ON A GEOMETRIC INTERPRETATION OF THE POSITIVE DEFINITE SECANT UPDATES BFGS AND DFP FROM THE WEDDERBURN FORMULA

MOODY T. CHU *, R. E. FUNDERLIC † AND GENE H. GOLUB ‡

Abstract. The BFGS update and the DFP update perhaps are, respectively, the most successful Hessian and inverse Hessian approximations for unconstrained minimization problems. This paper describes a geometric meaning of these updates understood from the Wedderburn rank-one reduction formula.

Key words. Rank-one reduction, Wedderburn theorem, BFGS update, DFP update.

AMS(MOS) subject classifications. 65F30,65K05,49D15

1. Introduction.

A straightforward way of solving

$$\min_{x\in R^n} f \tag{1}$$

where $f : R^n \longrightarrow R$ is twice continuously differentiable is to apply Newton's method to the system of equations

$$g(x) := \nabla f(x) = 0. \tag{2}$$

Even when modified to ensure global convergence, however, Newton's method suffers from a main disadvantage in that the Jacobian $G(x)$ of $g(x)$ (i.e., the Hessian of $f(x)$) is often unavailable or very expensive to compute. For this reason, considerable efforts have been made to develop cheap and reasonable approximations to either $G(x)$ or its inverse.

Since the Hessian is always symmetric and often positive definite (especially near the optimal solution), it is important that an approximate Hessian should also possess these properties. Among existing secant methods, perhaps the two most successful Hessian and inverse Hessian approximations in the literature that preserve these properties are the BFGS update and the DFP update [3, 5].

To bring out the notation, we quickly recapitulate these two method as follows. Let x_c and x_+ denote, respectively, the current and the next approximate of the optimal solution. The secant equation for the approximation H_+ of $G(x_+)$ is

$$H_+ s_c = y_c \tag{3}$$

* Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205. This research was supported in part by National Science Foundation under grant DMS-9123448.

† Department of Computer Science, North Carolina State University, Raleigh, North Carolina 27695-8206.

‡ Department of Computer Science, Stanford University, Stanford, CA 94305.

where

$$\begin{aligned} s_c &:= x_+ - x_c \\ y_c &:= g(x_+) - g(x_c). \end{aligned}$$

The BFGS update H_+ from H_c is given by

$$H_+ := H_c + \frac{y_c y_c^T}{y_c^T s_c} - \frac{H_c s_c s_c^T H_c}{s_c^T H_c s_c}. \tag{4}$$

The secant equation for the approximation K_+ of $G(x_+)^{-1}$ is

$$K_+ y_c = s_c. \tag{5}$$

The DFP update K_+ from K_c is given by

$$K_+ := K_c + \frac{s_c s_c^T}{s_c^T y_c} - \frac{K_c y_c y_c^T K_c}{y_c^T K_c y_c}. \tag{6}$$

It is interesting to note the dual relationship between (4) and (6) with the interchanges $H \leftrightarrow K$ and $y_c \leftrightarrow s_c$. Since only the secant equation and the format of the update matter in our discussion that follows, the duality enables us not to differentiate between BFGS and DFP although in practice a general consensus is that the BFGS update performs better than the DFP update.

In this paper, we want to point out that both BFGS and DFP updates are using the Wedderburn rank-one reduction formula [7]. The Wedderburn formula sheds a light on a geometric meaning of the updates. In particular, we show that both the BFGS update and the DFP update are, respectively, the unique least squares approximations of the original approximate Hessian and inverse Hessian subject to some special linear constraints.

We begin in section 2 with a quick survey of the Wedderburn rank reduction formula. Naturally associated with the formula is a projection mapping. This oblique projection plays an important role in characterizing a special $n-1$ dimensional affine subspace. In section 3 we relate the Wedderburn formula to the BFGS and the DFP updates. Our intention is not to redraw these well-known updates, but rather to describe an interesting geometric meaning on what these updates are approximating.

2. Wedderburn Formula.

In this section we introduce the Wedderburn rank reduction formula for a general matrix $A \in R^{m \times n}$.

