MEASURES FOR SYMMETRIC RANK-ONE UPDATES

Henry Wolkowicz
University of Waterloo
Department of Combinatorics and Optimization
Waterloo, Ontario, N2L 3G1, Canada

October 19, 1993

Revision of Report CORR 90-03, Jan. 1990.

Abstract

Measures of deviation of a symmetric positive definite matrix from the identity are derived. They give rise to symmetric rank-one, SR1, type updates. The measures are motivated by considering the volume of the symmetric difference of the two ellipsoids, which arise from the current and updated quadratic models in quasi-Newton methods. The measure defined by the problem - maximize the determinant subject to a bound of 1 on the largest eigenvalue - yields the SR1 update. The measure $\sigma(A) = \frac{\lambda_1(A)}{\det(A)^{\frac{1}{n}}}$ yields the optimally conditioned, sized, symmetric rank-one updates, [1, 2]. The volume considerations also suggest a 'correction' for the initial stepsize for these sized updates. It is then shown that the σ-optimal updates, as well as the Oren-Luenberger self-scaling updates [3], are all optimal updates for the κ measure, the ℓ_2 condition number. Moreover, all four sized updates result in the same largest (and smallest) 'scaled' eigenvalue and corresponding eigenvector. In fact, the inverse-sized BFGS is the mean of the σ-optimal updates, while the inverse of the sized DFP is the mean of the inverses of the σ-optimal updates. The difference between these four updates is determined by the middle $n-2$ scaled eigenvalues. The κ measure also provides a natural Broyden class replacement for the SR1 when it is not positive definite.

Keywords: Conditioning, Least-change Secant Methods, Quasi-Newton Methods, Unconstrained Optimization, Sizing, Symmetric Rank-one Update, Volume of Ellipsoid, Condition Number.
Short Title: Measures for SR1 Updates.

1 Introduction

In this paper we consider several new measures of deviation, of a symmetric positive definite matrix, from the identity matrix. These measures yield some well-known quasi-Newton updates.

[0]The author would like to thank the Natural Sciences and Engineering Research Council of Canada for their support.

We consider the unconstrained minimization problem

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

where f is twice continuously differentiable. We let x_c denote the current approximation to a minimizer x^*, B_c is the current approximation to the true Hessian at x_c, and g_c is the gradient at x_c. Secant-type methods are based upon approximating Newton's method by accumulating Hessian approximations using gradient differences. These methods have the property that the next Hessian approximation, or update $B = B_+$, is symmetric positive definite (denoted s.p.d.) and satisfies the secant condition

$$Bs = y \equiv g_+ - g_c \text{ or } B^{-1}y = s \equiv x_+ - x_c.$$

Simultaneously, these methods preserve as much information as possible from the current Hessian approximation B_c. We will use the notation that $B^{-1} = H$ and

$$a = y^t H_c y, \; b = y^t s, \; c = s^t B_c s. \tag{1.1}$$

Throughout the paper we deal with the space of real symmetric $n \times n$ matrices, $n \geq 2$, equipped with the trace inner product, $< A, B >= \text{trace}\langle AB \rangle$, and the induced Frobenius norm, $||A||_F = \langle A, A \rangle^{\frac{1}{2}}$. We assume the curvature condition $b > 0$ and that the current Hessian approximation B_c is positive definite.

Various measures have been used to preserve current Hessian information. Minimizing these measures, subject to the secant condition being satisfied, yields many of the well known s.p.d. updates used to date. For example, the measure $||B_c - B_+||_F$ yields the PSB update; the measure $||H_c - H_+||_F$ yields the Greenstadt update. These updates may not preserve positive definiteness. The weighted Frobenius norm $||T(H_c - H_+)T^t||_F$, where $T^tTs = y$, yields the BFGS update; while $||T^{-t}(B_c - B_+)T^{-1}||_F$ yields the DFP update. They preserve positive definiteness if and only if the curvature condition $b = y^t s > 0$ holds. If the scaling matrix T above satisfies $(T^tT + B_c)s = y$, then we get the symmetric rank-one (denoted SR1) update. (For the above results see e.g. [4, 5, 2].) Each of the pairs of updates, say the pair B_1 and B_2, yields a one parameter family of updates formed from the linear combinations $tB_1 + (1-t)B_2$, $t \in \Re$. In particular, the DFP and BFGS updates yield the Broyden class. The BFGS and DFP updates can also be characterized as the optimal updates for the measure

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

where det denotes determinant and A is chosen to be the scaled updates $H_c^{\frac{1}{2}} B_+ H_c^{\frac{1}{2}}$ and $B_c^{\frac{1}{2}} H_+ B_c^{\frac{1}{2}}$, respectively, see [6]. The measure

$$\kappa(A) = \frac{\lambda_1(A)}{\lambda_n(A)}, \tag{1.3}$$

(the ℓ_2 condition number) where λ_1 and λ_n are the largest and smallest eigenvalues, respectively, has been used to choose an optimally conditioned update in the Broyden class, [7]. (See also [8]). In [9],Theorem 5.1, we showed that the measure

$$\omega(A) = \frac{\mathrm{trace}(A)/n}{\det(A)^{\frac{1}{n}}} \tag{1.4}$$

yields the inverse-sized BFGS update (the BFGS update of $\frac{a}{b}B_c$) and sized DFP update (the DFP update of $\frac{b}{c}B_c$). (The sized updates are often referred to as the Oren-Luenberger self-scaling updates, [3, 10].) The measure ω acts as a condition number in that it provides a deviation from a multiple of the identity as does the ℓ_2 condition number, κ.

As noted above, measures are important in providing theoretical derivations of well-known updates. These measures are also used to derive updates with special constraints, e.g. special sparsity patterns are required in [11]. Measures are also important in convergence analysis, e.g. the measure ψ is used in the convergence analysis in [12]. The new measures we discuss have some additional useful properties. They are similar to the potential functions used in interior point methods. In particular, like potential functions, the optimal points are interior points. In our case, this guarantees positive definiteness of the optimal updates.

The SR1 and sized SR1 updates are the focus of the measures in this paper. The SR1 update has a major drawback in that it is not necessarily positive definite. However, it has some very strong convergence properties. Under certain regularity conditions, the updates converge globally to the true Hessian [13]. Successful numerical tests - in a trust region framework to avoid the possible loss of positive definiteness - has resulted in a renewed interest in the SR1 update, see e.g. [14]. Another method of avoiding the loss of positive definiteness of the SR1, is to size the current update, see [1, 2]. The resulting updates are called the optimal conditioned sized SR1 udpates.

