OPTIMAL CONDITIONING IN THE CONVEX
CLASS OF RANK TWO UPDATES[†]

Robert B. Schnabel

TR 76-288

August 1976

Computer Science Department
Cornell University
Ithaca, New York 14853

[†]This research supported by NSF grants 73-03413, contract P04361 of the National Bureau of Economic Research, Cambridge Massachusetts, and a National Science Foundation Graduate Fellowship, forms a portion of the author's doctoral thesis at Cornell University directed by Professor J.E. Dennis.

OPTIMAL CONDITIONING IN THE CONVEX
CLASS OF RANK TWO UPDATES[†]

Robert B. Schnabel
Cornell University

Abstract:

Davidon's new quasi-Newton optimization algorithm selects the new inverse Hessian approximation $\bar{H}$ at each step to be the "optimally conditioned" member of a certain one-parameter class of rank two updates to the last inverse Hessian approximation H . In this paper we show that virtually the same goals of conditioning can be achieved while restricting $\bar{H}$ to the convex class of updates, which are bounded by the popular DFP and BFGS updates. We therefore suggest that Davidon's algorithm, and other algorithms using optimal conditioning, restrict the choice of $\bar{H}$ to members of the convex class.

[†] This research supported by NSF grants 73-03413, contract P04361 of the National Bureau of Economic Research, Cambridge, Massachusetts, and a National Science Foundation Graduate Fellowship, forms a portion of the author's doctoral thesis at Cornell University directed by Professor J.E. Dennis.

1. Introduction

Many algorithms for updating approximations to symmetric (and perhaps positive definite) matrices use a formula of the form

$$\bar{H} = H + \frac{ss^T}{\langle s, y \rangle} - \frac{Hy y^T H}{\langle y, Hy \rangle} + \phi \langle y, Hy \rangle \left[\frac{s}{\langle s, y \rangle} - \frac{Hy}{\langle y, Hy \rangle} \right] \left[\frac{s}{\langle s, y \rangle} - \frac{Hy}{\langle y, Hy \rangle} \right]^T, \quad (1.1)$$

$H, \bar{H} \in L(R^n)$, $s, y \in R^n$. (Fletcher (1970)). The most common application is to unconstrained minimization problems

$$\min f(x), \quad f: R^n \rightarrow R.$$

Typical algorithms start each iteration at point x with gradient $\nabla f(x)$ and H an approximation to $[\nabla^2 f(x)]^{-1}$; take a step

$$\bar{x} = x - \lambda \cdot H \cdot \nabla f(x), \quad \lambda > 0$$

designed so that $f(\bar{x}) < f(x)$; and then use (1.1) to form $\bar{H}$, an approximation to $\nabla^2 f(\bar{x})^{-1}$, with

$$s = \bar{x} - x, \quad y = \nabla f(\bar{x}) - \nabla f(x).$$

Parameter ϕ in (1.1) is a free parameter whose choice affects none of these important properties of the update: i) $\bar{H}y = s$; ii) $(\bar{H} - H)$ is a symmetric matrix of rank two; iii) the column and row spaces of $(\bar{H} - H)$ are spanned by s and Hy .

The two most successful choices of ϕ are $\phi = 0$, the Davidon-Fletcher-Powell (DFP) update (Davidon (1959), Fletcher and Powell (1963)) and $\phi = 1$, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update (Broyden (1970), Fletcher (1970), Goldfarb (1970), Shanno (1970)). The entire family of updates was

first noted by Broyden (1970) and is referred to as the Broyden family; while the $\bar{H}(\phi)$ corresponding to $\phi \in [0,1]$, convex combinations of the DFP and BFGS updates, are referred to as the convex class. All updates in the convex class have the desirable property that if H is positive definite and $\langle s, y \rangle > 0$ then $\bar{H}$ is positive definite.

Davidon (1975) recently introduced an algorithm of this type, but with several significant modifications, which has performed very well in numerical tests. First, he selects s and y in (1.1) so that for quadratic problems, $\bar{H}y_- = Hy_-$, y_- the change in gradients at any previous step. This enables him to reach the exact minimum of a quadratic function in $(n+1)$ or fewer iterations. Secondly, he selects the free parameter ϕ in (1.1) to minimize, among updates giving $\bar{H}$ positive definite, the ratio of the largest to the smallest generalized eigenvalue of $\bar{H}z = \lambda Hz$ (or equivalently for H symmetric and positive definite, the condition number of $H^{-1/2}\bar{H}H^{-1/2}$). This seems advantageous as $\bar{H} \cdot \nabla f(\bar{x})$ will be used in making the next step. Davidon calls his choice of ϕ "optimally conditioned"; it may result in $\bar{H}$ inside or outside the convex class.

Davidon (1975) points out that in some cases, the "optimally conditioned" choice of ϕ may result in a much smaller condition number for $H^{-1/2}\bar{H}H^{-1/2}$ than the common choices, $\phi = 0$ or 1 . It is the purpose of this paper to show that virtually the same improvement can be achieved by selecting the

"optimally conditioned" update within $\phi \in [0,1]$. We show that for cases when Davidon's strategy selects ϕ outside the convex class, our modified strategy picks either the DFP or the BFGS; and the resultant condition number of $H^{-1/2} \bar{H}^{-1/2}$ is less than 32/27 the value it would have been for Davidon's choice of $\bar{H}$. Since the convex class has desirable properties (Fletcher (1970), Stoer (1975)), we therefore recommend restricting optimal conditioning to $\phi \in [0,1]$. We also show that optimal conditioning is likely to become unimportant as we approach the solution to many problems.

The conclusions of this paper hold whether s and y in (1.1) are the normal $(\bar{x} - x)$ and $(\nabla f(\bar{x}) - \nabla f(x))$, the projections of these quantities used in Davidon's new algorithm, or indeed for a very general class of rank two updates (Schnabel (1976a)). Therefore when we speak of the DFP, BFGS, and convex class of updates, we refer to the choices $\phi = 0, 1$, and $\phi \in [0,1]$ in (1.1), independent of the meanings of s and y .

Section 2 reviews Davidon's optimal conditioning strategy, summarized in several graphs which help illustrate many points of this paper. Section 3 contains the results concerning the improvement from Davidon's optimal conditioning, and from optimal conditioning restricted to the convex class. Section 4 explains why these results combined with other properties of the convex class seem to make it desirable to restrict optimal conditioning to the convex class. Numerical testing of this

suggestion will be undertaken in (Schnabel (1976b)).

2. Davidon's Optimal Conditioning

Davidon's (1975) new optimization method considers the class of matrix updates,

$$\bar{H}(:) = H + \frac{ss^T}{\langle s, y \rangle} - \frac{Hy y^T H}{\langle y, Hy \rangle} + \phi \langle y, Hy \rangle \left[\frac{s}{\langle s, y \rangle} - \frac{Hy}{\langle y, Hy \rangle} \right] \left[\frac{s}{\langle s, y \rangle} - \frac{Hy}{\langle y, Hy \rangle} \right]^T, \quad (2.1)$$

$H, \bar{H}(:) \in L(R^n)$, $s, y \in R^n$. This one parameter class of updates contains the DFP ($\phi = 0$) and BFGS ($\phi = 1$) updates. Each update from this class changes H to $\bar{H}$ by adding to H a matrix of rank (at most) two, which is the minimum rank change to H consistent with the following desirable properties for the matrix update: i) $\bar{H}y = s$, the way in which new information is incorporated into the approximation; ii) $\bar{H}$ is symmetric and positive definite (s.p.d.) if 1) H is s.p.d. and 2) $\bar{H}$ p.d. is consistent with $\bar{H}y = s$: $\langle y, \bar{H}y \rangle = \langle y, s \rangle > 0$.