Wedderburn [7, p.69] has shown that

LEMMA 2.1. *If $x \in R^n$ and $y \in R^m$ are vectors such that $\omega = y^T A x \neq 0$, then the matrix*

$$B := A - \omega^{-1} A x y^T A \tag{7}$$

has rank exactly one less that the rank of A.

Householder [6, p.33] has observed that the converse is also true.

LEMMA 2.2. *Let $u \in R^m$ and $v \in R^n$. Then the rank of the matrix $B = A - \sigma^{-1}uv^T$ is less than that of A if and only if there are vectors $x \in R^n$ and $y \in R^m$ such that $u = Ax$, $v = A^Ty$ and $\sigma = y^TAx$, in which case* $\text{rank}(B) = \text{rank}(A) - 1$.

We shall refer to (7) henceforth as the Wedderburn rank-one reduction formula. The formula can lead to a general matrix factorization process that includes the LDU decomposition, the Lanczos algorithm and the singular value decomposition as special cases [1].

It is also worth mentioning that Cline and Funderlic [2] have further generalized the about results to include the case where a matrix of rank possibly greater than one is subtracted.

LEMMA 2.3. *Suppose $U \in R^{m\times k}$, $R \in R^{k\times k}$ and $V \in R^{n\times k}$. Then*

$$\text{rank}(A - UR^{-1}V^T) = \text{rank}(A) - \text{rank}(UR^{-1}V^T) \tag{8}$$

if and only if there exist $X \in R^{n\times k}$ and $Y \in R^{m\times k}$ such that

$$U = AX, \ \ V = A^TY \ \ \textit{and} \ \ R = Y^TAX. \tag{9}$$

The transformations A and B in (7) are related in an interesting way. Suppose x and y are fixed. For any $z \in R^n$ we may write

$$Bz = A\left(z - \frac{y^TAz}{y^TAx}x\right). \tag{10}$$

Define

$$\mathcal{S}_{A,x,y} := \left\{z - \frac{y^TAz}{y^TAx}x \mid z \in R^n\right\}. \tag{11}$$

Note that x cannot belong to $\mathcal{S}_{A,x,y}$ so long as $y^TAx \neq 0$ x and that $\mathcal{S}_{A,x,y}$ is an $(n-1)$ dimensional linear subspace. Thus the space R^n is the direct sum

$$R^n = \text{span}\{x\} \oplus \mathcal{S}_{A,x,y}. \tag{12}$$

When there is no danger of ambiguity, we sometimes write $\mathcal{S}_{A,x,y}$ as $\mathcal{S}_A$ for brevity. The transformation $\mathcal{P}_{A,x,y}$ (again, abbreviated as $\mathcal{P}_A$) defined by

$$\mathcal{P}_{A,x,y} := I - \frac{xy^TA}{y^TAx} \tag{13}$$

represents the projection mapping from R^n onto $\mathcal{S}_A$ along the vector x and satisfies the relationships

$$B = A\mathcal{P}_A, \tag{14}$$

$$\mathcal{P}_A(R^n) = \mathcal{S}_A. \tag{15}$$

This projection will play a major role in providing us with an insight to the geometric meaning of the BFGS and the DFP updates.

3. A New Look of BFGS.

Since BFGS and DFP are dual to each other in the sense of interchanging $H \leftrightarrow K$ and $y_c \leftrightarrow s_c$, the discussion considered below for the BGFS update applies equally well to the DFP update.

Upon substituting (3) into (4), we observe that

$$H_+ - \frac{H_+ s_c s_c^T H_+}{s_c^T H_+ s_c} = H_c - \frac{H_c s_c s_c^T H_c}{s_c^T H_c s_c}. \tag{16}$$

Comparing with (7), we realize that each side of (16) is employing a Wedderburn rank reduction formula with $x = y = s_c$. Equivalently, (16) may be written as

$$H_+ \mathcal{P}_{H_+} = H_c \mathcal{P}_{H_c} \tag{17}$$

where we shall denote henceforth $\mathcal{P}_H = \mathcal{P}_{H,s_c,s_c}$ and $\mathcal{S}_H = \mathcal{S}_{H,s_c,s_c}$ for any given $H \in R^{n\times n}$ satisfying $s_c^T H s_c \neq 0$.