The primary motivation for this paper is to find the 'best' new update B_+, i.e. this update should satisfy the secant equation while preserving the most information from the current udpate B_c. With this aim in mind, we first show that minimizing the volume of the symmetric difference between the two ellipsoids corresponding to B_+ and B_c, is a valid measure for preserving the most information. This hard problem is not solved but rather relaxed in several ways. This leads to the main results in this paper, which are measures yielding SR1 type updates. Adding the restriction that the ellipsoid for B_+ contains (or is contained in) the normalized ellipsoid for B_c, yields the measure

$$\sigma(A) = \frac{\lambda_1(A)}{\det(A)^{\frac{1}{n}}}. \tag{1.5}$$

The optimal updates for this measure are the optimal conditioned, sized, SR1 updates. The sizing factor for the ellipsoids corresponding to B_+ implies that the initial stepsize of 1 may be wrong for these methods. (A similar conclusion, for the Oren-Luenberger self-scaling BFGS method is presented in [15].) Rather than try and correct the stepsize, we normalize both ellipsoids. This yields the measure

$$\max\{\det(A) : \lambda_1(A) \leq 1\}. \tag{1.6}$$

The optimal update for this measure is the SR1 update. In fact, except for the trivial degenerate case when B_c satisfies the secant equation, loss of definiteness of the SR1 update is equivalent to loss of feasibility of the above optimization problem. The final measure that we consider is the ℓ_2 condition number, κ. We characterize completely the optimal updates for this measure. In fact, we show that the ω and σ optimal updates are all κ optimal. Moreover, there is a very close spectral relationship between these four updates. In addition, the existence of a Broyden class κ optimal update complements positive definiteness of the SR1 update, and so it provides a natural replacement for an indefinite SR1.

The rest of the paper is organized as follows. In Section 2 we present several results on volumes of ellipsoids which lead to the measure σ and the measure defined by the maximum determinant problem (1.6). In Section 3 we show that the optimal conditioned SR1 updates arise from the measure σ and so have an optimal volume interpretation. In Section 4 we first show that the volume considerations suggest that the initial stepsize for the optimally conditioned SR1 updates should not be 1. Equivalently, these updates should be resized after they are evaluated. We then show that the SR1 update comes from (1.6) and has an optimal volume interpretation. In Section 5, we present the κ-optimal updates and the interesting spectral relationships among the various sized updates.

2 Volume as a Measure for Least Change

In this section we derive two measures of least change. Both measures arise from relaxations of the problem: approximate a given ellipsoid by another ellipsoid, from within a given set, by minimizing the volume of their symmetric difference. These measures involve the singular values of the product of two s.p.d. matrices. A further relaxation results in more tractable measures involving eigenvalues.

Least change secant methods attempt to find an update B_+ that satisfies the secant equation while simultaneously preserving as much information as possible from the current Hessian approximation B_c. If we assume that the gradient vector g_+ can be a random direction (of norm 1 say), then we can consider that B_+ is preserving the information from B_c when the search directions H_+g_+ and H_cg_+ are close. Thus B_+ is a least change update of B_c if the ellipsoids formed from the images of the unit ball under H_+ and H_c are close. Let us now use the volume as a measure of closeness for ellipsoids. It would be best if we could find the update H_+ so that the volume of the symmetric difference (set union minus intersection) of the updated and current ellipsoids is minimized. With this aim in mind, we first consider two 'optimal' updated ellipsoids. The first ellipsoid minimizes the volume over all ellipsoids containing the current ellipsoid, while the second one maximizes the volume over all ellipsoids contained within the current ellipsoid.

Suppose that B is s.p.d. Denote the ellipsoid for B of radius α by

$$E_\alpha(B) = \{x \in \Re^n : ||Bx|| \leq \alpha\} = \{x \in \Re^n : x^t B^2 x \leq \alpha^2\}; \tag{2.1}$$

denote the ellipsoid corresponding to the square root of B by

$$E_\alpha(B^{\frac{1}{2}}) = \{x \in \Re^n : x^t B x \leq \alpha^2\}. \tag{2.2}$$

Note that the image of the ball under H, $H(E_\alpha(I)) = E_\alpha(B)$. The volume of this ellipsoid is the determinant of H times the volume of $E_\alpha(I)$, i.e.

$$\mathrm{vol}(E_\alpha(B)) = \frac{\alpha^n}{\det(B)} \mathrm{vol}(E_1(I)).$$

Given B fixed, since

$$\lambda_1(B) = \max_{x \in E_1(I)} ||Bx||,$$

the ellipsoid of minimal volume containing $E_1(I)$ is $E_\alpha(B)$ with $\alpha = \lambda_1(B)$. The n-th root of the volume of this ellipsoid leads to our measure

$$\sigma(B) = \frac{\lambda_1(B)}{\det(B)^{\frac{1}{n}}}.$$

The measure σ has several interesting properties similar to the measure ω used in [9].

Proposition 2.1 *The measure $\sigma(B)$ satisfies*

1. $1 \leq \sigma(B) \leq n\omega(B) \leq n\kappa(B) \leq 4n\omega^n(B) \leq 4n\sigma^n(B)$;
2. $\sigma(\alpha B) = \sigma(B)$, for all $\alpha > 0$;
3. *σ is a pseudoconvex function on the set of s.p.d. matrices and thus any stationary point is a global minimizer.*

Proof. The result 1., without the function σ, is given in [9, Prop 2.1]. Including σ follows from the definitions, as does 2. On the set of s.p.d. matrices, the largest eigenvalue is a convex function, see e.g. [16, 17], while $\det(B)^{\frac{1}{n}}$ is an increasing concave function, see e.g. [16, pg 475]. (By increasing we mean isotonic with the Loewner order, i.e. the order $A \geq B$ if $A - B$ is s.p.d.) Thus σ is pseudoconvex, see e.g. [18]. (For our purposes we need to only know that the ratio of a convex and (positive) concave function is pseudoconvex and that a stationary point of a pseudoconvex function is a global minimum.) □

The inequalities in 1. of the Proposition show that σ and ω act as condition numbers in the sense that they provide bounds on the amplification factors for relative errors. Moreover, since the two measures bound each other from below and above, minimizing one would be a compromise for minimizing the other, i.e. minimizing ω would give good approximations for minimizing σ.