Equation (2.1) satisfies i) for any ϕ ; and (see Fletcher (1970), or below) there exists $k < 0$ such that (2.1) satisfies ii) for all $\phi > k$. Therefore (2.1) satisfies i) and ii) for the DFP and BFGS updates, for the entire convex class ($\phi \in [0, 1]$), and for other choices of ϕ . While this freedom in choosing ϕ has been noted in theory, most algorithms have used either $\phi = 0$, $\phi = 1$, or a combination (Fletcher (1970)). Davidon's new method uses the full freedom in choosing ϕ to achieve a property called optimal conditioning.

By optimal conditioning, Davidson means minimizing among all positive definite $\bar{H}(\phi)$ in (2.1), the ratio of the largest to the smallest generalized eigenvalues of $\bar{H}(\phi) \cdot z = \lambda H z$ (or equivalently, for H s.p.d., the ratio of the largest to the smallest eigenvalue of $H^{-1/2} \bar{H}(\phi) H^{-1/2}$ -- known as the condition number of $H^{-1/2} \bar{H}(\phi) H^{-1/2}$, and denoted by $K(H^{-1/2} \bar{H}(\phi) H^{-1/2})$). By doing this we are trying to control the growth in $\bar{H}$ at each step. This is likely to be advantageous, as $\bar{H}$ is almost certainly used in matrix-vector multiplication, and perhaps in the solution of a linear system.

Davidson therefore solved the following problem: for updates (2.1) with H s.p.d., use the free parameter ϕ to minimise $K(H^{-1/2} \bar{H}(\phi) H^{-1/2})$ among s.p.d. $\bar{H}(\phi)$. In this section we solve this problem by explicitly considering the eigenvalues of $H^{-1/2} \bar{H}(\phi) H^{-1/2}$. This gives us additional information about $K(H^{-1/2} \bar{H}(\phi) H^{-1/2})$, which is reflected in the graphs at the end of this section, and is used in section 3.

Because the lemmas of this section are proven by straightforward calculation, and because they are largely a review of material in (Davidson (1975)), the proofs are omitted.

We first calculate the eigenvalues of $H^{-1/2} \bar{H}(\phi) H^{-1/2}$. From (2.1) we find that $H^{-1/2} \bar{H}(\phi) H^{-1/2}$ is of the form $I + \Delta$, Δ a symmetric matrix of rank at most two. Therefore $H^{-1/2} \bar{H}(\phi) H^{-1/2}$ has $n - 2$ eigenvalues $= 1$, and two remaining eigenvalues which can be determined analytically:

Lemma 2.1 Let H be a s.p.d. matrix $\in L(R^n)$; s and y vectors $\in R^n$, $\langle s, y \rangle \neq 0$; $\bar{H}(\phi) \in L(R^n)$ defined by (2.1). Define

$$a = \langle y, Hy \rangle, \quad b = \langle y, s \rangle, \quad c = \langle s, H^{-1}s \rangle, \quad (2.2)$$

Then $H^{-1/2}\bar{H}(\phi)H^{-1/2}$ exists, has $n - 2$ eigenvalues $= 1$, and two remaining eigenvalues

$$\lambda_{\pm}(\phi) = \frac{b+c}{2b} + \frac{\phi(ac-b^2)}{2b^2} \pm \sqrt{\left[\frac{c-b+\phi(ac-b^2)}{2b}\right]^2 + \frac{(ac-b^2)}{ab}\left[1+\phi\frac{(a-b)}{b}\right]} \quad (2.3)$$

The matrix $\bar{H}(\phi)$ is p.d. if and only if $H^{-1/2}\bar{H}(\phi)H^{-1/2}$ is p.d.. $H^{-1/2}\bar{H}(\phi)H^{-1/2}$ is p.d. if and only if λ_+ and λ_- are positive. Therefore using (2.3) we find:

Lemma 2.2 Let H be a s.p.d. matrix $\in L(R^n)$; s and y vectors $\in R^n$; $\bar{H}(\phi) \in L(R^n)$ defined by (2.1); a, b, c defined by (2.2). Then $\bar{H}(\phi)$ is p.d. if and only if

$$b > 0 \quad \text{and} \quad 1 + \phi \frac{(ac-b^2)}{b^2} > 0. \quad (2.4)$$

Note that if s and Hy are linearly dependent, then each $\bar{H}(\phi)$ in (2.1) reduces to the same symmetric rank one update. Therefore for the remainder of this section we assume s, Hy linearly independent. In this case, since $a = \langle H^{1/2}y, H^{1/2}y \rangle$, $b = \langle H^{1/2}y, H^{-1/2}s \rangle$, $c = \langle H^{-1/2}s, H^{-1/2}s \rangle$, we find by the Cauchy-Schwartz relation that

$$ac - b^2 > 0. \quad (2.5)$$

Therefore (2.4) shows that for H s.p.d. and $\langle s, y \rangle > 0$, $\bar{H}(\phi)$ is s.p.d. for some negative and all non-negative ϕ .

Having identified the eigenvalues of $H^{-1/2}\bar{H}(\phi)H^{-1/2}$ and the conditions for $\bar{H}(\phi)$ p.d., we wish to find the p.d. $\bar{H}(\phi)$ for which $K(H^{-1/2}\bar{H}(\phi)H^{-1/2})$ is minimized. For the remainder of this paper we assume $n > 2$ (we comment briefly on the case $n = 2$ in section 4). Therefore the distinct eigenvalues of $H^{-1/2}\bar{H}(\phi)H^{-1/2}$ are λ_+ , λ_- , and 1, and we wish to find

$$\min \frac{\max\{\lambda_+(\phi), 1\}}{\min\{\lambda_-(\phi), 1\}} \quad \text{subject to } \lambda_-(\phi) > 0. \quad (2.6)$$

We first show that $\lambda_+(\phi)$, $\lambda_-(\phi)$ are monotonically increasing functions of ϕ . This is useful because it shows that (2.6) is equivalent to

$$\min \frac{\lambda_+(\phi)}{\lambda_-(\phi)} \quad \text{subject to } \begin{array}{l} 0 < \lambda_-(\phi) \leq 1 \\ \lambda_+(\phi) \geq 1. \end{array} \quad (2.7)$$

Lemma 2.3 Let $\lambda_{\pm}(\phi)$ be defined by (2.3), $ac-b^2 > 0$. Then

$$\frac{d\lambda_{\pm}(\phi)}{d\phi} = \frac{ac-b^2}{2b^2} \left[1 \pm \frac{\epsilon_1(\phi)}{\sqrt{\epsilon_1(\phi)^2 + \frac{ac-b^2}{b^2}}} \right], \quad \epsilon_1(\phi) = \frac{ab+ac-2b^2}{2ab} \pm \frac{(ac-b^2)}{2b^2}$$

is positive for all ϕ . Therefore $\lambda_{\pm}(\phi)$ are monotonically increasing functions of ϕ .

Hence we solve the optimal conditioning problem by considering (2.7). Note that if there exists a ϕ with $\lambda_-(\phi) = 1$, or with $\lambda_+(\phi) = 1$ and $\lambda_-(\phi) > 0$, then it is an endpoint to the feasible region of (2.7). In either case, $H^{-1/2}\bar{H}(\phi)H^{-1/2}$ would have $n - 1$ unit eigenvalues, and its remaining eigenvalue positive. Therefore $\bar{H}(\phi)$ would be the unique update in the

class (2.1) which changes H by a matrix of rank one. This is the symmetric rank one (SRI) update,

$$\bar{H} \begin{pmatrix} b \\ b-a \end{pmatrix} = H + \frac{(s-Hy)(s-Hy)^T}{\langle s-Hy, y \rangle}. \quad (2.8)$$

Therefore if the SRI update is p.d., then it is an endpoint to the feasible region of (2.7). This is reflected in the graphs at the end of this section.