Motivated by (16), we may regard BFGS as being constructed from two prospects:

(a) We first *tear* the matrix H_c down to a matrix $\tilde{H}_c$ of rank $n-1$ by *pruning* away a rank one matrix. The Wedderburn formula provides us with a recipe on how this should be done. To maintain symmetry, it is necessary that the x and y in (7) are proportional to each other. This pruning process is unique if the null space (spanned by x) of $\tilde{H}_c$ is specified. We choose $x = s_c$.

(b) We then *rebuild* a matrix H_+ by *grafting* a rank one matrix to $\tilde{H}_c$ so that the secant equation is satisfied. Furthermore, the rebuilt matrix must be symmetric and must be such that if it is pruned again according to the process in (a), then the remainder is the same as $\tilde{H}_c$.

We will elaborate more on these points in the following. In particulary, we will show that H_+ is unique.

Let $\{c_1, \ldots, c_{n-1}\}$ of $\mathcal{S}_{H_c}$ be an arbitrary but fixed basis of $\mathcal{S}_{H_c}$. Such a basis is fairly easy to construct. For example, one could simply project an arbitrary basis to $\mathcal{X}_{H_c}$ through $\mathcal{P}_{H_c}$ (See below). Define

$$x_i := \mathcal{P}_{H_+} c_i. \tag{18}$$

for $i = 1, \ldots, n-1$. It is easy to see that $\{x_1, \ldots, x_{n-1}\}$ forms a basis for $\mathcal{S}_{H_+}$. The geometry is illustrated in Figure 1. From (17) we realize that the BFGS update must satisfy, in addition to (3), the equations

$$H_+ x_i = H_c c_i \tag{19}$$

for $i = 1, \ldots, n-1$. Let

$$S := [H_c c_1, \ldots, H_c c_{n-1}, y_c] \tag{20}$$

$$T_{H_+} := [x_1, \ldots, x_{n-1}, s_c] \tag{21}$$

Note that T_{H_+} is nonsingular. Thus H_+ is uniquely determined by

$$H_+ = S T_{H_+}^{-1}. \tag{22}$$

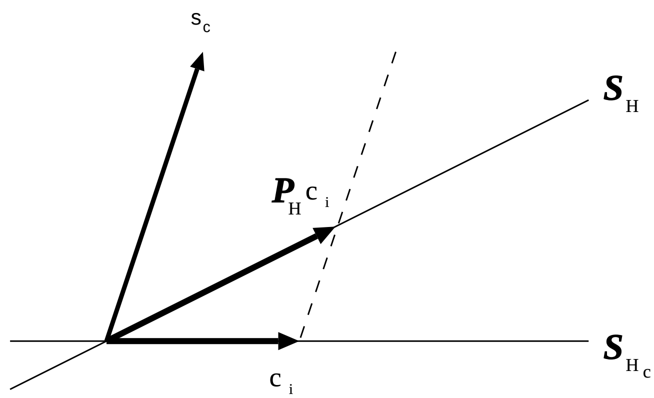

FIG. 1. *Geometry of Basis for $\mathcal{S}_H$.*

The disadvantage of (22) is that T_{H_+} is implicitly defined by H_+. Indeed, given any $H \in R^{n \times n}$ for which $s_c^T H s_c \neq 0$, we may always repeat similarly to (18) to produce a basis $\{\mathcal{P}_H c_1, \ldots, \mathcal{P}_H c_{n-1}\}$ for the subspace $\mathcal{S}_H$ and hence uniquely determine a matrix

$$\Phi(H) := S T_H^{-1} \tag{23}$$

with

$$T_H := [\mathcal{P}_H c_1, \ldots, \mathcal{P}_H c_{n-1}, s_c] \,. \tag{24}$$

It is interesting to note from (22) that the BFGS update H_+ is a fixed point of the mapping Φ.

The following is an important property of $\Phi(H)$.