We also need the measure

$$s(B) = \frac{||B||_2}{\det(B)^{\frac{1}{n}}},$$

where $||B||_2$ denotes the spectral norm of B, i.e. the largest singular value of B. Note that both measures σ and s are well defined if B has real positive eigenvalues. Moreover, in this case, $\sigma(B) \leq s(B)$ and equality holds if B is s.p.d. However, it can be shown that the two measures are not necessarily isotonic.

We now state the relationship between the measures s, σ and volume. Note that the true measure s, in the theorem, involves singular values, while the relaxed measure σ, in the corollary, involves eigenvalues and is more tractable. The relationship between the two measures is $s(B) = \sqrt{\sigma(B^tB)}$, with $s(B) = \sqrt{\sigma(B^2)}$, if B is s.p.d.

Theorem 2.1 *Suppose that B_c s.p.d. is given and Ω is a closed set of symmetric matrices which intersects the set of s.p.d. matrices. Then:*

a) When it exists, the s.p.d. B_+ in Ω, which yields the ellipsoid $E_\delta(B_+)$ of minimal volume that contains $E_1(B_c)$, is the solution of

$$\min_{\substack{B_+\in\Omega \\ B_+ s.p.d.}} s(H_cB_+) \quad (= \sigma(B_+H_c^2B_+)), \tag{2.3}$$

with $\delta = ||(H_cB_+)||_2$. Moreover, the volume ratio

$$\frac{vol(E_\delta(B_+))}{vol(E_1(B_c))} = s(H_cB_+)^n.$$

b) When it exists, the s.p.d. B_+ in Ω, which yields the ellipsoid $E_\beta(B_+)$ of maximal volume contained in $E_1(B_c)$, is the solution of

$$\min_{\substack{B_+\in\Omega \\ B_+ s.p.d.}} s(H_+B_c) \quad (= \sigma(H_+B_c^2H_+)), \tag{2.4}$$

with $\beta = 1/||(H_+B_c)||_2$. Moreover, the volume ratio

$$\frac{vol(E_1(B_c))}{vol(E_\beta(B_+))} = s(H_+B_c)^n.$$

c) If

$$\Omega = \{B : Bs = y\}, \tag{2.5}$$

then the minimum in both minimization problems is attained.

Proof. Given B_c and B_+ fixed, the ellipsoid of minimal volume for B_+ that contains $E_1(B_c)$ is $E_\delta(B_+)$, where δ is the solution of the generalized eigenvalue problem

$$\delta = \max_{||B_c x|| \leq 1} ||B_+ x||.$$

After substituting $x = H_c y$ with $||y|| \leq 1$, we get $\delta = ||(H_c B_+)||_2$. Therefore, with only B_c fixed, we minimize $\frac{||(H_c B_+)||_2^n}{\det(B_+)}$ or equivalently $s(H_c B_+)$ to find the ellipsoid of minimal volume.

Similarly, the ellipsoid of maximum volume for B_+ that is contained in $E_1(B_c)$ is $E_\beta(B_+)$, where the fact that $E_1(B_+) \subset E_{\frac{1}{\beta}}(B_c)$ if and only if $E_\beta(B_+) \subset E_1(B_c)$ implies that $\beta = 1/||(H_+ B_c)||_2$, i.e. we maximize $\frac{\frac{1}{||(H_+ B_c)||_2^n}}{\det(B_+)}$ or equivalently minimize $s(H_+ B_c)$.

The volume ratios follow from the definitions. To prove c), note that the functions are bounded below by 1 and bounded above by the fact that there is a feasible s.p.d. matrix, since we assume that $b > 0$. Attainment follows from the restriction on the singular values by the secant equation. (The proof of attainment follows the same argument as that given in [9, Lemma 2.1] for the measure ω.) □

Corollary 2.1 *Under the hypotheses of the above Theorem, if the matrices B_+, B_c, H_+, H_c are replaced by their respective square roots, with $\Omega = \{B^{\frac{1}{2}} : Bs = y\}$ in (2.5), then the Theorem holds with $\delta = \sqrt{\lambda_1(H_c B_+)}$ and $\beta = 1/\sqrt{\lambda_1(H_+ B_c)}$ and with the measure s replaced by $\sqrt{\sigma}$, the square root of the measure σ; so that the two objective functions are $\sigma(B_+^{\frac{1}{2}} H_c B_+^{\frac{1}{2}})$ and $\sigma(H_+^{\frac{1}{2}} B_c H_+^{\frac{1}{2}})$, respectively, and the two volume ratios are $\sigma(H_c B_+)^{\frac{n}{2}}$ and $\sigma(H_+ B_c)^{\frac{n}{2}}$, respectively.*

The above Theorem provides updates H_+ so that $H_+(E_\delta(I))$ and $H_+(E_\beta(I))$ approximate $H_c(E_1(I))$. However, for $H_+ g_+$ to properly approximate $H_c g_+$, we would like δ and β to be 1. This is done in the following Theorem and Corollary which is then used in Section 4 to derive the SR1 update.

Theorem 2.2 *Suppose that B_c s.p.d. is given and Ω is a closed set of symmetric matrices which intersects the set of s.p.d. matrices. Then:*

a) If there exists an s.p.d. B in Ω such that $||(BH_c)||_2 \leq 1$, then there exists an s.p.d. update B_+ in Ω which yields the ellipsoid $E_1(B_+)$ of minimal volume that contains $E_1(B_c)$ and it is the solution of

$$\max\{\det(H_c B_+) : B_+ \in \Omega,\ B_+ s.p.d.,\ ||(H_c B_+)||_2 \leq 1\}. \tag{2.6}$$

Moreover, the volume ratio

$$\frac{\text{vol}(E_1(B_+))}{\text{vol}(E_1(B_c))} = \det(B_c H_+).$$

b) If there exists an s.p.d. B in Ω such that $||(HB_c)||_2 \leq 1$, then there exists an s.p.d. update B_+ in Ω which yields the ellipsoid $E_1(B_+)$ of maximal volume that is contained in $E_1(B_c)$ and it is the solution of

$$\max\{\det(H_+B_c) : B_+ \in \Omega,\ B_+ s.p.d.,\ ||(H_+B_c)||_2 \leq 1\}. \tag{2.7}$$

