ELSEVIER

Journal of Computational and Applied Mathematics 124 (2000) 61–95

JOURNAL OF
COMPUTATIONAL AND
APPLIED MATHEMATICS

www.elsevier.nl/locate/cam

Variable metric methods for unconstrained optimization and nonlinear least squares ☆

Ladislav Lukšan[a], Emilio Spedicato[b,*]

[a] *Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2,
182 07 Prague 8, Czech Republic*

[b] *Department of Mathematics, University of Bergamo, 24129 Bergamo, Italy*

Received 20 May 1999; received in revised form 13 November 1999

Abstract

Variable metric or quasi-Newton methods are well known and commonly used in connection with unconstrained optimization, since they have good theoretical and practical convergence properties. Although these methods were originally developed for small- and moderate-size dense problems, their modifications based either on sparse, partitioned or limited-memory updates are very efficient on large-scale sparse problems. Very significant applications of these methods also appear in nonlinear least-squares approximation and nonsmooth optimization. In this contribution, we give an extensive review of variable metric methods and their use in various optimization fields. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Quasi-Newton methods; Variable metric methods; Unconstrained optimization; Nonlinear least squares; Sparse problems; Partially separable problems; Limited-memory methods

1. Introduction

This paper reviews the efficient class of methods known as variable metric methods or quasi-Newton methods for local unconstrained minimization, i.e., for finding a point $x_* \in \mathbb{R}^n$ such that $F(x_*) = \min_{x \in \mathbb{R}^n} F(x)$ (we consider only local minima). Here $F : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable objective function and $\mathbb{R}^n$ is an n-dimensional vector space.

☆ Work partly supported by grant GA ČR No. 201/00/0080, by CNR-GNIM visiting professorship program and by MURST Programma di Cofinanziamento.

[*] Corresponding author.

E-mail addresses: luksan@uivt.cas.cz (L. Lukšan), emilio@unibg.it (E. Spedicato).

PII: S0377-0427(00)00420-9

Methods for unconstrained minimization are iterative. Starting with an initial point $x_1 \in \mathbb{R}^n$, they generate a sequence $x_i \in \mathbb{R}^n$, $i \in \mathcal{N}$, by the simple process

$$x_{i+1} = x_i + \alpha_i d_i, \tag{1.1}$$

where $d_i \in \mathbb{R}^n$ is a direction vector and $\alpha_i \geqslant 0$ is a scalar, the stepsize ($\mathcal{N}$ is the set of natural numbers). The most efficient optimization methods belong to three classes: the modified Newton, variable metric and conjugate gradient methods. We mention basic properties of these classes here in order to clarify the application of variable metric methods in particular cases.

Modified Newton methods are based on a local quadratic model

$$Q_i(d) = \tfrac{1}{2} d^{\mathrm{T}} G_i d + g_i^{\mathrm{T}} d, \tag{1.2}$$

where $G_i = G(x_i)$ and $g_i = g(x_i)$ are, respectively, the Hessian matrix and the gradient of the objective function $F : \mathbb{R}^N \to \mathbb{R}$ at the point $x_i \in \mathbb{R}^n$. The direction vector $d_i \in \mathbb{R}^n$, $i \in \mathcal{N}$, is chosen to minimize $Q_i(d)$ (approximately) on $\mathbb{R}^n$ or on some subset of $\mathbb{R}^n$. Modified Newton methods converge fast, if they converge, but they have some disadvantages. Minimization of $Q_i(d)$ requires $\mathrm{O}(n^3)$ operations and computation of second-order derivatives can be difficult and time consuming. Moreover, if the Hessian matrices are not positive definite, then simple implementations of modified Newton methods need not be globally convergent. Nevertheless, modified Newton methods can be very efficient for large-scale problems. If $Q_i(d)$ is minimized iteratively, then the matrix–vector products involving G_i can be replaced by numerical differentiation. This leads to truncated Newton methods which do not require computation of second-order derivatives. Moreover, if G_i is sparse, then we need substantially less than $\mathrm{O}(n^3)$ operations for minimization of $Q_i(d)$.