To solve (2.7) we now require $\frac{d}{d\phi} \frac{\lambda_+(\phi)}{\lambda_-(\phi)}$:

Lemma 2.4 Let $\lambda_{\pm}(\phi)$ be defined by (2.3), $ac-b^2 > 0$, $b > 0$.

Then

$$\begin{aligned} \frac{d}{d\phi} \left(\frac{\lambda_+(\phi)}{\lambda_-(\phi)} \right) &= \varepsilon_2(\phi) [(ac-b^2)\phi - b(c-b)], \quad \varepsilon_2(\phi) > 0 \quad \forall \phi, \\ \varepsilon_2(\phi) &= \lambda_-(\phi)^{-2} \times [\text{discriminant of } \lambda_{\pm}(\phi)]^{-1/2} \times \frac{ac-b^2}{2ab^3}. \end{aligned}$$

Therefore

$$\frac{d}{d\phi} \left(\frac{\lambda_+(\phi)}{\lambda_-(\phi)} \right) \begin{cases} = 0, & \phi = \frac{b(c-b)}{ac-b^2} \triangleq \phi^* \\ > 0, & \phi > \phi^* \\ < 0, & \phi < \phi^* \end{cases} \quad (2.9)$$

Equation (2.9) is the key to the solution and graphical representation of the optimal conditioning problem. It remains only to check whether $\lambda_{\pm}(\phi^*)$ satisfy the constraints of (2.7), $0 < \lambda_-(\phi^*) \leq 1 \leq \lambda_+(\phi^*)$:

Lemma 2.5 Let $\lambda_{\pm}(\phi)$ be defined by (2.3), $ac-b^2 > 0$, $b > 0$,

$\phi^* = \frac{b(c-b)}{ac-b^2}$. Then $\lambda_-(\phi^*) > 0$; and $\lambda_-(\phi^*) \leq 1 \leq \lambda_+(\phi^*)$ if and only if

$$b \leq \frac{2ac}{a+c} \quad (2.10)$$

Probably the easiest way to see what the optimally conditioned (OC) update is when ϕ^* is infeasible to (2.7) is to refer to Figs. 2.1a-c. They explain what we have found from lemmas (2.3), (2.4), (2.5): that if (2.10) is not satisfied, then the constrained problem (2.7) has its minimum point at the boundary to the feasible region of (2.7) where one of $\lambda_{\pm}(\phi) = 1$, so that the OC update is the SR1 (2.8).

Therefore we can now state the solution to the optimal conditioning problem:

Theorem 2.6 (Davidon (1975)) Let H be s.p.d. matrix $\in L(R^n)$, $n > 2$; s, y vectors $\in R^n$, s, y linearly independent, $\langle s, y \rangle > 0$; $\bar{H}(\phi) \in L(R^n)$ defined by (2.1); a, b, c defined by (2.2). Then among all p.d. $\bar{H}(\phi)$, $K(H^{-1/2}\bar{H}(\phi)H^{-1/2})$ is minimized by $\bar{H}(\phi_{OC})$

$$\phi_{OC} = \begin{cases} \frac{b(c-b)}{ac-b^2} & b \leq \frac{2ac}{a+c} \\ \frac{b}{b-a} & b > \frac{2ac}{a+c} \end{cases} \quad (2.11)$$

Figs. 2.1a-c summarize the solution to the optimal conditioning problem. They show the shape of the curve $\frac{\lambda_+(\phi)}{\lambda_-(\phi)}$ as a function of ϕ , for all ϕ satisfying $0 < \lambda_-(\phi) \leq 1 \leq \lambda_+(\phi)$. The point PD, determined from (2.4), divides these $\bar{H}(\phi)$ to its right which are p.d. from the re-

maining $\bar{H}(\cdot)$ which are not. The three figures correspond to three distinct cases into which we find it convenient to treat the optimal conditioning problem. Note that in Case 1, λ_{OC} is in the convex class, while in Cases 2 and 3 it is not. In section 3 we determine, for Cases 2 and 3, how much smaller $\frac{\lambda_+(\lambda_{OC})}{\lambda_-(\lambda_{OC})}$ is than the smallest $\frac{\lambda_+(\phi)}{\lambda_-(\phi)}$, $\phi \in [0,1]$.

Case 1

$$b \leq a, c$$

$$\phi_{OC} = \frac{b(c-b)}{ac-b^2}$$

$$b < \frac{2ac}{a+c} \text{ always}$$

Fig. 2.1a

Case 2

$$a > b > c$$

$$\text{solid line: } b \leq \frac{2ac}{a+c}$$

$$\phi_{OC1} = \frac{b(c-b)}{ac-b^2}$$

$$\text{dashed line: } b > \frac{2ac}{a+c}$$

$$\phi_{OC2} = \frac{b}{b-a}, \text{ SR1}$$

Fig. 2.1b

Case 3

$$a < b < c$$

$$\text{solid line: } b \leq \frac{2ac}{a+c}, \phi_{OC1} = \frac{b(c-b)}{ac-b^2}$$

$$\text{dashed line: } b > \frac{2ac}{a+c}, \phi_{OC2} = \frac{b}{b-a}, \text{ SR1}$$

Fig. 2.1c

$$\text{DFP: } \phi = 0$$

$$\text{PD: } \phi = \frac{-b^2}{ac-b^2}$$

$$\text{SR1: } \phi = \frac{b}{b-a}$$

$$\text{BFGS: } \phi = 1$$

3. The Improvement From Optimal Conditioning

Figures 2.1a-c show the shape of the curve $\frac{\lambda_+(\phi)}{\lambda_-(\phi)}$, $\lambda_{\pm}(\phi)$ the non-unity eigenvalues of $H^{-1/2}\bar{H}(\phi)H^{-1/2}$, for the range of updates $\bar{H}(\phi)$ in the optimal conditioning problem (2.7). Within this range of updates, $\frac{\lambda_+(\phi)}{\lambda_-(\phi)}$ is the condition number of $H^{-1/2}\bar{H}(\phi)H^{-1/2}$, $K(H^{-1/2}\bar{H}(\phi)H^{-1/2})$. The important property of the graphs, given by lemma 2.4, is that $\frac{\lambda_+(\phi)}{\lambda_-(\phi)}$ decreases monotonically for $\phi < \phi_{OC}$, and increases monotonically for $\phi > \phi_{OC}$.

In this section we examine how much improvement we have gotten in $K(H^{-1/2}\bar{H}(\phi)H^{-1/2})$ by selecting $\phi = \phi_{OC}$, rather than the common choices $\phi = 0$ or 1 , the DFP and BFGS updates. In essence we are asking whether the curves in Figs. 2.1a-c are gently or steeply sloped. We let $K(\text{DFP})$, $K(\text{BFGS})$, and $K(\text{OC})$ equal $K(H^{-1/2}\bar{H}(\phi)H^{-1/2})$ with $\phi = 0$, 1 , and ϕ_{OC} respectively. In section 3.1 we show that in Case 1 (Fig. 2.1a), $K(\text{OC})$ may be much less than $K(\text{DFP})$ and $K(\text{BFGS})$. Davidon (1975) has noted this fact. In section 3.2 we show that in Case 2 (Fig. 2.1b), $K(\text{DFP}) < \frac{32}{27} \cdot K(\text{OC})$, but that we may still have $K(\text{BFGS})$ much greater than $K(\text{DFP})$. Arguments of symmetry show that the same conclusions are valid in Case 3 (fig. 2.1c), with the role of the DFP and BFGS updates reversed. In section 3.3, we show that, on a problem where our algorithm achieves superlinear convergence, we expect $K(H^{-1/2}\bar{H}(\phi)H^{-1/2})$ to approach 1 for the optimally conditioned (OC) update and the entire convex class. Recall that our conclusions are independent of the definitions

of s and y in (2.1); and that by DFP and BFGS we simply mean the choices $\phi = 0, 1$ in (2.1), independent of the meanings of s and y .

