adj\ B_+$.

We can now differentiate the Lagrangian with respect to B_+, equate the derivative to 0, and solve for B_+.

$$0 = \frac{1}{(\det\ B_+)^{2/n}} \left\{ (\det\ B_+)^{1/n} H_c - \frac{\text{trace}(\ H_c B_+)}{n} (\det\ B_+)^{1/n-1} (adj\ B_+) \right\} + \frac{s\lambda^t + \lambda s^t}{2}$$

or

$$0 = \frac{n}{\text{trace}(\ H_c B_+)} H_c - B_+^{-1} + \frac{n(\det\ B_+)^{1/n}}{\text{trace}(\ H_c B_+)} \frac{s\lambda^t + \lambda s^t}{2}. \tag{5.1}$$

Let

$$\bar{\alpha} = \frac{n}{\text{trace}(\ H_c B_+)}, \qquad \bar{u} = \frac{n(\det\ B_+)^{1/n}}{2\ \text{trace}(\ H_c B_+)} \lambda. \tag{5.2}$$

Then (5.1) becomes

$$B_+^{-1} = \bar{\alpha} H_c + s\bar{u}^t + \bar{u} s^t, \tag{5.3}$$

and so we must show that $\bar{\alpha} = \alpha$ and $\bar{u} = u$ as defined in the theorem. Since $B_+^{-1} y = s$, we obtain

$$\bar{\alpha} H_c y + s\bar{u}^t y + \bar{u} s^t y = s \tag{5.4}$$

or

$$\bar{u} = \frac{s}{s^t y} - \frac{\bar{\alpha} H_c y}{s^t y} - s\left(\frac{\bar{u}^t y}{s^t y}\right). \tag{5.5}$$

Let

$$\beta = \frac{\bar{u}^t y}{s^t y} \tag{5.6}$$

and substitute (5.5) in (5.4). We get

$$\bar{\alpha} H_c y + s\frac{s^t y}{s^t y} - \frac{s y^t \bar{\alpha} H_c y}{s^t y} - s s^t \beta y + \frac{s s^t y}{s^t y} - \frac{\bar{\alpha} H_c y}{s^t y} s^t y - s s^t \beta y = s$$

or

$$-s(2 s^t \beta y) + s\left(1 - \frac{y^t \bar{\alpha} H_c y}{s^t y}\right) = 0.$$

Therefore,

$$\beta = \frac{1}{2 s^t y}\left(1 - \frac{y^t \bar{\alpha} H_c y}{s^t y}\right). \tag{5.7}$$

From (5.5), (5.6), and (5.7), we conclude

$$\bar{u} = \frac{1}{s^t y}(s - \bar{\alpha} H_c y) - \left(\frac{s^t y - y^t \bar{\alpha} H_c y}{2(s^t y)^2}\right) s.$$

We can now substitute for $\bar{u}$ in (5.3) and obtain

$$\begin{aligned}
B_+^{-1} &= \bar{\alpha} H_c + \bar{u} s^t + s \bar{u}^t \\
&= \bar{\alpha} H_c + \frac{s s^t}{s^t y} - \frac{\bar{\alpha} H_c y s^t}{s^t y} - \frac{s s^t}{2 s^t y} + \frac{s s^t}{2 s^t y}\frac{y^t \bar{\alpha} H_c y}{s^t y} \\
&\quad + \frac{s s^t}{s^t y} - \frac{s y^t \bar{\alpha} H_c}{s^t y} - \frac{s s^t}{2 s^t y} + \frac{s s^t}{2 s^t y}\frac{y^t \bar{\alpha} H_c y}{s^t y} \\
&= \bar{\alpha} H_c + \frac{s s^t}{s^t y} - \frac{\bar{\alpha} H_c y s^t + s y^t \bar{\alpha} H_c}{s^t y} + \frac{s s^t y^t \bar{\alpha} H_c y}{(s^t y)^2}.
\end{aligned} \tag{5.8}$$