Moreover, the volume ratio

$$\frac{\mathrm{vol}(E_1(B_c))}{\mathrm{vol}(E_1(B_+))} = \det(B_+H_c).$$

Proof. The proof for a) follows by noting that the volume of $E_1(B_+)$ is the volume of $E_1(I)$ divided by $\det(B_+)$ and that

$$E_1(B_c) \subset E_1(B_+) \iff ||(B_+H_c)||_2 \leq 1.$$

The proof of b) is similar. Attainment follows by the constraint on the norm and the fact that, on s.p.d. matrices, the function $\det^{\frac{1}{n}}$ is concave and increasing, or the function $\log\det$ is concave and strictly increasing. (See the proof of Prop. 2.1 and [16].). □

Corollary 2.2 *Under the hypotheses of the above Theorem, if the matrices B, B_+, B_c, H_+, H_c are replaced by their respective square roots, then the Theorem holds with the spectral norm $||\cdot||$ replaced by the square root of the largest eigenvalue $\sqrt{\lambda_1(\cdot)}$. Thus problems (2.6) and (2.7) become (4.2) and (4.3), respectively.*

3 The σ-Optimal Updates

We now show that the best s.p.d. updates for our measure σ are the sized, optimally conditioned, SR1 updates in [1, 2]. Thus these updates provide ellipsoids of minimum (maximum) volume containing (contained in) the current normalized ellipsoid. We again assume that $b > 0$ and B_c is s.p.d.

Theorem 3.1 *Let*

$$\alpha_{\pm} = \frac{c}{b} \pm \{\frac{c^2}{b^2} - \frac{c}{a}\}^{\frac{1}{2}}. \tag{3.1}$$

Then the SR1 update of $\frac{1}{\alpha}B_c$,

$$H_+ = \alpha H_c + vv^t/(v^t y),\ v = s - \alpha H_c y,\ \alpha = \alpha_-, \tag{3.2}$$

is the unique solution of

$$\min_{\substack{B_+ s = y \\ B_+ s.p.d.}} \sigma(H_c B_+).$$

Moreover, $\frac{1}{\alpha} = \lambda_1(H_cB_+)$ is of multiplicity $n-1$ and the other eigenvalue of H_cB_+ is $\lambda_n(H_cB_+) = 1/\alpha_+$.

Proof. The optimum s.p.d. update exists by Corollary 2.1. We first show that we can restrict the problem to sized SR1 updates. The Lagrangian for our problem is identical to the one in the proof of [9, Theorem 5.1], with the possibly nondifferentiable function σ replacing ω, i.e.

$$L(u, B) = \sigma(B) + u^t(BB_c^{\frac{1}{2}}s - H_c^{\frac{1}{2}}y).$$

We have replaced the matrix H_cB_+ with

$$B = H_c^{\frac{1}{2}}B_+H_c^{\frac{1}{2}} \tag{3.3}$$

for simplicity. The secant equation becomes

$$B(B_c^{\frac{1}{2}}s) = (H_c^{\frac{1}{2}}y). \tag{3.4}$$

(The current Hessian approximate becomes the identity, I.) The function $\lambda_1(B)$ is not necessarily a differentiable function in the case of a multiple eigenvalue. However, since it is a real valued convex function, it is continuous and subdifferentiable, see e.g. [19]. The subgradient consists of the convex hull of the normalized m eigenvectors, where m is the multiplicity of the eigenvalue λ_1, see e.g. [16, 17]. We can now find a subgradient of the Lagrangian, using the calculations in the proof of Theorem 5.1 in [9], and set it equal to zero. Using the fact that the derivative of the determinant is the adjoint matrix and that Cramer's rule states that the inverse is the adjoint divided by the determinant, we get

$$H = tY + B_c^{\frac{1}{2}}su^t + u(B_c^{\frac{1}{2}}s)^t, \tag{3.5}$$

where u is a multiple of the Lagrange multiplier vector, Y is a subgradient of $\lambda_1(B)$, and $t = n/\lambda_1(B)$. The subgradient works for these fractional pseudoconvex functions just as it does for convex functions, see e.g. [20, pg 48]. (Y replaces H , the gradient of the trace, and $t = n/\lambda_1(B)$ replaces trace(B). Otherwise the details for the decomposition of H in (3.5) are unchanged from those in [9, Theorem 5.1]. Note that the gradient of u^tBw on the space of symmetric matrices is $\frac{wu^t+uw^t}{2}$.) Moreover, the calculations show that $Y = \sum_i \theta_i x_i x_i^t$, where x_i are the normalized eigenvectors for λ_1, $\theta_i \geq 0$, $\sum_i \theta_i = 1$, $i = 1, \cdots, m$. The above decomposition of H implies that the rank of $Y \geq n - 2$ and so the multiplicity of the largest eigenvalue of B is $\geq n - 2$. Therefore B is at most a rank-two update of a multiple of the identity, λI, where $\lambda = \lambda_1(B)$. If it is rank-two, let us first assume that it is a Broyden class update. Then $\frac{1}{\lambda}B$ has $n - 2$ unit eigenvalues and the other (smaller) two eigenvalues of $\frac{1}{\lambda}B$ are, see [8, pg 111] or [9],

$$\lambda_\pm = f_1(\phi) \pm (f_1(\phi)^2 - f_2(\phi))^{\frac{1}{2}}, \tag{3.6}$$

where ϕ is the scalar for the parameterization of the Broyden class of rank-two updates, and

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

Note that $\frac{1}{\lambda}B$ is a rank-two update of I with secant equation

$$(\frac{1}{\lambda}B)(\lambda^{\frac{1}{2}}B_c^{\frac{1}{2}}s) = (\frac{1}{\lambda^{\frac{1}{2}}}H_c^{\frac{1}{2}}y).$$

Thus a and c are changed to $\frac{1}{\lambda}a$ and λc, respectively. Therefore, since $\sigma(\frac{1}{\lambda}B) = \sigma(B)$, we are minimizing the function

$$\sigma(\frac{1}{\lambda}B)^n = \frac{1}{f_2} = 1/(\frac{ac - \phi(ac - b^2)}{\lambda bc}).$$