Variable metric methods are based on the local quadratic model

$$Q_i(d) = \tfrac{1}{2} d^{\mathrm{T}} B_i d + g_i^{\mathrm{T}} d, \tag{1.3}$$

where B_i is some positive-definite approximation of G_i. Matrices B_i, $i \in \mathcal{N}$, are constructed iteratively so that B_1 is an arbitrary positive-definite matrix and B_{i+1} is determined from B_i in such a way that it is positive definite, is as close as possible to B_i and satisfies the quasi-Newton condition

$$B_{i+1} s_i = y_i,$$

where $s_i = x_{i+1} - x_i$ and $y_i = g_{i+1} - g_i$. The BFGS formula

$$B_{i+1} = B_i + \frac{y_i y_i^{\mathrm{T}}}{y_i^{\mathrm{T}} s_i} - \frac{B_i s_i (B_i s_i)^{\mathrm{T}}}{s_i^{\mathrm{T}} B_i s_i}$$

is widely used (cf. (2.13) and (2.17)). Variable metric methods have some advantages over modified Newton methods. The matrices B_i are positive definite and so variable metric methods can be forced to be globally convergent. Moreover, we can update the inverse $H_i = B_i^{-1}$ or the Cholesky decomposition $L_i D_i L_i^{\mathrm{T}} = B_i$, instead of B_i itself, using only $\mathrm{O}(n^2)$ operations per iteration. Even when variable metric methods require more iterations than modified Newton methods, they are usually more efficient for small- and moderate-size dense problems.

Conjugate gradient methods, see [51,37,73], use only n-dimensional vectors. Direction vectors $d_i \in \mathbb{R}^n$, $i \in \mathcal{N}$ are generated so that $d_1 = -g_1$ and

$$d_{i+1} = -g_{i+1} + \beta_i d_i, \tag{1.4}$$

where $g_{i+1} = g(x_{i+1})$ is the gradient of the objective function $F : \mathbb{R}^N \to \mathbb{R}$ at the point x_{i+1} and β_i is a suitably defined scalar parameter. Conjugate gradient methods require only $\mathrm{O}(n)$ storage elements and O(n) operations per iteration, but they use more iterations than variable metric methods. Of course, these iterations are less expensive. Conjugate gradient methods are intended for large-scale problems.

In this paper, we review variable metric methods for basic unconstrained optimization problems. Our approach is mainly devoted to the computational aspects, i.e., to the derivation of efficient methods and their implementation; therefore, while we quote a number of fundamental convergence results in the field, the difficult and partly still open field of analysis of convergence is not dealt with at great length. Section 2 is devoted to variable metric methods for dense (small- and moderate-size) problems. In Section 3, we describe various modifications of variable metric methods for large-scale problems. Section 4 concerns the use of variable metric updates for improving the efficiency of methods for nonlinear least squares.

In this paper, properties of variable metric methods are sometimes demonstrated by computational experiments. For this purpose, we used FORTRAN codes TEST14 (22 test problems for general unconstrained optimization), TEST15 (22 test problems for nonlinear least squares) and TEST18 (30 test problems for systems of nonlinear equations) which are described in [62] and can be downloaded from the web homepage `http://www.uivt.cas.cz/~luksan#software`. Computational experiments were realized by using the optimization system UFO [61] (see also the above web homepage).

Optimization methods can be realized in various ways which differ in direction determination and stepsize selection. Line-search and trust-region realizations are the most popular, especially for variable metric methods. A basic framework for these methods is given in the following subsection. (Readers already familiar with this material may wish to skip it.)

1.1. Line-search methods

Line-search methods require the vectors $d_i \in \mathbb{R}^n$, $i \in \mathcal{N}$, to be descent directions, i.e.,

$$c_i \triangleq -g_i^{\mathrm{T}} d_i / \|g_i\| \|d_i\| > 0. \tag{1.5}$$

Then the stepsizes α_i, $i \in \mathcal{N}$, can be chosen in such a way that $\alpha_i > 0$ and

$$F_{i+1} - F_i \leqslant \varepsilon_1 \alpha_i g_i^{\mathrm{T}} d_i, \tag{1.6}$$

$$g_{i+1}^{\mathrm{T}} d_i \geqslant \varepsilon_2 g_i^{\mathrm{T}} d_i, \tag{1.7}$$

where $0 < \varepsilon_1 < \frac{1}{2}$ and $\varepsilon_1 < \varepsilon_2 < 1$ (here $F_{i+1} = F(x_{i+1})$, $g_{i+1} = g(x_{i+1})$, where x_{i+1} is defined by (1.1)). The following theorem, see [32], characterizes the global convergence of line-search methods.