The remainder of this section will make much use of $K(\text{DFP})$, $K(\text{BFGS})$, and $K(\text{OC})$. These condition numbers are calculated using the general equation (2.3) for $\lambda_{\pm}(\phi)$, the non-unity eigenvalues of $H^{-1/2}\bar{H}(\phi)H^{-1/2}$, with ϕ_{OC} determined by (2.11). In section 2 it was shown that in each of these cases, $0 < \lambda_{-}(\phi) \leq 1 \leq \lambda_{+}(\phi)$, and therefore $K(H^{-1/2}\bar{H}(\phi)H^{-1/2}) = \frac{\lambda_{+}(\phi)}{\lambda_{-}(\phi)}$.

Lemma 3.1 Let H be a s.p.d. matrix $\in L(R^n)$; s and y vectors $\in R^n$, $\langle s, y \rangle > 0$; $\bar{H}(\phi)$ defined by (2.2), ϕ_{OC} defined by (2.11). Then the condition numbers of the DFP, BFGS, and OC updates are:

$$K(\text{DFP}) = [a(b+c)^2 - 2b^3 + (b+c)\sqrt{a^2(b+c)^2 - 4ab^3}]/2b^3 \quad (3.1a)$$

$$K(\text{BFGS}) = [(a+b)^2 c - 2b^3 + (a+b)\sqrt{(a+b)^2 c^2 - 4b^3 c}]/2b^3 \quad (3.1b)$$

$$K(\text{OC}) = \begin{cases} [2ac - b^2 + 2\sqrt{a^2 c^2 - ab^2 c}]/b^2, & b \leq \frac{2ac}{a+c} \\ (a-b)/(b-c), & b > \frac{2ac}{a+c} \quad (\text{OC} = \text{SR1}) \text{ \& } a > b \\ (c-b)/(b-a), & b > \frac{2ac}{a+c} \quad (\text{OC} = \text{SR1}) \text{ \& } b > a \end{cases} \quad \begin{matrix} (3.1c) \\ (3.1d) \\ (3.1e) \end{matrix}$$

3.1 Optimal conditioning in Case 1

Davidon (1975) mentions that if $b \ll a, c$ then $K(\text{OC}) \ll$

$K(\text{DFP})$, $K(\text{BFGS})$. It is important to note that the condition $b \ll a$, c can occur only in Case 1 (Fig. 2.1a), when the OC update is in the convex class. In this case $b < \frac{2ac}{a+c}$, so that $K(\text{OC})$ is given by (3.1c). Therefore Davidon's observation follows directly from (3.1a-c):

Lemma 3.2 Let $K(\text{DFP})$, $K(\text{BFGS})$, $K(\text{OC})$ be given by (3.1a-c), $a, b, c > 0$. Let $\frac{b}{a}, \frac{b}{c} = 0$. Then $K(\text{DFP}) = \frac{ac^2}{b^3}$, $K(\text{BFGS}) = \frac{a^2c}{b^3}$, $K(\text{OC}) = \frac{ac}{b^2}$. Therefore $K(\text{OC}) \ll K(\text{DFP}), K(\text{BFGS})$.

For example, if $\frac{b}{a} = \frac{b}{c} = \epsilon \ll 1$, then $K(\text{DFP}) = K(\text{BFGS}) = \epsilon^{-3}$, $K(\text{OC}) = 4\epsilon^{-2}$, $K(\text{OC}) \ll K(\text{DFP}), K(\text{BFGS})$.

The conditions of lemma 3.2 may occur in practice, especially in an algorithm such as Davidon's where s and y are not the change in point and gradient, but projections of these vectors into some subspace. Therefore in Case 1, from the consideration of conditioning, it may be important to select the OC update rather than the DFP or BFGS. We will show that this can be the case only in Case 1, when the OC update is in the convex class.

Mei (1975) has shown that $K(\text{DFP}) < K(\text{BFGS})$ if and only if $c < a$. In Cases 2 and 3 this is clear from the graphs. It is also true in Case 1, but both the DFP and BFGS may be far more poorly conditioned than the OC update.

3.2 Optimal conditioning in Cases 2 and 3.

We now consider Cases 2 and 3, when the OC update is out-

side the convex class. It is clear from Figs. 2.1b,c that the DFP and BFGS updates are, respectively, the best conditioned updates within the convex class in these cases. We wish to show that in fact they are almost as well conditioned as the OC update.

Comparison of $K(\text{DFP})$ to $K(\text{OC})$ in Case 2 or $K(\text{BFGS})$ to $K(\text{OC})$ in Case 3 turn out to be algebraically equivalent (with the roles of the terms a and c interchanged throughout). Therefore we consider only Case 2 in detail. We must consider both $b \leq \frac{2ac}{a+c}$, when the OC update has $\phi = \frac{b(c-b)}{ac-b^2}$ (solid line in Fig. 2.1b), and $b > \frac{2ac}{a+c}$, when the OC update is the SRL, $\phi = \frac{b}{b-a}$ (dashed line in Fig. 2.1b). We show that for any values of a, b, c consistent with H s.p.d. and $\langle s, y \rangle > 0$, that $K(\text{DFP}) < \frac{9}{8} \cdot K(\text{OC})$ if the OC update isn't the SRL; and that $K(\text{DFP}) < \frac{32}{27} \cdot K(\text{OC})$ if it is.

Lemmas 3.3 and 3.4 assume s and $H y$ linearly independent, as otherwise all updates become the SRL and our results are trivially true. Therefore from section 2, $ac - b^2 > 0$.

Lemma 3.3 Let $H \in L(R^n)$ be a s.p.d. matrix, $n > 2$; s, y vectors $\in R^n$, $s, H y$ linearly independent, $\langle s, y \rangle > 0$; $\bar{H}(\phi) \in L(R^n)$ defined by (2.2). Then if $a > b > c$ and $b \leq \frac{2ac}{a+c}$, the DFP is the best conditioned update in the convex class; $\phi = \frac{b(c-b)}{ac-b^2}$ is the best conditioned p.d. update in the entire class; and $K(\text{DFP}) < \frac{9}{8} \cdot K(\text{OC})$.

Proof The fact that the DFP is best conditioned in the convex class, and that $\kappa = \frac{b(c-b)}{ac-b^2}$ is best conditioned overall, is shown in Section 2 and represented by the solid line in Fig. 2.1b. Their condition numbers, $K(\text{DFP})$ and $K(\text{OC})$, are given by (3.1a) and (3.1c).

$$\text{Let } a = \alpha c, \quad b = \beta c. \quad (3.2)$$

Then $a > b > c$ implies $\alpha > \beta > 1$, and $ac - b^2 > 0$ implies $\alpha > \beta^2$. From $\alpha, \beta > 1$ and $b \leq \frac{2ac}{a+c}$ we find $\beta \leq \frac{2\alpha}{\alpha+1} < 2$, $\alpha \geq \frac{\beta}{2-\beta}$. Therefore we require the maximum value of

$$f_1(\alpha, \beta) \triangleq \frac{K(\text{DFP})}{K(\text{OC})} = \frac{\alpha(\beta+1)^2 - 2\beta^3 + (\beta+1)\sqrt{\alpha^2(\beta+1)^2 - 4\alpha\beta^3}}{2\beta[2\alpha - \beta^2 + 2\sqrt{\alpha^2 - \alpha\beta^2}]} \quad (3.3)$$

$$\text{subject to } \beta \in (1, 2), \alpha \in \left[\frac{\beta}{2-\beta}, \infty\right). \quad (3.4)$$

We show that this maximum has a least upper bound of $\frac{9}{8}$, which is $\lim_{\beta \rightarrow 2} f\left(\frac{\beta}{2-\beta}, \beta\right)$.