Note that (5.8) is the BFGS update of (αH_c) and is equivalent to (3.2) with $\hat{\phi} = 0$ (cf. [19, p. 269]). Since the update B_+ is the best possible for our measure and since it is a rank-two update of $(1/\bar{\alpha})B_c$, we conclude that $(1/\bar{\alpha})$ is the constant that makes the BFGS update the best among all rank-two updates, which includes the Broyden class. We can now apply Theorem 3.1, i.e., we want ϕ in (3.7) to be one in order to get the BFGS update. Since $b > 0$, this is equivalent to scaling B_c so that the new a equals b, i.e., $y^t \bar{\alpha} H_c y = b$ or

$$\bar{\alpha} = \frac{b}{a}.$$

The values for the Lagrange multiplier λ and for $\bar{\alpha}$ are given in (5.2). The optimal value

$$\omega(H_c B_+) = \omega((\bar{\alpha} H_c) B_+),$$

by Proposition 2.1(i). Therefore, B_+ is the BFGS update of the sized B_c, i.e., $\phi = 1$, $\tilde{a} = b$, $\tilde{b} = b$, $\tilde{c} = ac/b$, and Lemma 3.1 yields

$$\begin{aligned}\omega(\bar{\alpha} H_c B_+) &= \frac{(2f_1(1) + n - 2)/n}{(f_2(1))^{1/n}} \\ &= \frac{(2 + n - 2)/n}{(b^2/ac)^{1/n}}. \qquad \square\end{aligned}$$

COROLLARY 5.1. *Assume* B_c s.p.d. *and* $ac \neq b^2$. *Then for*

$$\alpha = \frac{b}{c},$$

the DFP *update of* αB_c,

$$B_+ = \alpha B_c + uy^t + yu^t, \qquad u = \frac{y - \alpha B_c s}{b},$$

is the unique solution of

$$\begin{array}{ll}\min & \omega(B_c H_+) \\ \textit{subject to} & H_+ y = s, \quad H_+ \text{s.p.d.}\end{array}$$

In addition, the Lagrange multiplier for the secant equation is uniquely

$$\frac{2(y - \alpha B_c y)}{\alpha b n (\det(B_c B_+^{-1}))^{1/n}},$$

while $\alpha = (n/\text{trace}(H_+ B_c))$, *and the value of the measure is equal to the value in Theorem* 5.1.

Proof. We need only exchange the roles of B_c and H_c in the theorem and note that the optimal value does not change. $\square$

This theorem and corollary state that we obtain the same updated matrix B_+ whether we apply the ω optimal ϕ_* formula to B_c or the sized B_c, or indeed any $\sigma B_c, \sigma > 0$. Therefore, the ω-optimal update of σB_c is also the inverse sized BFGS update of B_c. Similarly, the ω-optimal inverse update of σH_c is the sized DFP update of H_c.

6. Concluding remarks and numerical tests. In this paper we studied the measure ω as it relates to the derivation of updates of the Hessian in least-change secant methods. We have seen, in §5, that sizing of the Hessian arises naturally from this measure. In particular, the inverse-sized BFGS and sized DFP are the optimal ω updates, over all s.p.d. updates that satisfy the secant equation. In §3 we found the optimal values, with respect to the measure ω, of the parameter ϕ that define the Broyden family of rank-two updates. We have also considered weak secant updating in place of sizing in §4. The motivation for this is to avoid large changes in the spectrum after the first sizing step. It is interesting to note that our numerical tests on the standard set of test problems in the literature (see below) show that the best update method is given by inverse sizing at the first step, with weak inverse updating at the subsequent steps, in conjunction with the optimal ϕ at each step. By our results in §4, this is equivalent to inverse sizing at the first step only and using the update in Theorem 4.5 (iv), i.e., the Broyden class update with $\phi = 1/((b/c) - ((b^2/ac) - 1))$ or

$\hat{\phi} = 1 - (b/a)$. The most robust methods in terms of failures due to overflow or too many iterations were the two ω-optimal updates. In fact, they were the most robust even when tested without any line search, i.e., with direct prediction step length one.