Theorem 1.1. *Let the objective function* $F : \mathbb{R}^N \to \mathbb{R}$ *be bounded from below and have bounded second-order derivatives. Consider the line-search method* (1.1) *with* d_i *and* α_i *satisfying* (1.5)–(1.7). *If*

$$\sum_{i \in \mathcal{N}} c_i^2 = \infty, \tag{1.8}$$

then $\liminf_{i \to \infty} \|g_i\| = 0$.

If d_i is determined by minimizing (1.3), i.e., $d_i = B_i^{-1} g_i$ with B_i positive definite, then (1.8) can be replaced by

$$\sum_{i \in \mathcal{N}} \frac{1}{\kappa_i} = \infty, \tag{1.9}$$

where $\kappa_i = \kappa(B_i)$ is the spectral condition number of the matrix B_i. Note that (1.8) (or (1.9)) is satisfied if a constant $\underline{c} > 0$ (or $\bar{c} > 0$) and an infinite set $\mathcal{M} \subset \mathcal{N}$ exist so that $c_i \geqslant \underline{c}$ (or $\kappa_i < \bar{c}$) $\forall i \in \mathcal{M}$.

Variable metric methods in a line-search realization require the direction vectors to satisfy condition (1.5) and $\|B_i d_i + g_i\| \leqslant \omega_i \|g_i\|$, where $0 \leqslant \omega_i \leqslant \bar{\omega} < 1$ is a prescribed precision (the additional condition $\omega_i \to 0$ is required for obtaining a superlinear rate of convergence). Such vectors can be obtained in two basic ways. If the original problem is of small or moderate size or if it has a suitable sparsity pattern, we can set

$$d_i = -H_i g_i, \tag{1.10}$$

where $H_i = B_i^{-1}$, or use back substitution to solve

$$L_i D_i L_i^{\mathrm{T}} d_i = -g_i \tag{1.11}$$

after Cholesky decomposition of B_i. Otherwise, an iterative method may be preferable. The preconditioned conjugate gradient method is especially suitable. It starts with the vectors $s_1 = 0$, $r_1 = -g_i$, $p_1 = C_i^{-1} r_1$ and uses the recurrence relations

$$\begin{aligned}
q_j &= B_i p_j, \\
\alpha_j &= r_j^{\mathrm{T}} C_i^{-1} r_j / p_j^{\mathrm{T}} q_j, \\
s_{j+1} &= s_j + \alpha_j p_j, \\
r_{j+1} &= r_j - \alpha_j q_j, \\
\beta_j &= r_{j+1}^{\mathrm{T}} C_i^{-1} r_{j+1} / r_j^{\mathrm{T}} C_i^{-1} r_j, \\
p_{j+1} &= C_i^{-1} r_{j+1} + \beta_j p_j
\end{aligned} \tag{1.12}$$

for $j \in \mathcal{N}$. This process is terminated if either $\|r_j\| \leqslant \omega_i \|g_i\|$ (sufficient precision) or $p_j^{\mathrm{T}} q_j \leqslant 0$ (non-positive curvature). In both cases we set $d_i = s_j$. The matrix C_i is a preconditioner which should be chosen to make $B_i C_i$ as well conditioned as possible. Very efficient preconditioners can be based on incomplete Cholesky decomposition, see [5].

1.2. Trust-region methods

Trust-region methods use direction vectors $d_i \in \mathbb{R}^n$, $i \in \mathcal{N}$, which satisfy

$$\|d_i\| \leqslant \Delta_i, \tag{1.13}$$

$$\|d_i\| < \Delta_i \quad \Rightarrow \quad \|B_i d_i + g_i\| \leqslant \omega_i \|g_i\|, \tag{1.14}$$

$$-Q_i(d_i) \geqslant \underline{\sigma} \|g_i\| \min(\|d_i\|, \|g_i\|/\|B_i\|), \tag{1.15}$$

where $0 \leqslant \omega_i \leqslant \bar{\omega} < 1$ and $0 < \underline{\sigma} < 1$ (we consider spectral norms here, but $\|d_i\|$ can be an arbitrary norm). Steplengths $\alpha_i \geqslant 0$, $i \in \mathcal{N}$, in (1.1)) are chosen so that