We show first that for any fixed $\beta \in (1, 2)$, $f_1(\alpha, \beta)$ is a monotonically decreasing function of α for $\alpha \geq \frac{\beta}{2-\beta}$:

$$\frac{df_1(\alpha, \beta)}{d\alpha} = -\frac{[f_2(\alpha, \beta) + f_3(\alpha, \beta) + f_4(\alpha, \beta)]}{f_5(\alpha, \beta)}, \quad (3.5)$$

$$f_2(\alpha, \beta) \triangleq (\beta-1)^2 [\alpha^2(\beta+1) + \sqrt{\alpha^2 - \alpha\beta^2} \cdot \sqrt{\alpha^2(\beta+1)^2 - 4\alpha\beta^3}],$$

$$f_3(\alpha, \beta) \triangleq [\alpha(\beta-1)^2 + 2\beta^3] \sqrt{\alpha^2(\beta+1)^2 - 4\alpha\beta^3},$$

$$f_4(\alpha, \beta) \triangleq (\beta+1) [\alpha(\beta-1)^2 - 2\beta^3] \sqrt{\alpha^2 - \alpha\beta^2},$$

$$f_5(\alpha, \beta) \triangleq 2\beta^{-1} [2\alpha - \beta^2 + 2\sqrt{\alpha^2 - \alpha\beta^2}]^2 \sqrt{\alpha^2 - \alpha\beta^2} \sqrt{\alpha^2(\beta+1)^2 - 4\alpha\beta^3}.$$

With (α, β) subject to (3.4), the two expressions appearing under square root signs are positive, and

$$f_2(\alpha, \beta), f_3(\alpha, \beta), f_5(\alpha, \beta) > 0 \quad (3.6)$$

Also

$$f_3(\alpha, \beta)^2 - f_4(\alpha, \beta)^2 = \alpha \beta^2 (\beta - 1)^2 [8\alpha \beta [\alpha(\beta + 1)^2 - 4\beta^3] + [\alpha(\beta - 1) - 2\beta^3]^2]. \quad (3.7)$$

The right hand side of (3.7) is positive if (α, β) is subject to (3.4) so that

$$f_3(\alpha, \beta) > |f_4(\alpha, \beta)|. \quad (3.8)$$

Substituting (3.6), (3.8) in (3.5) gives $\frac{df_1(\alpha, \beta)}{d\alpha} < 0$, so that for any fixed $\beta \in (1, 2)$, $f_1(\alpha, \beta)$ is monotonically decreasing for $\alpha \geq \frac{\beta}{2-\beta}$.

Therefore the maximum of (3.5), (3.6) is the maximum of $f_1(\frac{\beta}{2-\beta}, \beta)$, $\beta \in (1, 2)$. We show that $f_1(\frac{\beta}{2-\beta}, \beta)$ is an increasing function of β in this region. Define

$$g_1(\beta) = f_1(\frac{\beta}{2-\beta}, \beta) = [(2\beta^3 - 3\beta^2 + 2\beta + 1) + (\beta^2 - 1)\sqrt{4\beta + 1}] / 2\beta^3$$

Then

$$\begin{aligned} \frac{dg_1(\beta)}{d\beta} &= \frac{g_2(\beta) + g_3(\beta)}{2\beta^4 \sqrt{4\beta + 1}}, \\ g_2(\beta) &\triangleq (-2\beta^3 - \beta^2 + 10\beta + 3), \\ g_3(\beta) &\triangleq (3\beta^2 - 4\beta - 3)\sqrt{4\beta + 1}. \end{aligned}$$

For $\beta \in (1, 2)$, $(4\beta + 1) > 0$, $g_2(\beta) > 0$, and $g_2(\beta) > |g_3(\beta)|$ as

$$g_2(\beta)^2 - g_3(\beta)^2 = 4\beta^4 (2-\beta)(6-\beta) > 0.$$

Therefore $\frac{dg_1(\beta)}{d\beta} > 0$ for $\beta \in (1, 2)$. Therefore the maximum of (3.5), (3.6) has a least upper bound of $g_1(2) = \frac{9}{8}$.

Lemma 3.4 Let H be a s.p.d. matrix $\in L(\mathbb{R}^n)$, $n > 2$; s, y vectors $\in \mathbb{R}^n$, s, Hy linearly independent, $\langle s, y \rangle > 0$; $\bar{H}(\cdot) \in L(\mathbb{R}^n)$

defined by (2.1), a, b, c defined by (2.2). Then if $a > b > c$ and $b \geq \frac{2ac}{a+c}$, the DFP is the best conditioned update in the convex class, the SRL is the best conditioned p.d. update in the entire class, and $K(\text{DFP}) < \frac{32}{27} \cdot K(\text{OC}) (= \frac{32}{27} \cdot K(\text{SRL}))$.

Proof That the DFP is best conditioned in the convex class, and the SRL is best conditioned overall, is shown in section 2 and represented by the dashed line in Fig. 2.1b. Their condition numbers, $K(\text{DFP})$ and $K(\text{OC})$, are given by (3.1a) and (3.1d).

We again define α, β by (3.2). Therefore again $\alpha > \beta > 1$, $\alpha > \beta^2$, and from also $b \geq \frac{2ac}{a+c}$, we find either $\beta \in (1, 2)$ and $\alpha \in (\beta^2, \frac{\beta}{2-\beta}]$ or $\beta \in [2, \infty)$ and $\alpha \in (\beta^2, \infty)$. Therefore we require the maximum value of

$$f_6(\alpha, \beta) = \frac{K(\text{DFP})}{K(\text{OC})} = \frac{\alpha(\beta+1)^2 - 2\beta^3 + (\beta+1)\sqrt{\alpha^2(\beta+1)^2 - 4\alpha\beta^3}}{2\beta^3} \cdot \frac{\beta-1}{\alpha-\beta} \quad (3.9)$$

$$\text{subject to } \beta \in (1, 2), \quad \alpha \in (\beta^2, \frac{\beta}{2-\beta}] \quad (3.10)$$

$$\text{or to } \beta \in [2, \infty), \quad \alpha \in (\beta^2, \infty). \quad (3.11)$$

We show that this maximum has a least upper bound of $\frac{32}{27}$, which is $\lim_{\beta \rightarrow \infty} f(\alpha, \beta)$.

We consider first $f_6(\alpha, \beta)$ as a function of α with $\beta \in (1, \infty)$ fixed. Then

$$\frac{df_6(\alpha, \beta)}{d\alpha} = \frac{(\beta-1)}{2\beta^3} \cdot \frac{1}{(\alpha-\beta)^2} \cdot \frac{1}{\sqrt{\alpha^2(\beta+1)^2 - 4\alpha\beta^3}} \cdot f_7(\alpha, \beta),$$

$$f_7(\alpha, \beta) \triangleq (\beta+1)[\alpha(\beta^2 - 2\beta - 1) + 2\beta^3] + (\beta^2 - 2\beta - 1)\sqrt{\alpha^2(\beta+1)^2 - 4\alpha\beta^3}.$$

The term under the square root is positive for (3.10) or (3.11).