Before considering the numerical tests on the standard set of test problems, we consider the following special two-dimensional problem. Powell [1] shows that the DFP update performs far worse than the BFGS update when applied with direct prediction steps to the simple quadratic function

$$f(x) = \tfrac{1}{2}(x_1^2 + x_2^2).$$

He shows that the DFP was far less effective than the BFGS at reducing large eigenvalues. Of course, sizing can reduce large eigenvalues immediately. In this respect, sizing can be considered a "fix" for the DFP. This is corroborated in the following numerical data which compares Powell's data with sized DFP and sized BFGS. The initial Hessian approximation is the diagonal matrix $\mathrm{diag}(1, \lambda_1)$, while the initial point is $x_1 = (\cos\psi_1 \ \sin\psi_1)^t$, where the parameters λ_1 and ψ_1 are given in Tables 6.1–6.4. The numbers represent the number of iterations needed to obtain the condition $||x_{k+1}|| < \epsilon ||x_1||$. The numbers in brackets are for the sized updates.

TABLE 6.1
Number of iterations for the BFGS (*sized* BFGS) *when* $\epsilon = 10^{-4}$.

λ_1 \ ψ_1	20 deg	40 deg	60 deg	70 deg	80 deg	85 deg	87 deg	88 deg
10	5 (9)	6 (10)	7 (7)	8 (6)	7 (3)	6 (9)	5 (9)	4 (9)
100	5 (10)	7 (14)	8 (16)	9 (10)	10 (7)	10 (6)	9 (7)	9 (6)
10^4	5 (14)	7 (27)	8 (30)	9 (20)	11 (12)	12 (8)	13 (11)	14 (13)
10^6	5 (19)	7 (15)	8 (15)	9 (34)	11 (14)	12 (10)	13 (8)	14 (6)
10^9	5 (12)	7 (14)	8 (17)	9 (10)	11 (8)	12 (21)	13 (10)	14 (8)

TABLE 6.2
Number of iterations for the DFP (*sized* DFP) *when* $\epsilon = 10^{-4}$.

λ_1 \ ψ_1	20 deg	40 deg	60 deg	70 deg	80 deg	85 deg	87 deg	88 deg
10	6 (8)	10 (5)	14 (5)	16 (5)	14 (5)	9 (4)	7 (6)	6 (7)
100	8 (8)	15 (5)	29 (6)	47 (6)	89 (8)	106 (8)	84 (7)	59 (6)
1000	10 (8)	19 (5)	45 (6)	83 (7)	230 (8)	549 (10)	855 (10)	1000 (10)
10^4	12 (8)	24 (5)	60 (6)	119 (7)	380 (9)	1141 (10)	2420 (11)	4102 (12)
10^6	15 (8)	34 (5)	92 (6)	181 (7)	752 (9)	3482 (10)	5162 (11)	9194 (11)

TABLE 6.3
Number of iterations for the BFGS (*sized* DFP) *when* $\epsilon = 10^{-6}$.

λ_1 \ ψ_1	20 deg	40 deg	60 deg	70 deg	80 deg	85 deg	87 deg	88 deg
10	6 (9)	7 (6)	9 (6)	9 (5)	8 (5)	7 (7)	6 (7)	5 (8)
100	6 (9)	8 (6)	9 (7)	10 (8)	11 (8)	11 (7)	11 (7)	10 (7)
10^4	6 (9)	8 (6)	10 (7)	11 (9)	12 (10)	14 (10)	15 (10)	15 (11)
10^6	6 (9)	8 (6)	10 (7)	11 (9)	12 (10)	14 (11)	15 (11)	16 (12)
10^9	6 (9)	8 (6)	10 (7)	11 (9)	12 (10)	14 (11)	15 (11)	16 (12)

From Theorems 4.1 and 5.1 and Corollary 5.1, the sized DFP, inverse-sized BFGS, optimal ϕ and inverse optimal $\hat{\phi}$, are all equal in the case $n = 2$. (In fact, Proposition 2.1 (iii) implies that they are also equal to the optimally conditioned sized

TABLE 6.4
Number of iterations for the BFGS (*sized* DFP) *when* $\epsilon = 10^{-9}$.