$$\rho_i(d_i) \leqslant 0 \quad \Rightarrow \quad \alpha_i = 0, \tag{1.16}$$

$$\rho_i(d_i) > 0 \quad \Rightarrow \quad \alpha_i = 1, \tag{1.17}$$

where $\rho_i(d_i) = (F(x_i + d_i) - F(x_i))/Q_i(d_i)$. Trust-region radii $0 < \Delta_i \leqslant \bar{\Delta}$, $i \in \mathcal{N}$, are chosen so that $0 < \Delta_1 \leqslant \bar{\Delta}$ is arbitrary and

$$\rho_i(d_i) < \underline{\rho} \quad \Rightarrow \quad \underline{\beta}\|d_i\| \leqslant \Delta_{i+1} \leqslant \bar{\beta}\|d_i\|, \tag{1.18}$$

$$\rho_i(d_i) \geqslant \underline{\rho} \quad \Rightarrow \quad \Delta_i \leqslant \Delta_{i+1} \leqslant \bar{\Delta}, \tag{1.19}$$

where $0 < \underline{\beta} \leqslant \bar{\beta} < 1$ and $0 < \underline{\rho} < 1$. The following theorem, see [75], characterizes the global convergence of trust-region methods.

Theorem 1.2. *Let the objective function $F : \mathbb{R}^N \to \mathbb{R}$ be bounded from below and have bounded second-order derivatives. Consider the trust-region method* (1.13)–(1.19) *and denote $M_i = \max(\|B_1\|, \ldots, \|B_i\|)$, $i \in \mathcal{N}$. If*

$$\sum_{i \in \mathcal{N}} \frac{1}{M_i} = \infty, \tag{1.20}$$

then $\liminf_{i\to\infty} \|g_i\| = 0$.

Note that (1.20) is satisfied if a constant $\bar{B}$ and an infinite set $\mathcal{M} \subset \mathcal{N}$ exist, so that $\|B_i\| \leqslant \bar{B}$, $\forall i \in \mathcal{M}$.

Trust-region methods require the direction vectors to satisfy conditions (1.13)–(1.15). Such vectors can be obtained in three basic ways. The most sophisticated way consists in solving the constrained minimization subproblem

$$d_i = \mathop{\arg\min}_{\|d\| \leqslant \Delta_i} Q_i(d), \tag{1.21}$$

where $Q_i(d)$ is given by (1.2) or (1.3). This approach, which leads to the repeated solution of the equation $(B_i + \lambda I)d_i(\lambda) + g_i = 0$ for selected values of λ, see [66], is time consuming since it requires, on average, 2 or 3 Cholesky decompositions per iteration. Moreover, an additional matrix has to be used. Therefore, easier approaches have been looked for.

One such approach consists in replacing (1.21) by the two-dimensional subproblem

$$d_i = \mathop{\arg\min}_{\|d(\alpha,\beta)\| \leqslant \Delta_i} Q_i(d(\alpha, \beta)), \tag{1.22}$$

where $d(\alpha, \beta) = \alpha g_i + \beta B_i^{-1} g_i$. Subproblem (1.22) is usually solved approximately by the so-called dog-leg methods [25,74].

If the original problem is large then the inexact trust-region method, [90], can be used. This method is based on the fact that the vectors s_j, $j \in \mathcal{N}$, determined by the preconditioned conjugate gradient method (1.12), satisfy the recurrence inequalities

$$s_{j+1}^{\mathrm{T}} C s_{j+1}^{\mathrm{T}} > s_j^{\mathrm{T}} C s_j,$$

$$Q(s_{j+1}) < Q(s_j),$$

where Q is the quadratic function (1.2) or (1.3). Thus a suitable path is generated in the trust region. If $\|s_j\| \leqslant \Delta_i$ and $\|r_j\| \leqslant \omega_i \|g_i\|$, then we set $d_i = s_j$. If $\|s_j\| \leqslant \Delta_i$ and $p_j^{\mathrm{T}} q_j \leqslant 0$, then we set $d_i = s_j + \lambda_j p_j$, where λ_j is chosen in such a way that $\|d_i\| = \Delta$. If $\|d_j\| \leqslant \Delta$ and $\|d_{j+1}\| > \Delta$, then we set $d = d_j + \lambda_j(d_{j+1} - d_j)$, where λ_j is chosen in such a way that $d = \Delta$. Otherwise we continue the conjugate gradient process.