THEOREM 3.1. *For any $H \in R^{n \times n}$ such that $s_c^T H s_c \neq 0$, it is true that $\mathcal{P}_H = \mathcal{P}_{\Phi(H)}$.*

Proof. We already know that $\mathcal{S}_H$ is spanned by the basis $\{\mathcal{P}_H c_1, \ldots, \mathcal{P}_H c_{n-1}\}$. Observe that

$$\mathcal{P}_{\Phi(H)}(\mathcal{P}_H c_i) = \mathcal{P}_H c_i \tag{25}$$

because

$$s_c^T \Phi(H) \mathcal{P}_H c_i = s_c^T H_c c_i = 0 \tag{26}$$

by (23). It follows that $\{\mathcal{P}_H c_1, \ldots, \mathcal{P}_H c_{n-1}\}$ is also a basis of $\mathcal{S}_{\Phi(H)}$ which is the same as $\mathcal{S}_H$. □

Consider the set

$$\begin{aligned} \mathcal{H} &:= \left\{\hat{H} \in R^{n\times n} \mid \Phi(\hat{H}) = \hat{H}\right\} \\ &= \left\{\hat{H} \in R^{n\times n} \mid \hat{H}\mathcal{P}_{\hat{H}} c_i = H_c c_i \text{ for } i = 1, \ldots, n-1,\ \hat{H}s_c = y_c\right\} \end{aligned} \tag{27}$$

It is easy to check that $\Phi(H) \in \mathcal{H}$ for any $H \in R^{n\times n}$ such that $s_c^T H s_c \neq 0$. But what is the topology of the overall $\mathcal{H}$? We claim that

LEMMA 3.2. *The set $\mathcal{H}$ is an affine subspace. In fact, the dimension of $\mathcal{H}$ is $n-1$.*

Proof. For any $\hat{H} \in \mathcal{H}$, observe that

$$\hat{H}\mathcal{P}_{\hat{H}} c_i = \hat{H}\left(c_i - \frac{s_c^T \hat{H} c_i}{s_c^T \hat{H} s_c} s_c\right) = \hat{H} c_i - \frac{s_c^T \hat{H} c_i}{s_c^T y_c} y_c = \left(I - \frac{y_c s_c^T}{s_c^T y_c}\right) \hat{H} c_i$$

Define

$$\mathcal{P} := I - \frac{y_c s_c^T}{s_c^T y_c}. \tag{28}$$

Then $\mathcal{H}$ may be written as

$$\mathcal{H} = \left\{\hat{H} \in R^{n\times n} \mid \mathcal{P}\hat{H} c_i = H_c c_i \text{ for } i = 1, \ldots, n-1,\ \hat{H}s_c = y_c\right\}. \tag{29}$$

Note $\mathcal{P}$ is now independent of $\hat{H}$.

If $\hat{H}_1, \hat{H}_2 \in \mathcal{H}$, then $\mathcal{P}(\hat{H}_1 - \hat{H}_2)c_i = 0$ for $i = 1, \ldots, n-1$ and $(\hat{H}_1 - \hat{H}_2)s_c = 0$. This shows $\mathcal{H}$ is an affine subspace. It also follows that the tangent space $\mathcal{T}_{\mathcal{H}}(\hat{H})$ of $\mathcal{H}$ at any $\hat{H}$ is given by

$$\mathcal{T}_{\mathcal{H}}(\hat{H}) = \left\{H \in R^{n\times n} \mid \mathcal{P}Hc_i = 0 \text{ for } i = 1, \ldots, n-1,\ Hs_c = 0\right\} \tag{30}$$

For any $H \in \mathcal{T}_{\mathcal{H}}(\hat{H})$, observe that $Hc_i = \alpha_i y_c$ with $\alpha_i := \frac{s_c^T H c_i}{s_c^T y_c}$. We may rewrite this relationship as

$$H = y_c\left[\alpha_1, \ldots, \alpha_{n-1}, 0\right] C \tag{31}$$

where

$$C^{-1} := \left[c_1, \ldots, c_{n-1}, s_c\right].$$

It is easy to see that the converse is also true, i.e., any H of the form (31) with arbitrary $\alpha_1, \ldots, \alpha_{n-1}$ is in $\mathcal{T}_{\mathcal{H}}(\hat{H})$. This shows $\mathcal{H}$ is of dimension $n-1$. □

At this point, there is no guarantee that $\Phi(H)$ will be either symmetric or positive definite even if H is. The theory of the BFGS update indicates that there should be an intersection of the affine subspace $\mathcal{H}$ and the linear subspace $S(n)$ of all symmetric $n \times n$ matrices. Are there any other matrices in the intersection of $\mathcal{H}$ and $S(n)$? We claim