If $\beta \in (1, 1+\sqrt{2})$, then $(\beta^2 - 2\beta - 1) < 0$, and $f_7(\alpha, \beta) = 0$ for $\alpha = -\beta(\beta+1)^2/2(\beta^2 - 2\beta - 1)$, which is the maximum point of (3.9), (3.10), (3.11) for fixed $\beta \in (1, 1+\sqrt{2})$. For this value of α , $f_6(\alpha, \beta) = \frac{4\beta}{3\beta+1}$, which for $\beta \in (1, 1+\sqrt{2})$ has a strict upper bound of $\lim_{\beta \rightarrow 1+\sqrt{2}} \frac{4\beta}{3\beta+1} = 2(2-\sqrt{2}) = 1.172$.

If $\beta \geq 1 + \sqrt{2}$, $(\beta^2 - 2\beta - 1) \geq 0$ and $f_7(\alpha, \beta) > 0$ for (α, β) subject to (3.11). Therefore in this case $\frac{df_6(\alpha, \beta)}{d\alpha} > 0$ for $\alpha > \beta^2$, so that the maximum of (3.9), (3.11) for fixed $\beta \geq 1 + \sqrt{2}$ has upper bound

$$\lim_{\alpha \rightarrow \infty} f_6(\alpha, \beta) = (\beta-1)(\beta+1)^2/\beta^3.$$

This function of β has maximum value $\frac{32}{27} = 1.185$ at $\beta = 3$.

Therefore the maximum value of (3.9) subject to (3.10) or (3.11) has a strict upper bound of $\lim_{\alpha \rightarrow \infty} f_6(\alpha, 3) = \frac{32}{27}$.

In Case 3, $c > b > a$, we prove the analog of lemmas 3.3, 3.4 where the BFGS is substituted for the DFP. The proofs proceed identically if we simply replace (3.2) by

$$c = \alpha a, \quad b = \beta a.$$

Therefore we have proven our main result:

Theorem 3.5 Let H be a s.p.d. matrix $\in L(R^n)$, $n > 2$; s and y vectors $\in R^n$, $\langle s, y \rangle > 0$; $\bar{H}(\phi) \in L(R^n)$ defined by (2.1), a, b, c defined by (2.2). Then for OC Davidson's optimally conditioned update, and OCX the optimally conditioned update among members

of the convex class, $K(OCX) < \frac{32}{27} \cdot K(OC)$.

Furthermore, the OC update is in the convex class, and therefore is the OCX update, if $b \leq a$ and $b \leq c$. Otherwise, the OC update is not in the convex class, and the OCX update is the BFGS if $a < c$, the DFP if $a > c$.

We have shown for the curves in Figs. 2.1b,c, that their lowest point is "hardly lower" than their lowest point in the convex class. We illustrate finally that in Cases 2 and 3, the highest point of the curve in the convex class may be "much higher" than the lowest point. This result is of interest because it shows that, while the optimally conditioned member of the convex class is virtually as well conditioned as the OC update, many other members of the convex class may be far more poorly conditioned. If we are interested in conditioning it may therefore be important to pick the OCX update, and not just any member of the convex class.

We prove a simple lemma which shows that in Case 2, it is possible that $K(DFP) \ll K(BFGS)$. The reverse is clearly true for Case 3 with the roles of the DFP and BFGS updates, and the terms a and c , interchanged.

Lemma 3.6 Let $K(DFP)$, $K(BFGS)$ be given by (3.1a,b), $a > b > c$. Let ϵ, δ be positive numbers $\ll 1$. If $\frac{b}{a} = \epsilon$, $\frac{b}{c} = 1 + \delta$, then $K(DFP) \ll K(BFGS)$.

Proof Write (3.1a,b) as

$$K(\text{DFP}) = \left[\left(1 + \frac{b}{c}\right)^2 - 2\left(\frac{b}{a}\right)\left(\frac{b}{c}\right)^2 + \left(1 + \frac{b}{c}\right) \sqrt{\left(1 + \frac{b}{c}\right)^2 - 4\left(\frac{b}{a}\right)\left(\frac{b}{c}\right)^2} \right] / 2\left(\frac{b}{a}\right)\left(\frac{b}{c}\right)^2 \quad (3.12a)$$

$$K(\text{BFGS}) = \left[\left(1 + \frac{b}{a}\right)^2 - 2\left(\frac{b}{a}\right)^2\left(\frac{b}{c}\right) + \left(1 + \frac{b}{a}\right) \sqrt{\left(1 + \frac{b}{a}\right)^2 - 4\left(\frac{b}{a}\right)^2\left(\frac{b}{c}\right)} \right] / 2\left(\frac{b}{a}\right)^2\left(\frac{b}{c}\right). \quad (3.12b)$$

Substituting $\frac{b}{a} = \epsilon$, $\frac{b}{c} = 1 + \delta$ into (3.12a,b) gives

$$K(\text{DFP}) = \frac{8 + \mathcal{O}(\epsilon + \delta)}{\epsilon^2 [2 + \mathcal{O}(\delta)]} = \epsilon^{-1} [4 + \mathcal{O}(\epsilon + \delta)],$$

$$K(\text{BFGS}) = \frac{2 + \mathcal{O}(\epsilon)}{\epsilon^2 [2 + 2\delta]} = \epsilon^{-2} [1 + \mathcal{O}(\epsilon + \delta)].$$

Therefore since ϵ^{-1} is a large positive number, $K(\text{DFP}) \ll K(\text{BFGS})$.

Therefore from the consideration of the conditioning of $H^{-1/2} \bar{H}(\phi) H^{-1/2}$, selecting the DFP in Case 2 and the BFGS in Case 3 is a good policy, while always selecting one or the other may not be.

3.3 Optimal conditioning and superlinear convergence

Our final aim is to show that for many problems, optimal conditioning will become unnecessary as we approach the solution, because the value of $K(H^{-1/2} \bar{H} H^{-1/2})$ at successive iterations will approach 1 for established choices of $\bar{H}$. Powell (1976) and Broyden, Dennis, and Moré (1973) have shown that the standard DFP and BFGS algorithms (update (2.1) with $s = \bar{x} - x$, $y = \nabla f(\bar{x}) - \nabla f(x)$, $\phi = 0, 1$ respectively) achieve Q-superlinear convergence to the solution x^* of minimization problems satisfying reasonable conditions (an important one being $\nabla^2 f(x^*)$ s.p.d.). We prove a simple lemma showing that, when we achieve

Q-superlinear convergence with the standard BFGS algorithm under the conditions assumed by Broyden, Dennis, and Moré, then the value of $\|\bar{H} - H\|$ at successive iterations approaches 0, and $K(H^{-1/2}\bar{H}H^{-1/2}) \rightarrow 1$. Virtually the same proof can be given for the DFP algorithm, using the formula for the DFP update to H^{-1} .

All norms are l_2 or L_2 .

Lemma 3.7 Let $f: R^n \rightarrow R$ be twice differentiable in the open convex set D , and suppose for some x^* in D and $p > 0$, $K \geq 0$

$$\|\nabla^2 f(x) - \nabla^2 f(x^*)\| \leq K \|x - x^*\|^p$$

where $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is s.p.d. Consider the sequence $(x_1, H_1, x_2, H_2, \dots)$ generated from (x_0, H_0) by the BFGS iteration

$$\begin{aligned} x_{i+1} &= x_i - H_i \nabla f(x_i), \\ H_{i+1} &= H_i + \frac{(s_i - H_i y_i) s_i^T + s_i (s_i - H_i y_i)^T}{\langle s_i, y_i \rangle} - \frac{\langle s_i - H_i y_i, y_i \rangle s_i s_i^T}{\langle s_i, y_i \rangle^2}. \end{aligned}$$

If $\{x_i\}$ converges Q-superlinearly to x^* , then $\lim_{i \rightarrow \infty} \|H_{i+1} - H_i\| = 0$.