This function is isotonic with ϕ, for $\phi < \frac{ac}{ac-b^2}$, which are also the values of ϕ for which the update B is s.p.d. But we must maintain the maximality of λ, i.e. $1 \geq \lambda_+$. In [8, Lemma 7.1.3] it is shown that $\lambda_\pm$ are isotonic with $-\phi$. The optimality conditions imply that λ is the largest eigenvalue. Therefore, we can decrease ϕ until $1 = \lambda_+$, i.e. the multiplicity of λ is at least $n-1$, i.e. B is an SR1 update of a multiple of the identity and equivalently, for some $\alpha > 0$,

$$H = \alpha I + rr^t/(r^t H_c^{\frac{1}{2}}y), \; r = (B_c^{\frac{1}{2}}s) - (\alpha I)(H_c^{\frac{1}{2}}y).$$

The eigenvalues of H are α and

$$\alpha + r^t r/(r^t H_c^{\frac{1}{2}}y) = \alpha + \frac{2\alpha b - c - \alpha^2 a}{\alpha a - b}.$$

The eigenvalue α is the smallest if and only if

$$h(\alpha) = \frac{2\alpha b - c - \alpha^2 a}{\alpha a - b} > 0$$

and the condition number is then the strictly pseudoconvex function

$$\kappa(H) = \kappa(\alpha) = \frac{h(\alpha) + \alpha}{\alpha} = \frac{c - \alpha b}{\alpha b - \alpha^2 a}.$$

Since the multiplicity of the smallest eigenvalue of H is at least $n-1$, cancellation shows that the measure σ is equivalent to κ on SR1 updates of a multiple of the identity, i.e.

$$\sigma(B) = \frac{1/\alpha}{((1/\alpha)^{n-1}(1/(\alpha + h(\alpha))))^{\frac{1}{n}}} = (\frac{\alpha + h(\alpha)}{\alpha})^{\frac{1}{n}}.$$

We can therefore replace σ by κ. Since the optimum of σ is characterized by the eigenvalue configuration and the stationary point property, we see that a unique stationary point α satisfying $h(\alpha) > 0$ must exist. Setting the derivative of κ to 0 yields the stationary points in (3.1). One of these points must correspond to the unique optimum. The numerator of h is ≤ 0, for all α. Therefore, we can assume that the denominator of h is < 0. Since $\alpha_- \leq \alpha_+$, we get that α_- corresponds to the unique optimum. (Note that the case $ac = b^2$ raises no difficulties.) Finally, note that the smaller eigenvalue of B is

$$1/(\alpha_- + h(\alpha_-)) = \frac{\alpha_- a - b}{\alpha_- b - c} = 1/\alpha_+.$$

The case when B is not a Broyden class rank-two update follows similarly. Note that if $B = \lambda I + K$, where K is rank-two, then $y - \alpha s$ is in the range of K; and so K can be written using the two vectors $y - \alpha s$, w, for some $w \in \Re^n$. Therefore we can parametrize the line of updates which join B with the SR1 update of λI. This yields a representation for the eigenvalues of the updates on this line, similar to the representation in (3.6). $\square$

Corollary 3.1 *Let*

$$\hat{\alpha}_{\pm} = \frac{a}{b} \pm \{\frac{a^2}{b^2} - \frac{a}{c}\}^{\frac{1}{2}}. \tag{3.7}$$

Then the SR1 update of $\frac{1}{\hat{\alpha}} H_c$*,*

$$B_+ = \hat{\alpha} B_c + \hat{v}\hat{v}^t / (\hat{v}^t s), \;\; \hat{v} = y - \hat{\alpha} B_c s, \;\; \hat{\alpha} = \hat{\alpha}_-, \tag{3.8}$$

is the unique solution of

$$\min_{\substack{B_+ s = y \\ B_+ s.p.d.}} \sigma(B_c H_+).$$

Moreover, $\frac{1}{\hat{\alpha}} = \lambda_1(B_c H_+)$ *is of multiplicity* $n-1$ *and equals* α_+ *in (3.1); the other eigenvalue of* $B_c H_+$ *is* $1/\hat{\alpha}_+$ *and equals* α_-*, the reciprocal from (3.1); the largest and smallest eigenvalues of* $H_c B_+$ *and the optimal value of the measure, from the theorem and the corollary, all have the same respective values.*

Proof. The proof follows by interchanging the roles of H and B. That the optimal values are the same for both problems can be seen by using the fact that the largest $n-1$ eigenvalues are equal at the optimum in the Theorem while the smallest $n-1$ are equal in the Corollary and $\kappa(B) = \kappa(B^{-1})$. $\square$

4 A Measure for SR1

The optimal conditioned SR1 updates discussed above arise from the volume considerations of Corollary 2.1. Our motivation was that, for H_+ to preserve built up information from H_c, we want $H_+ g_+$ and $H_c g_+$ close for random g_+, i.e. we want the images of the unit balls $H_+(E_1(I))$ and $H_c(E_1(I))$ close in volume. However, Theorem 2.1 a) finds H_+ so that $H_+(E_t(I))$ approximates $H_c(E_1(I))$, where $t = ||H_c B_+||$ is not necessarily 1. Therefore $H_+(tg_+) = (tH_+)(g_+)$ (rather than $H_+(g_+)$) approximates $H_c(g_+)$. This suggests that we scale H_+, or equivalently scale the search direction or initial stepsize, to get $t(H_+ g_+)$, i.e., based on a quadratic model argument which attempts to preserve current Hessian information, the initial stepsize of 1 for these updates is wrong. Note that the initial stepsize can determine whether one update is superior to another since relatively little effort is placed into the line search part of quasi-Newton methods.

For Corollary 2.1 a), we have that $H_+^{\frac{1}{2}}(E_t(I))$ approximates $H_c^{\frac{1}{2}}(E_1(I))$, where $t = \sqrt{\lambda_1(H_cB_+)}$. Therefore $tH_+^{\frac{1}{2}}g_+ \approx H_c^{\frac{1}{2}}g_+$ or

$$tH_c^{\frac{1}{2}}H_+^{\frac{1}{2}}g_+ \approx H_cg_+. \tag{4.1}$$

This suggests a further scaling correction rather than just a stepsize correction.