λ_1 \ ψ_1	20 deg	40 deg	60 deg	70 deg	80 deg	85 deg	87 deg	88 deg
10	7 (10)	9 (7)	10 (7)	10 (7)	10 (6)	8 (5)	7 (8)	6 (9)
100	7 (10)	9 (7)	11 (7)	12 (7)	13 (9)	13 (9)	12 (8)	11 (8)
10^4	7 (10)	9 (7)	11 (7)	12 (8)	14 (10)	15 (11)	16 (12)	17 (13)
10^6	7 (10)	9 (7)	11 (7)	12 (8)	14 (10)	15 (11)	16 (12)	17 (12)
10^9	7 (10)	9 (7)	11 (7)	12 (8)	14 (10)	15 (11)	16 (12)	17 (12)

symmetric rank-one update.) This clearly appears to be the best update in the above tables. We now decrease ϵ. The following results show that we do not appear to lose asymptotically when we use the ω optimal updates in the case $n = 2$.

We have also tested over 30 variations of methods on the quartic problem from [3] and the standard set of 18 test problems from [10]. The tests were done on a SUN SPARC station 1 using a MATLAB translation of the codes in [21]. We include some of the results below. The updating methods we include are:

1. BFGS.
2. Optimal ϕ.
3. Optimal $\hat{\phi}$.
4. Size at first step only; optimal ϕ.
5. Size at first step only; optimal $\hat{\phi}$.
6. Inverse size at first step only; optimal ϕ.
7. Inverse size at first step only; optimal $\hat{\phi}$.
8. Size at first step only; direct shift subsequently; optimal ϕ.
9. Size at first step only; direct shift subsequently; optimal $\hat{\phi}$.
10. Inverse size at first step only; weak inverse update subsequently; optimal ϕ. (equivalently, inverse size at the first step only and use the Broyden class update with $\phi = 1/((b/c) - ((b^2/ac) - 1))$, $\hat{\phi} = 1 - (b/a)$, from Theorem 4.5 (iv)).
11. Inverse size at first step only; weak inverse update subsequently; optimal $\hat{\phi}$.
12. Inverse size; BFGS.
13. Size; DFP.
14. Inverse size at first step only; BFGS.

Note that there is a Broyden class description for the various combinations of shifting and optimal ϕ updating in the above methods. We have only included the ϕ value for method 10, as this is our best method in the following testing.

We have three groups of tests. These groups are distinguished by the type of line search used.

Group 1. Line search with sufficient decrease only (from [21]).

Group 2. Line search with sufficient decrease and curvature condition (from [21]).

Group 3. Direct prediction only, i.e., no line search.

Each group consists of 12 series of tests where each series is distinguished by the scaling used for the initial estimates of the solution. We have scaled the initial estimates with: .1, .5, 1, 2, 3, 4, 5, 6, 7, 8, 10, 20. Each series consists of the 14 methods and the 19 problems. We have used the priority theory of Lootsma and Saaty as done by Hock and Schittkowski in [22]. More precisely, for each series of tests, let

$$S_i = \{j : \text{ problem } j \text{ is solved successfully by method } i\}.$$

The priority theory is based on a pairwise comparison of the method iteration and function evaluation count on the various problems. For example, let t_{ij} be the iteration count of method i on problem j and define the matrix $R = (r_{ik})$ by

$$r_{ik} = \frac{\sum_{j \in S_i \cap S_k} t_{ij}}{\sum_{j \in S_i \cap S_k} t_{kj}}.$$

The matrix R is an approximation of the (reciprocal) matrix $P = (p_{ik})$ with $p_{ik} = (w_i/w_k)$, where the entries w_i are the true expected values for the iteration count of method i on a problem. We normalize so that the sum of the expected values $\sum w_i = 14$, and set $w = (w_1 \cdots w_n)^T$. Thus, if all 14 methods were equivalent, the expected number of iterations would be one for each method. The matrix P is a rank-one positive matrix with $Pw = Nw$, i.e., w is the unique positive eigenvector corresponding to the largest eigenvalue N. We determine the positive eigenvector corresponding to the largest positive eigenvalue of the approximating matrix R. After normalization, this vector w determines the scores presented below to estimate the performance criterion. (Please see [22] for more details.)