2. Variable metric methods for dense problems

2.1. Derivation of variable metric methods

Variable metric methods were originally developed for general unconstrained minimization of objective functions with dense Hessian matrices. As mentioned above, these methods use positive-definite matrices B_i, $i \in \mathcal{N}$, which are generally constructed iteratively using a least-change update satisfying the quasi-Newton condition $B_i s_i = y_i$, where $s_i = x_{i+1} - x_i$ and $y_i = g_{i+1} - g_i$. This condition is fulfilled by the matrix

$$\tilde{G}_i = \int_0^1 G(x_i + ts_i)\,\mathrm{d}t \tag{2.1}$$

which can be considered as a good approximation of the matrix $G_{i+1} = G(x_{i+1})$. Roughly speaking, the least-change principle guarantees that as much information from previous iterations as possible is saved while the quasi-Newton condition brings new information because it is satisfied by matrix (2.1). Notice that there are many least-change principles based on various potential functions and also that it is not necessary to satisfy the quasi-Newton equation accurately (see Theorem 3.1 and [98]).

More sophisticated quasi-Newton conditions are sometimes exploited, based on the fact that the matrix $G(x_{i+1})$ satisfies the condition

$$G(x_{i+1}) \frac{\mathrm{d}x(t)}{\mathrm{d}t}\bigg|_{t=1} = \frac{\mathrm{d}g(t)}{\mathrm{d}t}\bigg|_{t=1}, \tag{2.2}$$

where $x(t)$ is a smooth curve such that $x(0) = x_i$ and $x(1) = x_{i+1}$, say, and $g(t) = g(x(t))$. Starting from (2.2), Ford and Moghrabi [40] used a polynomial curve $x(t)$ interpolating the most recent iterates together with the gradient curve $g(t)$ determined by using the same interpolation coefficients. In the quadratic case when $x(t_{i-1}) = x_{i-1}$, $x(0) = x_i$ and $x(1) = x_{i+1}$, this approach gives the quasi-Newton equation

$$B_{i+1}\left(s_i + \frac{1}{t_{i-1}(t_{i-1} - 2)} s_{i-1}\right) = y_i + \frac{1}{t_{i-1}(t_{i-1} - 2)} y_{i-1},$$

where $s_{i-1} = x_i - x_{i-1}$ and $y_{i-1} = g_i - g_{i-1}$. The efficiency of this approach strongly depends on the value $t_{i-1} < 0$. Some ways of choosing this value are described in [39,41].

Another approach based on (2.2) was used in [99]. In this case, $x(t) = x_i + ts_i$ and $g(t)$ is a quadratic polynomial interpolating $g(0) = g_i$, $g(1) = g_{i+1}$ and satisfying the condition

$$F_{i+1} - F_i = \int_0^1 s_i^{\mathrm{T}} g(t)\,\mathrm{d}t.$$

This approach leads to the quasi-Newton equation

$$B_{i+1}s_i = y_i + \gamma_i \frac{s_i}{\|s_i\|},$$

where $\gamma_i = 3(g_{i+1} + g_i)^{\mathrm{T}} s_i - 6(F_{i+1} - F_i)$.

The simplest way to incorporate function values into the quasi-Newton equation, known as the nonquadratic correction, was introduced in [7]. Consider the function $\phi(t) = F(x_i + ts)$. Using the backward Taylor expansion, we can write $\phi(0) = \phi(1) - \phi'(1) + (1/2)\phi''(\tilde{t})$, where $0 \leqslant \tilde{t} \leqslant 1$. On the other hand, if we write the quasi-Newton condition as

$$B_{i+1}s_i = \frac{1}{\rho_i} y_i, \tag{2.3}$$

then $s_i^{\mathrm{T}} B_{i+1} s_i = s_i^{\mathrm{T}} y_i/\rho_i$. Approximating $s_i^{\mathrm{T}} B_{i+1} s_i$ by $\phi''(\tilde{t})$ obtained from the backward Taylor expansion, we get

$$\rho_i = \frac{s_i^{\mathrm{T}} y_i}{2(F_i - F_{i+1} + s_i^{\mathrm{T}} g_{i+1})}. \tag{2.4}$$

Formula (2.4) was derived in [84]. Similar formulas are also proposed in [7,8]. Alternatively instead of matrices B_i, $i \in \mathcal{N}$, we can construct matrices $H_i = B_i^{-1}$, since the equation $B_i d_i = -g_i$ can easily be solved in this case by setting

$$d_i = -H_i g_i \tag{2.5}$$

To simplify the notation, we now omit the index i and replace the index $i+1$ by $+$ so that (2.3) can be rewritten in the form

$$H_+ y = \rho s. \tag{2.6}$$

Moreover, we define the scalars a, b, c by

$$a = y^{\mathrm{T}} H y, \quad b = y^{\mathrm{T}} s, \quad c = s^{\mathrm{T}} H^{-1} s. \tag{2.7}$$

In what follows, we will take the nonquadratic correction (2.6) into account, together with a suitable scaling.