Rather than correct the stepsize using the optimally conditioned updates that arise from Corollary 2.1 which results in the secant equation not being satisfied, we can try to find the correct minimal volume updates using Theorem 2.2 and Corollary 2.2, i.e. let us find the updates so that $H_+(E_t(I))$ approximates $H_c(E_1(I))$ with $t = 1$. We now see that when the SR1 is s.p.d., then Corollary 2.2 yields the SR1 update,

$$B_+ = B_c + \hat{v}\hat{v}^t/\hat{v}^ts,$$

where $\hat{v} = y - B_cs$. We again assume that $b > 0$ and B_c is s.p.d. (Note that the formula for the SR1 changes if $b^2 = ac$. In this case, the entire Broyden class reduces to the SR1; or it reduces to the rank-zero update B_c, if the latter satisfies the secant equation.)

Theorem 4.1 *Consider the two maximum determinant problems from Corollary 2.2*

$$(i) \qquad \max\{\det(H_cB_+) : B_+ \in \Omega,\ B_+ s.p.d.,\ \lambda_1(H_cB_+) \le 1\}, \tag{4.2}$$

and

$$(ii) \qquad \max\{\det(H_+B_c) : B_+ \in \Omega,\ B_+ s.p.d.,\ \lambda_1(H_+B_c) \le 1\}, \tag{4.3}$$

where Ω, as given by (2.5), defines the set of symmetric matrices satisfying the secant equation. Then:

a) $b > a$ if and only if the SR1 update B_+ is the unique solution of problem (i); in which case $\lambda_1(B_+H_c) = 1$, and problem (ii) is infeasible;

b) $b > c$ if and only if the SR1 update B_+ is the unique solution of problem (ii); in which case $\lambda_n(B_+H_c) = 1$, and problem (i) is infeasible;

c) $b \le \min\{a, c\}$ if and only if the SR1 update is not s.p.d. if and only if the feasible set of both problems (i) and (ii) is the empty set or contains B_c.

Proof. The proof is very similar to the proof of Theorem 3.1. (We refer the reader there for missing details.) Let us consider the first maximization problem (i) given in a). Suppose that $b > a$. For simplicity, we again use the matrix B in (3.3) with corresponding secant equation (3.4) . (Note that the SR1 update is s.p.d. when $\frac{-c}{b-c} < \frac{ac}{ac-b^2}$ or equivalently $\frac{-a}{b-a} < \frac{ac}{ac-b^2}$, see e.g. [8],[9]. Therefore $b > a$ implies that the SR1 update is s.p.d. and in addition, that $b < c$ since $b^2 \le ac$. Moreover the SR1 update has $n-1$ unit eigenvalues and the other eigenvalue is smaller than 1 if $\hat{v}^ts = b - c < 0$, or equivalently if $(s - H_cy)^ty = b - a > 0$. Thus $B \ne I$.) In this case we can take the SR1 update of $(1-\epsilon)I$, where $\epsilon > 0$ is small, and get a feasible update with largest eigenvalue < 1. Therefore the generalized Slater constraint qualification holds, i.e. Lagrange multipliers exist for the problem. The Lagrangian for this first problem is

$$L(\alpha, u, B) = \det(B)^{\frac{1}{n}} - \alpha\lambda_1(B) - u^t(BB_c^{\frac{1}{2}}s - H_c^{\frac{1}{2}}y),$$

where we have added the power $1/n$ to the determinant. Differentiating yields

$$0 = \frac{\det(B)^{\frac{1}{n}}}{n}(\frac{\mathrm{adj}(B)}{\det(B)} - \alpha Y - B_c^{\frac{1}{2}}su^t - u(B_c^{\frac{1}{2}}s)^t). \tag{4.4}$$

If we let the Lagrange multipliers absorb the constants, we get the same decomposition as in the proof of Theorem 3.1

$$H = \alpha Y + B_c^{\frac{1}{2}}su^t + u(B_c^{\frac{1}{2}}s)^t.$$

Since H is s.p.d., we conclude that $\alpha > 0$ or $n = 2$. Therefore, by complementary slackness with the eigenvalue constraint, we see that $\lambda_1(B) = 1$ or $n = 2$. We now conclude that B is at most a rank-two update of I, where $1 = \lambda_1(B)$ if $n > 2$. We first assume that B is Broyden class. (As in Theorem 3.1, we can then generalize this argument to arbitrary rank-two updates.) Therefore we can explicitly write down the objective function to be maximized, i.e.

$$\det(B) = \frac{ac - \phi(ac - b^2)}{bc}.$$

This function is isotonic with $-\phi$, for appropriate ϕ. Therefore, if $\lambda_+ < 1$, we can decrease ϕ and increase $\det(B)$. But this increases the other two eigenvalues $\lambda_\pm(\phi)$ of B. We must maintain the maximality of $\lambda_1 = 1$. We conclude that B is the SR1 update of the identity. This proves necessity and the eigenvalue statement in a). Conversely, if the SR1 is the unique solution of (i), then all the eigenvalues of B are ≤ 1 and, as seen above, this implies that $b > a$.

The optimum solution of the second problem given in b) is similarly solved by the SR1 if and only if $b > c$. We still have to prove the infeasibility claims in a) and b), i.e. that there are no other solutions of (i) (or (ii)) when the SR1 is infeasible. Now suppose that $b > c$ so that $b < a$ and the SR1 update can not solve problem (i) as it is an infeasible point. Then problem (i) is either infeasible or, if there exists a feasible solution B, it cannot have largest eigenvalue < 1. For if it did, then the above argument implies that the SR1 update exists and is optimal. Thus a feasible solution B exists if and only if $\lambda_1(B) = 1$, i.e. there are no strictly feasible points. The generalized Slater constraint qualification fails and, in fact, there can be no Lagrange multipliers at the optimum. (Or, the above implies the existence of a rank-two update which again leads to the SR1.) If the feasible set is a single point, then it is also the optimal point. Otherwise, the feasible set consists of the intersection of the (convex) set of s.p.d. matrices with largest eigenvalue $\lambda_1 \leq 1$ and the (linear manifold) set of matrices satisfying the secant equation. This intersection must be a (convex) subset of the set of matrices with largest eigenvalue $\lambda_1 = 1$. To complete the proof we need only show that this set is empty. If B is any optimal matrix with normalized eigenvectors x_i for the eigenvalue 1, then we can orthogonally decompose

$$B = [X\ V] \begin{bmatrix} I & 0 \\ 0 & \bar{B} \end{bmatrix} [X\ V]^t, \tag{4.5}$$

where X is $n \times k$, $k < n-2$ (otherwise, either $B = I$ or the SR1 can be shown to be s.p.d., by the above rank-two update argument), and X consists of the k orthonormal eigenvectors of B corresponding to the eigenvalue 1, $[X\ V]$ is an orthogonal matrix, and $\lambda_1(\bar{B}) < 1$. The secant equation now becomes

$$\begin{bmatrix} I & 0 \\ 0 & \bar{B} \end{bmatrix} ([X\ V]^t B_c^{\frac{1}{2}} s) = ([X\ V]^t H_c^{\frac{1}{2}} y).$$