THEOREM 3.3. *Given any $s_c \neq 0$ and y_c, there is one and only matrix in the intersection of $\mathcal{H}$ and $S(n)$, i.e., the one given by (4).*

Proof. We have already proved that $H_+ \in \mathcal{H} \cap S(n)$. On the other hand, any $H \in \mathcal{H}$ satisfies $H\mathcal{P}_H x = H_c \mathcal{P}_{H_c} x$ for all $x \in R^n$. Upon expansion, we have

$$H\mathcal{P}_H x = Hx - \frac{Hs_c s_c^T H}{s_c^T H s_c} x.$$

If H is also symmetric, then we may replace $Hs_c = (s_c^T H)^T$ by y_c and hence prove the assertion. □

It is important to note that $\mathcal{H}$ depends on the basis $c_1, \ldots, c_{n-1}$ of $\mathcal{S}_{H_c}$, but H_+ does not. We should also point out that Theorem 3.3 guarantees that $\mathcal{H} \cap S(n) \neq \emptyset$ and that the intersection is of the form (4), but it still does not guarantee the positive definiteness. As a matter of fact, it is a well known result that H_+ is positive definite if and only if $s_c^T y_c > 0$ [3, 5]. Regardless of the definiteness, however, we now know that the BFGS update is the unique symmetric matrix that acts on the subspace $\mathcal{S}_{H_+}$ in exactly the same way as H_c does on $\mathcal{S}_{H_c}$. We may interpret H_+ as the unique least squares approximation of H_c subject to the constraints that $H_+ \in \mathcal{H}$ and $H_+ \in S(n)$.

It is also interesting to point out an alternative approach for finding $\mathcal{H} \cap S(n)$. The idea is similar to the so called alternating projection method for two convex sets [4]. We define a sequence of matrices $\{H^{(k)}\}$ in $S(n)$ as follows:

$$\begin{aligned} H^{(1)} &:= H_c, \\ H^{(k+1)} &:= \frac{\Phi(H^{(k)}) + \Phi(H^{(k)})^T}{2} \text{ for } k \geq 1. \end{aligned} \tag{32}$$

Observe that for $k \geq 1$,

$$s_c^T H^{(k+1)} s_c = s_c^T \Phi(H^{(k)}) s_c \equiv s_c^T y_c. \tag{33}$$

We may therefore further restrict our attention to the affine subspace $\mathcal{W}$ of $S(n)$ where

$$\mathcal{W} := \left\{ H \in S(n) \mid s_c^T H s_c \equiv s_c^T y_c \right\}. \tag{34}$$

Obviously $H^{(k+1)}$ is the Frobenius norm projection of $\Phi(H^{(k)})$ into the subspace $S(n)$ (and $\mathcal{W}$). Unfortunately $\Phi(H^{(k)})$ is not the Frobenius norm projection of $H^{(k)}$ into $\mathcal{H}$. However, by Theorem 3.1, we notice that $\Phi(H^{(k)})$ is the approximation that makes the least changes of the subspace $\mathcal{S}_{H^{(k)}}$. It can be proved that $H^{(k)}$ converges to H_+.

REFERENCES

[1] M. T. Chu, R. E. Funderlic and G. H. Golub, Wedderburn Decomposition and Its Relations to Other Matrix Factorizations, North Carolina State University, preprint, 1993.

[2] R. E. Cline and R. E. Funderlic, The rank of a difference of matrices and associated generalized inverses, Linear Alg. Appl., 24(1979), 185-215.

[3] J. E. Dennis, Jr. and R. B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice Hall, Englewood Cliffs, 1983.

[4] F. Deutsch, Von Neumann's alternating method: the rate of convergence, Approximation Theory IV, C. Chui, L. Schumaker and J. Ward, ed., Academic Press, New York, 1983, 427-434.
[5] R. Fletcher, Practical Methods of Optimization, 2nd ed., Wiley, Chichester, 1990.
[6] A. S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell, New York, 1964; and Dover, New York, 1975.
[7] J. H. M. Wedderburn, Lectures on Matrices, Colloquium Publications, vol. XVII, A merican Mathematical Society, New York, 1934; and Dover, New York 1964.