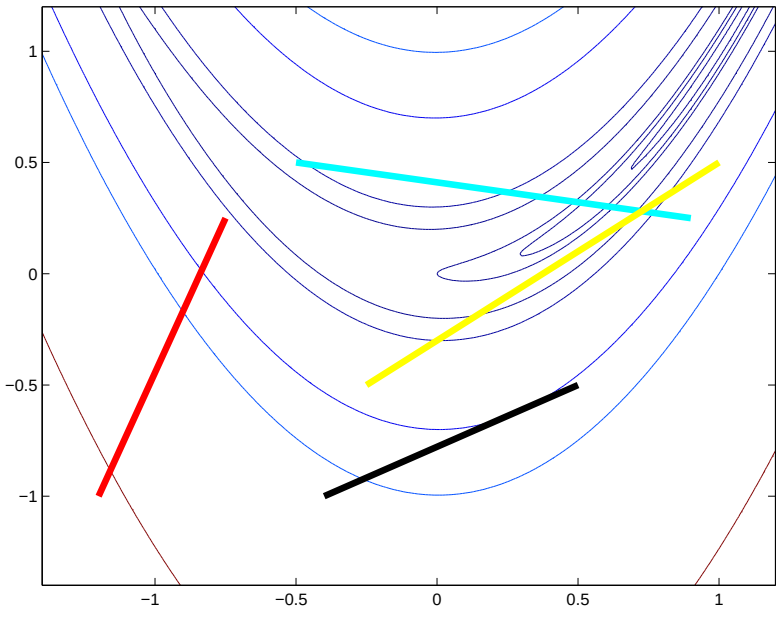
UNIVERSITÉ DE NAMUR

naxys
Namur Center for Complex Systems

Quasi-Newton updates with weighted secant equations

by S. Gratton, V. Malmedy and Ph. L. Toint

Report NAXYS-09-2013 19 October 2013



ENSIEEHT-IRIT, 2, rue Camichel, 31000 Toulouse (France)

LMS SAMTECH, A Siemens Business,15-16, Lower Park Row, BS1 5BN Bristol (UK)

University of Namur, 61, rue de Bruxelles, B5000 Namur (Belgium)

http://www.unamur.be/sciences/naxys

Quasi-Newton updates with weighted secant equations

S. Gratton*, V. Malmedy† and Ph. L. Toint‡

19 October 2013

Abstract

We provide a formula for variational quasi-Newton updates with multiple weighted secant equations. The derivation of the formula leads to a Sylvester equation in the correction matrix. Examples are given.

1 Introduction

Quasi-Newton methods have long been at the core of nonlinear optimization, both in the constrained and unconstrained case. Such methods are characterized by their use of an approximation for the Hessian of the underlying nonlinear function(s) which is typically recurred by low rank modifications of an initial estimate. The corresponding updates are derived variationally to enforce minimal correction in a suitable norm while guaranteeing the preservation of available "secant equations" which provide information about local curvature. By far the most common case is when a single such secant equation is considered for updating the Hessian approximation, and several famous formulae have been derived in this context, such as the DFP (Davidon, 1959, Fletcher and Powell, 1963), BFGS (Broyden, 1970, Fletcher, 1970, Goldfarb, 1970, Shanno, 1970) or PSB (Powell, 1970) updates. More unusual is the proposal by Schnabel (1983) (see also Byrd, Nocedal and Schnabel, 1994) to consider the simultaneous enforcement of several secant equations.

The purpose of this small note, whose content follows up from Malmedy (2010), is to pursue this line of thought, but in a slightly relaxed context where the deviation from the (multiple) secant equations is penalized rather than strictly forced to zero. Our development is primarily motivated by an attempt to explain the influence of the order in which secant pairs are considered in limited-memory BFGS, as described in Malmedy (2010), where a weighed combination of secant updates is used to conduct this investigation. This motivation was recently reinforced by a presentation of a stochastic context in which secant equations are enforced on average (Nocedal, 2013).

This short paper is organized as follows. The new penalized multiple secant update is derived in Section 2, while a few examples are considered in Section 3. Conclusions are drawn in Section 4.

2 Penalized multisecant quasi-Newton updates

We consider deriving a new quasi-Newton approximation B^+ for the Hessian of a nonlinear function from $\mathbb{R}^n$ into $\mathbb{R}$ from an existing one B, while at the same time attempting to satisfy m secant equations. More specifically, we assume that m "secant" pairs of the form (s_i, y_i) $(i = 1, \ldots, m)$ are at our disposal. These pairs are supposed to contain useful curvature information and are typically derived by computing differences y_i in the gradient of the underlying nonlinear function corresponding to steps s_i in the variable space $\mathbb{R}^n$. The matrix E holds the correction $B^+ - B$. The vectors s_i and y_i form the columns of the $n \times m$ matrices S and Y, respectively.