Scaling of the matrix H was first introduced in [69]. A simple heuristic idea for scaling is the replacement of H by γH before updating to make the difference $H_+ - \gamma H$ as small as possible. One possibility is to derive γ from (2.6) after premultiplying it by a vector and replacing H_+ by γH. Using the vector y, we obtain

$$\gamma/\rho = b/a. \tag{2.8}$$

Similarly, using the vector $H^{-1}s$, we obtain

$$\gamma/\rho = c/b. \tag{2.9}$$

Another useful value is the geometric mean

$$\gamma/\rho = \sqrt{c/a}. \tag{2.10}$$

It is interesting that these simple values often considerably improve the efficiency of variable metric methods, while more sophisticated formulae, derived by minimization of certain potential functions, usually give worse results, see [57]. Scaling applied in every iteration is inefficient in general, see [78], but can be very useful on very difficult functions, see [81]. Therefore, some selective scaling strategies have been developed. The simplest possibility, scaling only in the first iteration (or preliminary scaling, PS), is proposed in [78]. In [18], it is recommended to use the scaling parameter $\gamma = \max(1, \min(\tilde{\gamma}, \bar{\gamma}))$ in every iteration, where $\tilde{\gamma}$ is a theoretically computed value (e.g. (2.8)–(2.10)) and $\bar{\gamma}$ is a suitable upper bound. This choice follows from the fact that global convergence can be proved in this case (cf. Theorem 2.2). A slightly modified strategy, interval scaling IS, is proposed in [58]. Here the value $\gamma = \tilde{\gamma}$ is used, if $\underline{\gamma} \leqslant \tilde{\gamma} \leqslant \bar{\gamma}$. Otherwise we set $\gamma = 1$. Recommended values $0 < \underline{\gamma} < 1 < \bar{\gamma}$, corresponding to individual formulae (2.8)–(2.10), are also given in [58].

Now, we are in a position to derive a class of scaled variable metric methods satisfying the generalized quasi-Newton condition (2.6). Our problem can be formulated as finding a symmetric least-change update $\Delta H = H_+ - \gamma H$, satisfying the condition $\Delta H y = \rho s - \gamma H y$. We can intuitively suppose that the rank of this update should be as small as possible. Since two vectors s and Hy appear in the generalized quasi-Newton condition (2.6), we restrict our attention to rank two updates of the form $\Delta H = \gamma U M U^{\mathrm{T}}$, where $U = [s, Hy]$ and M is a symmetric 2×2 matrix. Substituting this expression into the quasi-Newton condition and comparing the coefficients, we obtain, with η a free parameter

$$\frac{1}{\gamma} H_+ = H + \frac{\rho}{\gamma} \frac{1}{b} s s^{\mathrm{T}} - \frac{1}{a} Hy (Hy)^{\mathrm{T}} + \frac{\eta}{a} \left(\frac{a}{b} s - Hy \right) \left(\frac{a}{b} s - Hy \right)^{\mathrm{T}}. \tag{2.11}$$

Formula (2.11) defines a three-parameter class, the so-called Huang–Oren class of variable metric updates, see [53,69,84]. If we assume ρ and γ to be fixed or computed by (2.4) and (2.8)–(2.10), we get a one-parameter class, the so-called scaled Broyden class (the original Broyden class corresponds to the values $\rho = 1$ and $\gamma = 1$). Three classic values of the parameter η are very popular. Setting $\eta = 0$, we get the scaled DFP [19,36] update

$$\frac{1}{\gamma} H_+ = H + \frac{\rho}{\gamma} \frac{1}{b} s s^{\mathrm{T}} - \frac{1}{a} Hy (Hy)^{\mathrm{T}}. \tag{2.12}$$