We now present some of the numerical results for the 14 methods. The following tables contain the scores for the number of iterations, function evaluations and failures for each of the 14 methods, after taking the average over all of the 12 different scalings. Note that we do not include function evaluations for group 3 as no line search is used.

GROUP 1. *Line search with sufficient decrease.*

Methods:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Iters.:	.862	.862	1.38	.776	1.42	.769	1.35	.922	1.24	.725	1.31	.801	.805	.784
Fn. eval.:	.887	.878	1.34	.794	1.37	.787	1.31	.926	1.21	.746	1.3	.805	.859	.802
Fails.:	4.17	4.33	4.33	4.00	3.75	3.42	3.25	5.17	5.25	3.167	3.92	1.92	2.42	3.08

order of best methods for iterations: 10 6 4 14 12 13 2 1 8 9 11 7 3 5
order of best methods for function evaluations: 10 6 4 14 12 13 2 1 8 9 11 7 3 5
order of best methods for failures: 12 13 14 10 7 6 5 4 11 1 2 3 8 9

GROUP 2. *Line search with sufficient decrease and curvature condition.*

Methods:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Iters.:	.888	.863	1.4	.763	1.38	.731	1.35	.953	1.29	.67	1.33	.812	.815	.743
Fn. eval.:	.910	.876	1.36	.778	1.34	.751	1.31	.951	1.25	.721	1.29	.824	.875	.761
Fails.:	4.	4.	4.33	3.83	4.	3.5	3.33	4.33	4.42	3.42	3.42	2.	2.08	3.25

order of best methods for iterations: 10 6 14 4 12 13 2 1 8 9 11 7 5 3
order of best methods for function evaluations: 10 6 14 4 12 13 2 1 8 9 11 7 5 3
order of best methods for failures: 12 13 14 7 10 11 6 4 1 2 5 9 3 8

GROUP 3. *No line search.*

Methods:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Iters.:	1.02	.917	1.39	.733	1.39	.753	1.35	.921	1.15	.692	1.36	.717	.771	.83
Fails.:	6.17	5.58	6.08	5.08	4.42	4.58	3.84	8.17	7.67	5.75	5.25	3.75	4.75	4.42

order of best methods for iterations: 10 12 4 6 13 14 2 8 1 9 7 11 3 5
order of best methods for failures: 12 7 5 14 6 13 4 2 11 10 3 1 9 8

We now present the averages, for each method, of the different orders from the above three groups of results.

Averages over the orders.

Methods:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Iters.:	8.33	7.	13.3	3.33	13.7	2.67	11.7	8.67	10.	1.	11.3	4.	5.67	4.33
Fn. eval:	8.	7.	13.5	3.5	13.5	2.	12.	9.	10.	1.	11.	5.	6.	3.5
Fails.:	10.3	9.67	12.	7.67	7.	6.	3.67	13.7	13.	6.33	8.	1.	3.33	3.33
Total:	9.	8.	12.9	5.	11.1	3.75	8.75	10.7	11.1	3.	10.	3.13	4.88	3.75

The best method, with respect to iterations and function evaluations, in each group of tests, was method 10, i.e., inverse sizing at the first step only and using the selfscaling Broyden class update from Theorem 4.5 (iv). Thus this new selfscaling update from the Broyden class appears to be very interesting. Also, the methods using optimal ϕ were better than the ones using BFGS. This confirms the conjecture that the weak sizing after the first step is a correction for sizing. It also shows that the optimal ϕ is an improvement over BFGS. The methods using optimal ϕ did very well. The sized methods 12 and 13 also did very well. This shows that sizing is a fix for the DFP method even for higher-dimensional problems, just as it was for the two-dimensional Powell example given above. Moreover, the sizing methods did the best when it came to failures. This is particularly evident when no line search was used. Thus the sizing seems to help the robustness of the updates.