*ENSEEIHT-IRIT, 2, rue Camichel, 31000 Toulouse, France. Email: serge.gratton@enseeiht.fr

†LMS SAMTECH, A Siemens Business, 15-16, Lower Park Row, BS1 5BN Bristol (UK). Email: vincent.malmedy@gmail.com

‡Namur Center for Complex Systems (naXys) and Department of Mathematics, University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium. Email: philippe.toint@unamur.be

As is well-known (see Dennis and Schnabel, 1983, for instance), most secant updates may be derived by a variational approach. To define multisecant versions of the classical formulae, it would therefore make sense to solve the constrained variational problem

$$\min_E \quad \frac{1}{2}\|W^{-T}EW^{-1}\|_F^2 \tag{2.1a}$$

$$\text{s.t.} \quad (B+E)s_i = y_i \text{ for } i = 1,\ldots,m, \quad \text{and } E = E^T, \tag{2.1b}$$

using some particular nonsingular weighting matrix W, whose choice determines the formula that is obtained. For instance, choosing $W = I$ yields the multisecant PSB update, while any matrix W such that $W^TWS = Y$ yields the multisecant DFP update. The multisecant BFGS may then be obtained by duality from its DFP counterpart. This is the approach followed by Schnabel (1983) (see also Byrd et al., 1994) for a multisecant analog of the BFGS formula.

We propose here to consider a slightly different framework : instead of solving (2.1), we aim to solve the penalized problem

$$\min_E \quad \frac{1}{2}\|W^{-T}EW^{-1}\|_F^2 + \frac{1}{2}\sum_{i=1}^m \omega_i \|(B+E)s_i - y_i\|_{\hat{W}^{-1}}^2 \tag{2.2a}$$

$$\text{s.t.} \quad E = E^T, \tag{2.2b}$$

with $\hat{W} = W^TW$, in effect relaxing the secant equations by penalizing their violation with independent nonnegative penalty parameters ω_i (for $i = 1,\ldots,m$).

2.1 Solving the penalized variational problem

Zeroing the derivative of the Lagrangian function of problem (2.2) with respect to E in the (matrix) direction K, we obtain that

$$\mathrm{tr}(W^{-T}EW^{-1}W^{-T}KW^{-1}) + \mathrm{tr}K^T(M - M^T) + \sum_{i=1}^m \omega_i s_i^T K^T \hat{W}^{-1}((B+E)s_i - y_i) = 0, \tag{2.3}$$

where M is the Lagrange multiplier corresponding to the symmetry constraint. Using the vectorization operator $\mathrm{vec}(\cdot)$, this equation may be rewritten as

$$\langle \mathrm{vec}(\hat{W}^{-1}E\hat{W}^{-1} + M - M^T), \mathrm{vec}(K)\rangle + \sum_{i=1}^m \omega_i \langle s_i \otimes \hat{W}^{-1}Es_i - s_i \otimes \hat{W}^{-1}r_i, \mathrm{vec}(K)\rangle = 0, \tag{2.4}$$

where $r_i = y_i - Bs_i$ is the residual on the i-th secant equation at the current Hessian. As the equation (2.4) must hold for every matrix K, we thus have that

$$\mathrm{vec}(\hat{W}^{-1}E\hat{W}^{-1} + M - M^T) + \sum_{i=1}^m \omega_i (s_i \otimes \hat{W}^{-1}Es_i - s_i \otimes \hat{W}^{-1}r_i) = 0,$$

which is equivalent to

$$\hat{W}^{-1}E\hat{W}^{-1} + M - M^T + \sum_{i=1}^m \omega_i (\hat{W}^{-1}Es_is_i^T - \hat{W}^{-1}r_is_i^T) = 0.$$

By summing this equation with its transpose and then, pre- and post-multiplying by $\hat{W}$, we eliminate the Lagrangian multiplier M, and obtain that

$$E\left(I + \sum_{i=1}^m \omega_i s_i s_i^T \hat{W}\right) + \left(I + \sum_{i=1}^m \omega_i \hat{W} s_i s_i^T\right) E = \sum_{i=1}^m \omega_i (\hat{W}s_ir_i^T + r_is_i^T\hat{W}).$$

If we define the $n \times m$ matrices

$$R = Y - BS, \quad \bar{R} = R\,\Omega^{1/2}, \quad \bar{S} = S\,\Omega^{1/2} \text{ and } \bar{Z} = \hat{W}\bar{S},$$

where $\Omega = \mathrm{diag}(\omega_i)$, we thus have to solve the Lyapunov equation

$$AE + EA^T = C, \tag{2.5}$$

where $A = I + \bar{Z}\bar{S}^T$ and $C = \bar{R}\bar{Z}^T + \bar{Z}\bar{R}^T$.