We see that we have reduced the problem to a $n-k$ dimensional problem, since $\det(B) = \det(\bar{B})$. But then the optimum must have largest eigenvalue 1, which contradicts the decomposition. The infeasibility statement for problem (ii) in a) follows similarly.

We now prove c). If $b \leq \min\{a, c\}$, then the above infeasibility proof holds step by step except for the statement that $B \neq I$ in (4.5), which required that $b > c$. Since $B = I$ is equivalent to the current update B_c satisfying the secant equation, we have shown that the feasible set of problem (ii) contains B_c. The converse is clear from the definitions. The result for problem (i) follows similarly. □

An alternate interpretation of Theorem 4.1 can be obtained from the fact that, for B_c, B_+ s.p.d., we have

$$\begin{aligned} \lambda_1(B_c^{-1}B_+) \leq 1 \quad &\iff \quad B_c^{-\frac{1}{2}} B_+ B_c^{-\frac{1}{2}} - I \leq 0 \\ &\iff \quad B_+ - B_c \leq 0, \end{aligned}$$

This follows by Sylvester's Theorem of Inertia, see [16] and Section 2. (A similar result holds for strict inequality.) Furthermore, $B_+ - B_c \leq 0$ implies that

$$\lambda_k(B_+) \leq \lambda_k(B_c),\ \forall k, \tag{4.6}$$

which implies that the same ordering holds for the trace and determinant. (Similarly, for $\lambda_1(B_+^{-1}B_c) \leq 1$ we get $B_c - B_+ \leq 0$ and the implication

$$\lambda_k(B_c) \leq \lambda_k(B_+),\ \forall k.) \tag{4.7}$$

5 the κ Measure

We now derive the optimal updates for the κ measure and show that there is a strong relationship between these updates and the various SR1 updates discussed above. In fact, we show that the σ-optimal updates in Section 3 and the ω-optimal updates in [9] are actually κ-optimal as well and have a common spectral property.

Each of the measures ω, ψ, σ lead to a pair of BFGS and DFP type updates. Our measures are motivated by the volume considerations. As mentioned earlier, ideally we would like to minimize the volume of the symmetric difference. One point about the symmetric difference is that if we found a measure for it, then the measure should only lead to a single update rather than a pair of updates. One such measure, that yields only a single update rather than a pair, is the ℓ_2 condition number κ, since the condition numbers of a matrix and its inverse are equal. In [21] it has been shown that the measure κ yields a scaled Broyden class update. We can apply the techniques from the proof of Theorem 3.1 to the measure κ to obtain an explicit representation for the optimal κ update. (As seen in the above proofs, we can assume that $B_c = I$.) We get the stationary point condition

$$0 = \lambda_1' \lambda_n - \lambda_n' \lambda_1 + su^t + us^t,$$

for some u, where $\lambda_i' = x_i x_i^t$ is an element in the subdifferential of λ_i and x_i is a normalized eigenvector. (A rank argument implies that the subdifferentials have to be rank-one, since the rank of $su^t + us^t$ is at most 2.) We now conclude that the span of $\{u, s\}$ equals the span of $\{x_1, x_n\}$, i.e. $s \in \text{span}\{x_1, x_n\}$. The secant condition now implies that y is in this span also. Our problem is reduced to the 2-dimensional subspace span$\{s, y\}$. But the measures κ, ω, σ all have the same optimum in 2-dimensions. (This can be seen from the eigenvalue expansion and has been shown in [9].) Therefore we can use an arbitrary orthonormal basis of span$\{s, y\}$ and find the optimal update, restricted to the 2-dimensional subspace, using the results in Section 3 or the ω-optimal updates in [9]. This yields a rank-two matrix on the 2-dimensional subspace. We can then add on the rank-$(n$-$2)$ matrix on the orthogonal complement and choose arbitrary eigenvalues between λ_1 and λ_n, e.g. we can add on λP where P is the orthogonal projection of rank $n-2$ and $\lambda = (\lambda_1 + \lambda_n)/2$.

To better illustrate the κ-optimal updates, we now characterize the case when there exists one in the Broyden class. Let $Q = I - P$ be the orthogonal projection onto the two dimensional subspace span$\{s, y\}$. In [9] it is shown that, in two dimensions, the optimal update of Q (the identity in the 2-dimensional subspace) for the measures κ, ω, σ is the Broyden class update

$$B_Q = Q - \frac{1}{c} ss^t + \frac{1}{b} yy^t + (1 - \phi) cww^t, \tag{5.1}$$

where $w = \frac{1}{b} y - \frac{1}{c} s$ and

$$\phi = 1 - \frac{(a-b)b}{ac - b^2}. \tag{5.2}$$

(In fact, this update, the inverse-sized BFGS, the sized DFP, and the optimal conditioned SR1 updates are all equal in two dimensions. Note that $Qs = s$ and $Qy = y$.) As in the proof of Theorem 3.1, we can evaluate the two functions f_1, f_2 and obtain the following values for the two nonzero eigenvalues of the scaled update in (5.1):

$$\begin{aligned} \lambda_\pm &= f_1(\phi) \pm (f_1(\phi)^2 - f_2(\phi))^{\frac{1}{2}} \\ &= \frac{a}{b} \pm \{\frac{a^2}{b^2} - \frac{a}{c}\}^{\frac{1}{2}}. \end{aligned} \tag{5.3}$$