Acknowledgment. The authors thank Natalia Alexandrov of Rice University and Cathy Bakos of University of Toronto for their many hours of computer work in helping with our numerical tests. The authors also thank Rick Beherens from the University of Colorado at Boulder for providing us with the MATLAB programs for the unconstrained optimization routines from [21]. The three groups of tests were done in Matlab using these routines. The authors are also indebted to Andy Conn from the University of Waterloo for many helpful conversations.

REFERENCES

[1] M. J. D. POWELL, *How bad are the* BFGS *and* DFP *methods when the objective function is quadratic*, Math. Programming, 34 (1986), pp. 34–47.

[2] W. C. DAVIDON, *Optimally conditioned optimization algorithms without line searches*, Math. Programming, 9 (1975), pp. 1–30.

[3] R. H. BYRD, R. H. NOCEDAL, AND Y. YUAN, *Global convergence of a class of quasi-Newton methods on convex problems*, SIAM J. Numer. Anal., 24 (1987), pp. 1171–1191.

[4] R. H. BYRD AND R. H. NOCEDAL, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal., 26 (1989), pp. 727–739.

[5] R. FLETCHER, *A new variational result for quasi-Newton formulae*, SIAM J. Optimization, 1 (1991), pp. 18–21.

[6] H. WOLKOWICZ, *Measures for symmetric rank-one updates*, Tech. Report CORR 90-03, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, 1990.

[7] S. S. OREN AND D. G. LUENBERGER, *Self-scaling variable metric* (SSVM) *algorithms, part* I. *Criteria and sufficient conditions for scaling a class of algorithms*, Management Sci., 20 (1974), pp. 845–862.

[8] K. H. PHUA AND D. F. SHANNO, *Numerical comparison of several variable metric algorithms*, J. Optim. Theory Appl., 25 (1978), pp. 507–518.

[9] S. HOSHINO, *A formulation of variable metric methods*, J. Inst. Math. Appl., 10 (1972), pp. 394–403.

[10] B. S. GARBOW, K. E. HILLSTROM, AND J. J. MORÉ, *User guide for* MINPACK-1, Tech. Report ANL-80-74, Argonne National Laboratory, Argonne, IL, 1980. Available from National Technical Information Service, Springfield, Virginia.

[11] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983.

[12] A. W. MARSHALL AND I. OLKIN, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, 1979.

[13] O. L. MANGASARIAN, *Nonlinear Programming*, McGraw-Hill, New York, 1969.

[14] N. KARMARKAR, *A new polynomial-time algorithm for linear programming*, Combinatorica, 4 (1984), pp. 373–395.

[15] R. B. SCHNABEL, *Analysing and improving quasi-Newton methods for unconstrained optimization*, Ph.D thesis, Department of Computer Science, Cornell University, Ithaca, NY, 1977. Also available as TR-77-320.

[16] S. S. OREN AND E. SPEDICATO, *Optimal conditioning of self-scaling variable metric algorithms*, Math. Programming, 10 (1976), pp. 70–90.

[17] J. STOER, *On the convergence rate of imperfect minimization algorithms in Broyden's β class*, Math. Programming, 9 (1975), pp. 313–335.

[18] Y. YUAN AND R. H. BYRD, *Non-quasi-Newton updates for unconstrained optimization*, Tech. Report, Department of Computer Science, University of Colorado at Boulder, Boulder, CO, 1991.

[19] D. G. LUENBERGER, *Linear and Nonlinear Programming*, 2nd ed., Addison-Wesley, Reading, MA, 1984.

[20] J. E. DENNIS JR., D. M. GAY, AND R. E. WELSCH, *An adaptive nonlinear least-squares algorithm*, TOMS, 7 (1981), pp. 348–368.

[21] J. E. DENNIS JR. AND R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983. (In English.) Mir Publishing Office, Moscow, 1988. (In Russian.)

[22] W. HOCK AND K. SCHITTKOWSKI, *A comparative performance evaluation of* 27 *nonlinear programming codes*, Computing, 30 (1983), pp. 335–358.