Setting $\eta = 1$, we get the scaled BFGS [11,31,46,77] update

$$\frac{1}{\gamma} H_+ = H + \left(\frac{\rho}{\gamma} + \frac{a}{b} \right) \frac{1}{b} s s^{\mathrm{T}} - \frac{1}{b} (Hy s^{\mathrm{T}} + s (Hy)^{\mathrm{T}}). \tag{2.13}$$

Setting $\eta = (\rho/\gamma)/(\rho/\gamma - a/b)$, we get the scaled symmetric rank-one (SR1) update

$$\frac{1}{\gamma} H_+ = H + \left(\frac{\rho}{\gamma} - \frac{a}{b} \right)^{-1} \frac{1}{b} \left(\frac{\rho}{\gamma} s - Hy \right) \left(\frac{\rho}{\gamma} s - Hy \right)^{\mathrm{T}}. \tag{2.14}$$

Formula (2.11) gives another idea for scaling. It can be proved, see [69], that if $0 \leqslant \eta \leqslant 1$ and $b/c \leqslant \rho/\gamma \leqslant a/b$, then $\kappa(\tilde{G} H_+) \leqslant \kappa(\tilde{G} H)$, where $\tilde{G}$ is the matrix defined by (2.1) (κ denotes the spectral condition number). It is clear that for (2.8)–(2.10) the inequality $b/c \leqslant \rho/\gamma \leqslant a/b$ holds ($b/c \leqslant a/b$

follows from the Schwartz inequality). A more sophisticated reason for scaling, based on optimal conditioning of the matrix $H^{-1}H_+$, will be mentioned later (see (2.24)).

Writing $\Delta B = B_+ - (1/\gamma)B$, we can write (2.6) in the form $\Delta Bs = (1/\rho)y - (1/\gamma)Bs$. Proceeding as above, we obtain

$$\gamma B_+ = B + \frac{\gamma}{\rho}\frac{1}{b}yy^{\mathrm{T}} - \frac{1}{c}Bs(Bs)^{\mathrm{T}} + \frac{\beta}{c}\left(\frac{c}{b}y - Bs\right)\left(\frac{c}{b}y - Bs\right)^{\mathrm{T}}, \tag{2.15}$$

see (2.11), if we replace H, s, y, η, ρ, γ by B, y, s, β, $1/\rho$, $1/\gamma$, respectively. Using the Woodbury formula, we can prove that $B = H^{-1}$ implies $B_+ = H_+^{-1}$ if and only if the parameters η and β are related by the following *duality* relation:

$$\beta\eta(ac - b^2) + (\beta + \eta)b^2 = b^2. \tag{2.16}$$

For example, setting $\beta = 0$, we get the scaled BFGS update

$$\gamma B_+ = B + \frac{\gamma}{\rho}\frac{1}{b}yy^{\mathrm{T}} - \frac{1}{c}Bs(Bs)^{\mathrm{T}}. \tag{2.17}$$

Variable metric methods for general unconstrained problems are usually realized in the form (2.11), but form (2.15) is also possible. In the second case, the Cholesky decomposition LDL^{T} of the matrix B is updated using $\mathrm{O}(n^2)$ operations by the numerically stable method described in [45]. This possibility is very attractive, since positive definiteness can be controlled. However, numerical experiments indicate that the form (2.11) is more efficient, measured by computational time, since cheaper operations are used and stability is not lost. Nevertheless, form (2.15) is the only possible one for sparse problems and for improving the Gauss–Newton method for nonlinear least squares.

2.2. Theoretical properties of variable metric methods

From now on we shall assume that the vectors s and Hy are linearly independent. Otherwise, the generalized quasi-Newton condition (2.6) can be fulfilled by simple scaling. Assuming γ and ρ to be fixed, we have one degree of freedom in the choice of the parameter η (or β). We introduce the critical values

$$\eta^{\mathrm{c}} = \beta^{\mathrm{c}} = \frac{b^2}{b^2 - ac} < 0. \tag{2.18}$$

We can then deduce from (2.16) that $\eta < \eta^{\mathrm{c}}$, $\eta^{\mathrm{c}} < \eta < 0$, $0 \leqslant \eta \leqslant 1$, $1 < \eta$, if and only if $\beta < \beta^{\mathrm{c}}$, $1 < \beta$, $0 \leqslant \beta \leqslant 1$, $\beta^{\mathrm{c}} < \beta < 0$, respectively. Moreover, one can prove, see [80], that the matrix H_+ (or B_+) is positive definite if and only if $b > 0$ and $\eta > \eta^{\mathrm{c}}$ (or $\beta > \beta^{\mathrm{c}}$). Value (2.18) is negative by the Schwartz inequality, since H is assumed to be positive definite and the vectors s and Hy are assumed to be linearly independent. The interval given by $0 \leqslant \eta \leqslant 1$ (or $0 \leqslant \beta \leqslant 1$) defines the so-called *restricted Broyden subclass*, whose updates can be written as convex combinations of the DFP and the BFGS update.