This agrees with the results obtained in both Theorem 3.1 and Corollary 3.1. We get the same results if we do the calculation for the eigenvalues of the scaled ω-optimal updates, i.e. the inverse-sized BFGS and sized DFP updates. This shows that both the σ-optimal and ω-optimal updates are actually optimal updates for the measure κ as well. Therefore, the above proof for the κ-optimal updates implies that the largest and smallest eigenvalues, with their corresponding eigenvectors, for each of these four updates have the same respective values, since this is true for all the κ-optimal updates (and their convex combinations). In fact, the values of the sizing factors show that the mean of the two σ-optimal updates is the inverse-sized BFGS update. (A similar result holds for the sized DFP. These means should provide better updates for minimizing the volume of the symmetric difference. We can continue this process and find two new means until a limit is reached.) Therefore, to get H_cB_+ as close to the identity as possible, we should choose the update for which the $n-2$ middle eigenvalues is closest to 1. One κ-optimal update is

$$\alpha P + Q - \frac{1}{c}ss^t + \frac{1}{b}yy^t + (1-\phi)cww^t = \alpha I + (1-\alpha)Q - \frac{1}{c}ss^t + \frac{1}{b}yy^t + (1-\phi)cww^t, \quad (5.4)$$

where ϕ is given in (5.2). This update is in the Broyden class when $\alpha = 1$. We can choose $\alpha = 1$ if and only if the convex hull of the two eigenvalues in (5.3) contains 1. This is equivalent to

$$2ac \geq (a+c)b. \quad (5.5)$$

Note that if (5.5) fails, then either $b > a$ or $b > c$, which by Theorem 4.1 implies that the SR1 is s.p.d., i.e. if there is no κ-optimal update in the Broyden class, then the SR1 update is s.p.d. (Condition (5.5) is the condition that determines the different cases for the optimal κ update restricted to the Broyden class, i.e. it determines when the middle $n-2$ scaled eigenvalues are equal to 1, see [8].)

We summarize some of the above discussion in the following. Note that, for simplicity of notation, we assume that $B_c = I$ in part 1.

Theorem 5.1 *Consider the measures ω, σ, κ and the corresponding four sized updates: the inverse-sized BFGS and sized DFP updates which are optimal for the measure ω, and the two sized, optimal conditioned, SR1 updates which are optimal for the measure σ. Then the following holds:*

1. *The κ-optimal updates (of I) are of the form $B_+ = B_Q + B$, where B_Q is given in (5.1), Q is the projection on span$\{s, y\}$, $P = I - Q$, $PBP = B$, and the eigenvalues of B lie between the eigenvalues of B_Q given in (5.3).*

2. *Each of the four sized updates mentioned above (and their convex combinations) is optimal for the κ measure.*

3. *Each of these four sized updates (and their convex combinations), denoted B_+, yields the same value for the largest (and smallest) eigenvalue, and corresponding eigenvector, for the scaled update $H_c^{\frac{1}{2}}B_+H_c^{\frac{1}{2}}$.*

4. *The mean of the two σ-optimal updates is the inverse-sized BFGS update. The mean of the inverses of the two σ-optimal updates is the inverse of the sized DFP update.*

5. *A κ-optimal update exists in the Broyden class if and only if (5.5) holds. Moreover, if (5.5) fails, then the SR1 is s.p.d.*

Acknowledgement: The author would like to thank John Dennis for many hours of helpful conversations. This paper follows the work done in [9] and a question posed by John Dennis on using the ellipsoids of the quadratic model as a measure. Thanks also go to two anonymous referees for their careful reading and many suggestions which improved the presentation of the paper. In particular, thanks go to one of the referees for correcting several errors, including an error in the volume motivation, and in helping to improve the statement of Theorem 4.1.

References

[1] C.M. IP and M.J. TODD. Optimal conditioning and convergence in rank one quasi-Newton updates. *SIAM J. Numerical Analysis*, 25:206–221, 1988.

[2] M.R. OSBORNE and L.P. SUN. A new approach to the symmetric rank-one updating algorithm. Technical report, Australian National University, 1989.

[3] 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. *Manage. Sci.*, 20:845–862, 1974.

[4] J.E. DENNIS Jr. and R.B. SCHNABEL. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, NJ, 1983. Russian edition, Mir Publishing Office, Moscow, 1988, O. Burdakov, translator.

[5] E. SPEDICATO. A class of rank-one positive quasi-Newton updates for unconstrained minimization. *Mathematische Operations und Statistic, Ser. Optimization*, 14:61–70, 1983.

[6] R. FLETCHER. A new variational result for quasi-Newton formulae. *SIAM J. on Optimization*, 1:18–21, 1991.

[7] W.C. DAVIDON. Optimally conditioned optimization algorithms without line searches. *Math. Prog.*, 9:1–30, 1975.

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

[9] J.E. DENNIS and H. WOLKOWICZ. Sizing and least change secant methods. Technical Report CORR 90-02, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, 1990. SIAM J. Numerical Analysis to appear.

[10] K.H. PHUA and D.F. SHANNO. Matrix conditioning and nonlinear optimization. *Math. Prog.*, 14:145–160, 1984.

[11] P.H. CALAMAI and J. J. MORE. Quasi-Newton updates with bounds. *SIAM J. Numer. Anal.*, 24:1434–1441, 1987.

[12] 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.*, 26:727–739, 1989.

[13] A.R. CONN, N.I.M. GOULD, and P.L. TOINT. Convergence of quasi-Newton matrices generated by the symmetric rank one update. *Math. Prog.*, 50:177–195, 1991.

[14] H.F.H. KHALFAN. *Topics in quasi-Newton methods for unconstrained optimization.* PhD thesis, University of Colorado, 1989.

[15] J. NOCEDAL and Y. YUAN. Analysis of a self-scaling quasi-Newton method. Technical report, Department of Electrical Engineering and Computer Science, Northwestern University, Chicago, Ill, 1991.

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

[17] R. FLETCHER. Semi-definite matrix constraints in optimization. *SIAM J. Control and Optimization*, 23:493–513, 1985.

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

[19] R.T. ROCKAFELLAR. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[20] F.H. CLARKE. *Optimization and Nonsmooth Analysis.* Canadian Math. Soc. Series of Monographs and Advanced Texts. John Wiley & Sons, 1983.

[21] R.H. BYRD. A multiobjective characterization of the Broyden class. In preparation. Presented at ORSA/TIMS, New York, Sept. 1990.