We know from that equation (2.5) has a unique solution whenever A does not have opposite eigenvalues (see p. 414 of Lancaster and Tismenetsky, 1985). But

$$A = I + \bar{Z}\bar{S}^T = I + \hat{W}\bar{S}\bar{S}^T,$$

and from the relation $sp(A) \cup \{1\} = sp(I + \bar{S}^T\hat{W}\bar{S}) \cup \{1\}$, we may deduce that A has only positive eigenvalues. Thus (2.5) has a unique solution. Transposing the left and right hand-side of (2.5), and using the symmetry of C, we furthermore see that E^T also satifies (2.5). Therefore, since the solution of this equation is unique, E is symmetric.

We now show that E can be written in the form

$$E = (\bar{R}, \overline{Z}) \begin{pmatrix} X_1 & X_2 \\ X_2^T & X_3 \end{pmatrix} \begin{pmatrix} \bar{R}^T \\ \bar{Z}^T \end{pmatrix} \tag{2.6}$$

where X_1 and X_3 are symmetric. Substituting this expression into (2.5) and using

$$\bar{R}\bar{Z}^T + \bar{Z}\bar{R}^T = (\bar{R}, \bar{Z}) \begin{pmatrix} 0 & I_m \\ I_m & 0 \end{pmatrix} \begin{pmatrix} \bar{R}^T \\ \bar{Z}^T \end{pmatrix}$$

yields

$$(\bar{R}, \overline{Z}) \left[2 \begin{pmatrix} X_1 & X_2 \\ X_2^T & X_3 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \bar{S}^T\bar{R}X_1 + \bar{S}^T\bar{Z}X_2^T & \bar{S}^T\bar{R}X_2 + \bar{S}^T\bar{Z}X_3 \end{pmatrix} \right.$$
$$\left. + \begin{pmatrix} 0 & X_1\bar{R}^T\bar{S} + X_2\bar{Z}^T\bar{S} \\ 0 & X_2^T\bar{R}^T\bar{S} + X_3\bar{Z}^T\bar{S} \end{pmatrix} - \begin{pmatrix} 0 & I_m \\ I_m & 0 \end{pmatrix} \right] \begin{pmatrix} \bar{R}^T \\ \bar{Z}^T \end{pmatrix} = 0.$$

A solution of (2.5), and thus the unique one, can thus be obtained by setting

$$X_1 = 0, \quad X_2 = \left(2I_m + \bar{Z}^T\bar{S}\right)^{-1} \tag{2.7}$$

and choosing X_3 to be the solution of the Lyapunov equation

$$(I_m + \bar{S}^T\bar{Z})X_3 + X_3(I_m + \bar{Z}^T\bar{S}) = -\bar{S}^T\bar{R}X_2 - X_2\bar{R}^T\bar{S}. \tag{2.8}$$

But the matrix

$$I_m + \bar{S}^T\bar{Z} = I_m + \bar{S}^T\hat{W}\bar{S} \tag{2.9}$$

is symmetric and positive definite. Hence equation (2.8) also has a unique solution, which can directly be computed using the eigen-decomposition of the $m \times m$ symmetric matrix $I_m + \bar{S}^T\bar{Z}$. Therefore, using (2.6), (2.7) and the fact that (2.9) implies the symmetry of X_2, we obtain that the solution of (2.3) can be written as

$$E = \bar{R}\left(2I_m + \bar{Z}^T\bar{S}\right)^{-1}\bar{Z}^T + \bar{Z}\left(2I_m + \bar{Z}^T\bar{S}\right)^{-1}\bar{R} + \bar{Z}X_3\bar{Z}^T, \tag{2.10}$$

where X_3 solves (2.8).

Note that the core of the computation to obtain E (the determination of X_2 and X_3) only involves forming $m \times m$ matrices from $m \times n$ factors and then working in this smaller space, resulting in an $O(m^2n + m^3)$ computational complexity.