If in addition, there exists constant I_1 such that for all $i > I_1$ the H_i 's are s.p.d. and $(\|H_i^{-1}\|)$ is uniformly bounded, then $\lim K(H_i^{-1/2} H_{i+1} H_i^{-1/2}) = 1$.

Proof Define $\nabla^2 f(x^*)^{-1} = A$. Broyden, Dennis, and Moré (1973) show that under the assumptions of lemma 3.7, there exist fixed constants I_2 and $\rho > 1$ such that for all $i > I_2$,

$$\langle s_i, y_i \rangle \geq \frac{2}{3} \|s_i\|^2 / \|A\|, \quad (3.12)$$

$$\frac{1}{\rho} \|s_i\| \leq \|y_i\| \leq \rho \|s_i\| \quad (3.13)$$

$$\frac{\|s_i - A y_i\|}{\|s_i\|} \leq K \|A\| \cdot \max\{(x_{i+1} - x^*)^P, (x_i - x^*)^P\}. \quad (3.14)$$

They also show that under these assumptions

$$\lim_{i \rightarrow \infty} \frac{||(\mathbf{A} - \mathbf{H}_i)\mathbf{y}_i||}{||\mathbf{y}_i||} = 0. \quad (3.15)$$

Therefore since

$$||\mathbf{H}_{i+1} - \mathbf{H}_i|| \leq \frac{2||\mathbf{s}_i - \mathbf{H}_i\mathbf{y}_i|| ||\mathbf{s}_i||}{\langle \mathbf{s}_i, \mathbf{y}_i \rangle} + \frac{||\mathbf{s}_i - \mathbf{H}_i\mathbf{y}_i|| ||\mathbf{y}_i|| ||\mathbf{s}_i||^2}{\langle \mathbf{s}_i, \mathbf{y}_i \rangle^2},$$

using (3.12), (3.13), and the triangle inequality we find

$$\begin{aligned} ||\mathbf{H}_{i+1} - \mathbf{H}_i|| &\leq \frac{||\mathbf{s}_i - \mathbf{H}_i\mathbf{y}_i|| \cdot 3||\mathbf{A}||}{||\mathbf{s}_i||} + \frac{||\mathbf{s}_i - \mathbf{H}_i\mathbf{y}_i|| \cdot \frac{9}{4}\rho ||\mathbf{A}||^2}{||\mathbf{s}_i||} \\ &\leq (3||\mathbf{A}|| + \frac{9}{4}\rho ||\mathbf{A}||^2) \frac{||\mathbf{s}_i - \mathbf{A}\mathbf{y}_i||}{||\mathbf{s}_i||} + \frac{\rho \cdot ||(\mathbf{A} - \mathbf{H}_i)\mathbf{y}_i||}{||\mathbf{y}_i||}. \end{aligned}$$

Therefore using (3.14), (3.15), and $\{\mathbf{x}_i\} \rightarrow \mathbf{x}^*$, $\lim_{i \rightarrow \infty} ||\mathbf{H}_{i+1} - \mathbf{H}_i|| = 0$.

If also $\mathbf{H}_i$ is s.p.d. and $\{||\mathbf{H}_i||\}$ uniformly bounded for $i > I_1$, then for $i > I_1$, $\mathbf{H}_i^{-1/2}$ exists; and since $\mathbf{H}_i^{-1/2} \mathbf{H}_{i+1} \mathbf{H}_i^{-1/2} = \mathbf{H}_i^{-1/2} [\mathbf{H}_i + (\mathbf{H}_{i+1} - \mathbf{H}_i)] \mathbf{H}_i^{-1/2} = \mathbf{I} + \mathbf{H}_i^{-1/2} (\mathbf{H}_{i+1} - \mathbf{H}_i) \mathbf{H}_i^{-1/2}$, we have $\lim_{i \rightarrow \infty} (\mathbf{H}_i^{-1/2} \mathbf{H}_{i+1} \mathbf{H}_i^{-1/2}) = \mathbf{I}$, $\lim_{i \rightarrow \infty} \kappa(\mathbf{H}_i^{-1/2} \mathbf{H}_{i+1} \mathbf{H}_i^{-1/2}) = 1$.

Therefore, $\kappa(\mathbf{H}^{-1/2} \mathbf{H} \mathbf{H}^{-1/2})$ is likely to approach 1 for many problems when we use the BFGS (or DFP) algorithm. We note that the assumption in lemma 3.7 that the $\mathbf{H}_i$'s are s.p.d. and bounded in norm is a very reasonable one for a superlinearly convergent algorithm--it was proven, for example, in the convergence proof of Dennis and Moré (1973).

Lemma 3.7 does not mean that, were we to use the OC update at a near final iteration of an otherwise BFGS (or DFP) algorithm, that the OC choice would be the BFGS (or DFP). Indeed,

the choice could fall into any possibility of Figs. 2.1a-c. However, the curves in Figs. 2.1a-c, while rising at their ends, would virtually equal the line $\frac{\lambda_+(\phi)}{\lambda_-(\phi)} = 1$ throughout a region including the OC update and the entire convex class. Therefore there is little point to the optimal conditioning strategy in this case, as $K(H^{-1/2}\bar{H}H^{-1/2}) \approx 1$ for established and successful choices of $\bar{H}$.

If we use other quasi-Newton classes of updates, such as Davidon's (1975) new method, we hope they will achieve Q-superlinear convergence on problems where the standard BFGS and DFP algorithms do. In this case it is likely that $K(H^{-1/2}\bar{H}H^{-1/2})$ also approaches 1 at successive iterations for all choices of $\bar{H}$ in the convex class, so that the optimal conditioning strategy would also become unimportant as we approach the solution to such problems. For example, in Davidon's new algorithm the "BFGS" update can also be written as

$$\bar{H} = H + \frac{(s-Hy)d^T + d(s-Hy)^T}{\langle d, y \rangle} - \frac{\langle s-Hy, y \rangle dd^T}{\langle d, y \rangle^2} \quad (3.16)$$

where $s = \bar{x} - x$, $y = \nabla f(\bar{x}) - \nabla f(x)$, and d is the projection of s onto some subspace. We then show, under the assumptions of lemma 3.7 but using update (3.16), that if our algorithm converges Q-superlinearly and also $\langle d_i, y_i \rangle \geq \tau \|d_i\| \|y_i\|$ for some $\tau \in (0, 1]$, then again $\lim_{i \rightarrow \infty} \|H_{i+1} - H_i\| = 0$ and

$\lim_{i \rightarrow \infty} K(H_i^{-1/2} H_{i+1} H_i^{-1/2}) = 1$. The assumption on d_i is very

reasonable, as will be discussed below.

Lemma 3.8 Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ satisfy the hypotheses of lemma 3.7.

Consider the sequence $(x, H, x, H, \dots)$ generated from (x, H) by the iteration

$$x_{i+1} = x_i - H_i \nabla f(x_i),$$

$$H_{i+1} \text{ given by (3.16).}$$

Then if $\{x_i\}$ converges Q-superlinearly to x^* , and if there exist constants $\tau \in (0,1)$ and I_0 such that for all $i > I_0$,

$$\langle d_i, y_i \rangle > \tau \|d_i\| \|y_i\| \quad (3.17)$$

then $\lim_{i \rightarrow \infty} \|H_{i+1} - H_i\| = 0$. If in addition, there exists constant

I_1 such that for all $i > I_1$ the H_i 's are s.p.d. and $\{\|H_i^{-1}\|\}$

is uniformly bounded, then $\lim_{i \rightarrow \infty} \kappa(H_i^{-1/2} H_{i+1} H_i^{-1/2}) = 1$.

Proof The proof is similar to that of lemma 3.7, using

(3.17) in place of (3.12). Relations (3.13-15) are still valid.

Therefore we have

$$\|H_{i+1} - H_i\| \leq \frac{2 \|s_i - H_i y_i\| \|d_i\|}{\langle d_i, y_i \rangle} + \frac{\|s_i - H_i y_i\| \|y_i\| \|d_i\|^2}{\langle d_i, y_i \rangle^2},$$

and using (3.12) and (3.13),

$$\|H_{i+1} - H_i\| \leq \frac{\|s_i - H_i y_i\|}{\tau \|y_i\|} + \frac{\|s_i - H_i y_i\|}{\tau^2 \|y_i\|}$$

The remainder of the proof proceeds identically to lemma 3.7.

The conclusions of lemma 3.8 can also be proven for the "DFP" update in Davidon's new method. Here we use the projection c_i of y_i onto some subspace, and require for all $i > I_0$,

$$\langle s_i, c_i \rangle \geq \tau \|s_i\| \|c_i\| \quad (3.18)$$

Note that in Davidon's algorithm $\langle s_i, c_i \rangle = \langle d_i, y_i \rangle$, and therefore (3.17) and (3.18) are closely related. Assumption (3.17) (or (3.18)) is quite likely to be enforced in implementations of Davidon's method-- $\langle d_i, y_i \rangle > 0$ is required for hereditary positive definiteness, and we want a condition like (3.17) to ensure that d_i and y_i don't get too close to being perpendicular. Therefore on problems for which Davidon's new method converges Q-superlinearly, we expect $K(H^{-1/2} \bar{H}^{-1/2})$ to approach 1 at successive iterations for the entire convex class, so that optimal conditioning becomes unimportant as we approach the solution.

4. Summary and recommendations

We have shown that, for a wide class of quasi-Newton updates including Davidon's new method, optimal conditioning will achieve virtually the same objective whether updates are chosen from the entire one-parameter family, or only from the convex class. Such a statement is only of interest, however,

if we have reasons for preferring to select our updates from the convex class. Past theoretical and experimental results indicate that we do.

When the class of updates is the standard Broyden family, vast experimental evidence favors the DFP and BFGS methods, and gives the edge to the BFGS. (Recall that the standard Broyden class is (2.1) with $s = \bar{x} - x$, $y = \nabla f(\bar{x}) - \nabla f(x)$; the DFP and BFGS the choices $\phi = 0, 1$; while other classes have different definitions of s and y , and we still call the choices $\phi = 0, 1$ the "DFP" and "BFGS".) Furthermore, proofs of convergence have been produced for the convex class (Stoer (1975)) but not for methods extending beyond it (except when, with exact line searches, they are all equivalent (Dixon (1972))); and Q-superlinear convergence has been shown only for the DFP and BFGS methods (Broyden-Dennis-Moré (1973), and Powell (1976)). When the class of updates is a different one, including the one in Davidon's new method, theoretical and experimental results are far more limited. However, our experience with the DFP, BFGS, and convex class in the standard Broyden family leads us to trust these updates in new situations; also, theoretical results in (Schnabel (1976b)) still favor the DFP and BFGS formulae as the minimum norm changes to H^{-1} and H , respectively, from the class of updates considered.

Therefore our results, which show that the benefits of optimal conditioning can be attained while restricting consideration to updates in the convex class, suggest to us the following

choice of $\bar{H}$ at each iteration: select the optimally conditioned $\bar{H}$ within the convex class, except select the BFGS if $K(\text{BFGS})$ is within a prescribed tolerance of $K(\text{OC})$. Such an updating scheme would have these advantages: i) it would achieve the objectives of optimal conditioning; ii) it would select the DFP or BFGS update often; iii) it would be likely (see section 3.3) to become the BFGS algorithm close to the solution.

We suggest this restricted optimal conditioning, rather than the simple choice of the DFP and/or BFGS, because our results show that either the DFP or BFGS may be far more poorly conditioned than the other, and that both may be far more poorly conditioned than some other members of the convex class. However, we really do not know yet whether the theoretical attractions of optimal conditioning are mirrored in its algorithmic performance. Experimentation is required to determine whether algorithms using the restricted OC update perform better than those using just the BFGS (and/or DFP); and whether they perform better or worse than algorithms using optimal conditioning from the entire class. Results of such experiments will be reported in (Schnabel (1976b)). Mei (1976) has tested OC in a related algorithm, and his results favor the BFGS.

We note finally that the results of this paper do not apply to problems in only two variables. This is because, in the cases represented by the dashed lines in Figs. 2.1b,c, the OC update is then no longer the SRL, but rather a choice of ϕ

such that both eigenvalues of $H^{-1/2} \bar{H}(\phi) H^{-1/2}$ are > 1 , or both $\in (0,1)$. In such a case the conclusions of section 3 also may not hold-- $K(OC)$ may be much smaller than the condition number of the optimally conditioned member of the convex class. We do not regard this exception to the work of both Davidon and this paper as important, simply because we can handle two variable problems well without requiring techniques like optimal conditioning.

Acknowledgement

My thanks to the National Bureau of Economic Research, Cambridge, Massachusetts, and the Atomic Energy Research Establishment, Harwell, England, for providing the facilities to perform this research.

References

- Broyden, C.G. (1970), "The convergence of a class of double-rank minimization algorithms," Journal of the Institute of Mathematics and its Applications 6, 76-90.
- Broyden, C.G., Dennis, J.E., and Moré, J.J. (1973), "On the local and superlinear convergence of quasi-Newton methods," Journal of the Institute of Mathematics and its Applications 12, 223-246.
- Davidon, W.C. (1959), "Variable metric method for minimization," Argonne National Laboratory Rept. ANL-5990 (Rev.)
- Davidon, W.C. (1975), "Optimally conditioned optimization algorithms without line searches," Mathematical Programming 9, 1-30.
- Dixon, L.C.W. (1972), "All the quasi-Newton family generate identical points," Mathematical Programming 2, 363-388.
- Fletcher, R. (1970), "A new approach to variable metric algorithms," The Computer Journal 13, 317-322.
- Fletcher, R. and Powell, M.J.D. (1963), "A rapidly convergent descent method for minimization," The Computer Journal 6, 163-168.
- Goldfarb, D. (1970), "A family of variable metric updates derived by variational means," Mathematics of Computation 24, 23-26.
- Mei, H.H-W. (1975), "On the conditioning of DFP and BFGS updates for unconstrained optimization," Cornell University Technical Rept. 75-265.

- Mei, H.H.-W. (1976), Ph.D. thesis, Cornell University, to appear.
- Powell, M.J.D. (1976), "Some global convergence properties of a variable metric algorithm for minimization without exact line searches," in Nonlinear Programming. Siam-Ams proceedings, Vol. IX, Cottle, R.W. and Lemke, C.E., eds.
- Schnabel, R.B. (1976a), "Properties of general classes of rank two updates," to appear.
- Schnabel, R.B. (1976b), Ph.D. thesis, Cornell University, to appear.
- Shanno, D.F. (1970), "Conditioning of quasi-Newton methods for function minimization," Mathematics of Computation 24, 647-656.
- Stoer, J. (1975), "On the convergence rate of imperfect minimization algorithms in Broyden's β -class," Mathematical Programming 9, 313-335.