An important particular case is the case where we have only one secant equation, i.e. $m = 1$. In this case, the Lyapunov equation (2.8) is a scalar equation, and, defining $r = y - Bs$, $\overline{r} = \sqrt{\omega} r$ and $\overline{s} = \sqrt{\omega} s$, we obtain that

$$\begin{aligned} E &= \frac{\overline{r}\overline{z}^T + \overline{z}\overline{r}^T}{2 + \overline{z}^T\overline{s}} - \frac{\overline{s}^T\overline{r}}{\left(2 + \overline{z}^T\overline{s}\right)\left(1 + \overline{z}^T\overline{s}\right)}\overline{z}\overline{z}^T, \\ &= \frac{rs^T\hat{W} + \hat{W}sr^T}{2\omega^{-1} + s^T\hat{W}s} - \frac{s^Tr}{\left(2\omega^{-1} + s^T\hat{W}s\right)\left(\omega^{-1} + s^T\hat{W}s\right)}\hat{W}ss^T\hat{W} \end{aligned} \tag{2.11}$$

As the (scalar) ω goes to infinity, we get the update

$$E_\infty = \frac{rs^T\hat{W} + \hat{W}sr^T}{s^T\hat{W}s} - \frac{s^Tr}{\left(s^T\hat{W}s\right)^2}\hat{W}ss^T\hat{W}.$$

3 Examples of penalized updates

3.1 The penalized PSB update

If we choose $\hat{W} = I$, we obtain from (2.10) and (2.8) that

$$B^+ = B + R\left(2\Omega^{-1} + S^TS\right)^{-1}S^T + S\left(2\Omega^{-1} + S^TS\right)^{-1}R^T + S\,\Omega^{1/2}X_3\,\Omega^{1/2}S^T, \tag{3.12}$$

where X_3 solves

$$(I_m + \Omega^{1/2}S^TS\,\Omega^{1/2})X_3 + X_3(I_m + \Omega^{1/2}S^TS\,\Omega^{1/2}) = -\,\Omega^{1/2}\,S^TR\,\Omega^{1/2}\,X_2 - X_2\,\Omega^{1/2}\,R^TS\,\Omega^{1/2} \tag{3.13}$$

with $X_2 = \left(2I + \Omega^{1/2}\,S^TS\,\Omega^{1/2}\right)^{-1}$. If we consider the single-secant case ($m = 1$), we derive from (2.11) that

$$B^+ = B + \frac{rs^T + sr^T}{2\omega^{-1} + \|s\|^2} + \left(\frac{1}{\omega^{-1} + \|s\|^2} - \frac{2}{2\omega^{-1} + \|s\|^2}\right)\frac{\langle s, r\rangle}{\|s\|^2}ss^T, \tag{3.14}$$

3.2 The penalized DFP update

If we now assume that the matrices B and Y^TS are symmetric, and that Y^TS is positive definite, we deduce that

$$B^+ = B + R\left(2\Omega^{-1} + Y^TS\right)^{-1}Y^T + Y\left(2\Omega^{-1} + Y^TS\right)^{-1}R^T + Y\,\Omega^{1/2}X_3\,\Omega^{1/2}Y^T, \tag{3.15}$$

where $X_2 = \left(2I + \Omega\,Y^TS\,\Omega\right)^{-1}$ and X_3 solves

$$(X_2^{-1} - I)X_3 + X_3(X_2^{-1} - I) = -\,\Omega^{1/2}\,S^TR\,\Omega^{1/2}\,X_2 - X_2\,\Omega^{1/2}\,R^TS\,\Omega^{1/2}. \tag{3.16}$$

with $X_2 = \left(2I + \Omega^{1/2}\,Y^TS\,\Omega^{1/2}\right)^{-1}$. Note also that a direct computation shows that

$$\begin{aligned} &(X_2^{-1} - I + I)\,X_2\,\Omega^{1/2}\,S^TR\,\Omega^{1/2}\,X_2 + X_2\,\Omega^{1/2}S^TR\,\Omega^{1/2}X_2(X_2^{-1} - I + I) \\ &\qquad = \Omega^{1/2}\,S^TR\,\Omega^{1/2}\,X_2 + X_2\,\Omega^{1/2}\,R^TS\,\Omega^{1/2}. \end{aligned} \tag{3.17}$$

To simplify and better understand the relation between the penalized and a standard DFP formula in block form, we set

$$X_4 = X_3 + X_2\,\Omega^{1/2}S^TR\,\Omega^{1/2}X_2, \tag{3.18}$$

sum up equations (3.17) and (3.16), and obtain that

$$\left(I + \Omega^{1/2}\,Y^TS\,\Omega^{1/2}\right)X_4 + X_4\left(I + \Omega^{1/2}\,Y^TS\,\Omega^{1/2}\right) = -2X_2\,\Omega^{1/2}S^TR\,\Omega^{1/2}X_2. \tag{3.19}$$

We also deduce from (3.18) that

$$\Omega^{1/2} X_3 \Omega^{1/2} = -\left(2\Omega^{-1} + Y^T S\right)^{-1} S^T R \left(2\Omega^{-1} + Y^T S\right)^{-1} + \Omega^{1/2} X_4 \Omega^{1/2},$$

and inserting this expression into the penalized DFP update (3.15), we obtain that

$$\begin{aligned} B^+ &= \left(I - Y\left(2\Omega^{-2} + Y^T S\right)^{-1} S^T\right) B \left(I - Y\left(2\Omega^{-2} + Y^T S\right)^{-1} S^T\right)^T \\ &\quad + Y\left(2\left(2\Omega^{-2} + Y^T S\right)^{-1} - \left(2\Omega^{-2} + Y^T S\right)^{-1} S^T Y \left(2\Omega^{-2} + Y^T S\right)^{-1} + X_5\right) Y^T \end{aligned} \tag{3.20}$$

where, in view of (3.19), $X_5 = \Omega^{1/2} X_4 \Omega^{1/2}$ solves

$$\left(I + \Omega Y^T S\right) X_5 + X_5 \left(I + Y^T S \Omega\right) = -2\left(2\Omega^{-1} + Y^T S\right)^{-1} S^T R \left(2\Omega^{-1} + Y^T S\right)^{-1}. \tag{3.21}$$

But, using $S^T Y = Y^T S$,

$$\begin{aligned} 2\left(2\Omega^{-1} + Y^T S\right)^{-1} &- \left(2\Omega^{-1} + Y^T S\right)^{-1} S^T Y \left(2\Omega^{-1} + Y^T S\right)^{-1} \\ &= \left(2\Omega^{-1} + Y^T S\right)^{-1} \left[2\left(2\Omega^{-1} + Y^T S\right)\left(2\Omega^{-1} + Y^T S\right)^{-1} - S^T Y \left(2\Omega^{-1} + Y^T S\right)^{-1}\right] \\ &= \left(2\Omega^{-1} + Y^T S\right)^{-1} \left(4\Omega^{-1} + Y^T S\right)\left(2\Omega^{-1} + Y^T S\right)^{-1}, \end{aligned}$$

and hence (3.20) finally becomes

$$\begin{aligned} B^+ &= \left(I - Y\left(2\Omega^{-1} + Y^T S\right)^{-1} S^T\right) B \left(I - Y\left(2\Omega^{-1} + Y^T S\right)^{-1} S^T\right)^T \\ &\quad + Y\left(\left(2\Omega^{-1} + Y^T S\right)^{-1} \left(4\Omega^{-1} + Y^T S\right)\left(2\Omega^{-1} + Y^T S\right)^{-1} + X_5\right) Y^T \end{aligned} \tag{3.22}$$

where X_5 solves (3.21). If we set $\Omega = \omega I$ and assume that $Y^T S$ is non singular, we obtain from (3.21) that $X_5 = O(\omega^{-1})$, and we see from (3.22) that the penalized update converges to the DFP update in block form when ω goes to infinity.

If $m = 1$, (2.11) gives that

$$B^+ = B + \frac{r y^T + y r^T}{2\omega^{-1} + \langle s, y\rangle} + \left(\frac{1}{\omega^{-1} + \langle s, y\rangle} - \frac{2}{2\omega^{-1} + \langle s, y\rangle}\right) \frac{\langle s, r\rangle}{\langle s, y\rangle} y y^T. \tag{3.23}$$

3.3 The penalized BFGS update

Let us assume again that $Y^T S$ is symmetric positive definite and set $P = S - HY$, with H being an approximation of the inverse Hessian. Exchanging the role of Y and S in the penalized DFP formula, we derive the BFGS penalized update defined by

$$\begin{aligned} H^+ &= \left(I - S\left(2\Omega^{-1} + Y^T S\right)^{-1} Y^T\right) H \left(I - S\left(2\Omega^{-1} + Y^T S\right)^{-1} Y^T\right)^T \\ &\quad + S\left(\left(2\Omega^{-1} + Y^T S\right)^{-1} \left(4\Omega^{-2} + Y^T S\right)\left(2\Omega^{-1} + Y^T S\right)^{-1} + X_6\right) S^T, \end{aligned} \tag{3.24}$$

where X_6 solves

$$\left(I + \Omega Y^T S\right) X_6 + X_6 \left(I + Y^T S \Omega\right) = -2\left(2\Omega^{-1} + Y^T S\right)^{-1} S^T P \left(2\Omega^{-1} + Y^T S\right)^{-1}. \tag{3.25}$$

If we set $\Omega = \omega I$, and consider large values for ω we see that the penalized update again converges to the BFGS formula. We also conclude that for ω sufficiently large, the penalized update will be positive definite. If $Y^T S$ is not symmetric and positive definite, an approximate update could be computed by replacing the matrix $\left(\left(2\Omega^{-1} + Y^T S\right)^{-1} \left(4\Omega^{-1} + Y^T S\right)\left(2\Omega^{-1} + Y^T S\right)^{-1} + X_6\right)$ between S and S^T in expression (3.24) by any symmetric and positive definite approximation.

In the case of a single secant pair ($m = 1$), we obtain that

$$H^+ = H + \frac{ps^T + sp^T}{2\omega^{-1} + \langle s, y\rangle} + \left(\frac{1}{\omega^{-1} + \langle s, y\rangle} - \frac{2}{2\omega^{-1} + \langle s, y\rangle}\right)\frac{\langle p, y\rangle}{\langle s, y\rangle} ss^T \tag{3.26}$$

$$= \left(I - \frac{\rho sy^T}{1 + 2\rho\omega^{-1}}\right) H \left(I - \frac{\rho ys^T}{1 + 2\rho\omega^{-1}}\right) + \theta\rho ss^T, \tag{3.27}$$

with $\rho = \langle s, y\rangle^{-1}$, $p = s - Hy$ and

$$\theta = \left(1 + \tfrac{\rho}{\omega}\right)^{-1} - \rho\langle y, Hy\rangle \left[\left(1 + \tfrac{2\rho}{\omega}\right)^{-2} + \left(1 + \tfrac{\rho}{\omega}\right)^{-1} - \left(\tfrac{1}{2} + \tfrac{\rho}{\omega}\right)^{-1}\right].$$

4 Conclusions

We have used variational techniques to derive a new algorithm for updating a quasi-Newton Hessian approximation using multiple secant equations. The cost of the associated updates has been discussed and remains reasonable in the sense that it does not involve terms in the cube of the problem dimension. It is hoped that the new algorithm may prove useful beyond its initial application in Malmedy (2010).

References

C. G. Broyden. The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications*, **6**, 76–90, 1970.

R. H. Byrd, J. Nocedal, and R. B. Schnabel. Representation of quasi-newton matrices and their use in limited memory methods. *Mathematical Programming*, **63**, 129–156, 1994.

W. C. Davidon. Variable metric method for minimization. Report ANL-5990(Rev.), Argonne National Laboratory, Research and Development, 1959. Republished in the SIAM Journal on Optimization, vol. 1, pp. 1–17, 1991.

J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1983. Reprinted as *Classics in Applied Mathematics 16*, SIAM, Philadelphia, USA, 1996.

R. Fletcher. A new approach to variable metric algorithms. *Computer Journal*, **13**, 317–322, 1970.

R. Fletcher and M. J. D. Powell. A rapidly convergent descent method for minimization. *Computer Journal*, **6**, 163–168, 1963.

D. Goldfarb. A family of variable metric methods derived by variational means. *Mathematics of Computation*, **24**, 23–26, 1970.

P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Academic Press, London, second edn, 1985.

V. Malmedy. *Hessian approximation in multilevel nonlinear optimization*. PhD thesis, Department of Mathematics, University of Namur, Namur, Belgium, 2010.

J. Nocedal. private communication, Toulouse, 2013.

M. J. D. Powell. A new algorithm for unconstrained optimization. *in* J. B. Rosen, O. L. Mangasarian and K. Ritter, eds, 'Nonlinear Programming', pp. 31–65, London, 1970. Academic Press.

R. B. Schnabel. Quasi-newton methods using multiple secant equations. Technical Report CU-CS-247-83, Departement of Computer Science, University of Colorado at Boulder, Boulder, USA, 1983.

D. F. Shanno. Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, **24**, 647–657, 1970.