First, we introduce some basic results concerning the scaled Broyden class of variable metric methods. We begin with the quadratic termination property, see [11].

Theorem 2.1. *Let the objective function* $F : \mathbb{R}^N \to \mathbb{R}$ *be quadratic with positive-definite Hessian matrix* G. *Consider the variable metric method* (1.1) *with stepsizes chosen so that* $g_{i+1}^{\mathrm{T}}d_i = 0$

(*perfect line search*) *and direction vectors determined by* (2.5) *and* (2.11). *Then there exists an index* i, $1 \leqslant i \leqslant n$, *such that the direction vectors* $d_j, 1 \leqslant j \leqslant i$, *are mutually G-conjugate* (*i.e.* $d_j^{\mathrm{T}} G d_k = 0$ *whenever* $j \neq k$ *and* $1 \leqslant j \leqslant i, 1 \leqslant k \leqslant i$) *and*, *moreover*, $g_{i+1} = 0$ *and* $x_{i+1} = x_*$.

In general, the quadratic termination property requires perfect line searches. Since this property seemed essential in the past, many authors proposed variable metric methods keeping this property even without perfect line searches (see [20]). These methods are not used presently since they require expensive computations while quadratic termination was shown to be unnecessary for obtaining a superlinear rate of convergence (cf. Theorem 2.3). Time-consuming perfect line searches are also not used even if they have nice theoretical implications: Dixon [30] proved that all variable metric methods from the Broyden class generate identical points when perfect line searches are used.

Very general global-convergence results for imperfect line searches can be found in [16]. We summarize and generalize them in the following theorem, see [60].

Theorem 2.2. *Consider the variable metric method* (1.1) *with* $B_i d_i = -g_i$, (1.6), (1.7) *and* (2.15) *with* $0 < \underline{\gamma} \leqslant \gamma_i \leqslant \bar{\gamma}$, $0 < \underline{\rho} \leqslant \rho_i \leqslant \bar{\rho}$ *and* $(1 - \underline{\delta})\beta_i^{\mathrm{c}} \leqslant \beta_i \leqslant 1 - \underline{\delta}$, *where* $0 < \underline{\delta} < 1$. *Let the initial point* $x_1 \in \mathbb{R}^n$ *be chosen so that the objective function* $F : \mathbb{R}^n \to \mathbb{R}$ *is uniformly convex and has bounded second-order derivatives on the convex hull of the level set* $\mathscr{L}_1 = \{x \in \mathbb{R}^n \colon F(x) \leqslant F(x_1)\}$. *If there exist* $k \in \mathscr{N}$ *such that* $\gamma_i \geqslant 1$ $\forall i \geqslant k$, *then* $\liminf_{i \to \infty} \|g_i\| = 0$.

The above theorem has some important consequences. First, it cannot be proved when $\beta \geqslant 1$, which may be related to the bad properties of the DFP method. Secondly, it confirms that values $\beta^{\mathrm{c}} < \beta < 0$ (or $1 < \eta$) are permissible (computational experiments have shown that some particular methods from this subclass are very efficient in practice). Third, the restriction $\gamma \geqslant 1$ has also a practical consequence and it was used in [18] as an efficient strategy for scaling.

The above theorem has a weakness, namely the fact that it requires uniform convexity of the objective function. Fortunately, global convergence of the line-search method can be controlled by using restarts of the iterative process. If the value c_i, defined by (1.5), is not sufficiently positive, we can replace the unsuitable matrix H_i by an arbitrary well-conditioned positive-definite matrix ($H_i = I$, say). Theorem 2.2 shows that restarting eventually does not occur if the objective function is uniformly convex in a neighborhood of the minimizer.

Another way to guarantee global convergence of the line-search method consists in turning the search direction towards the negative gradient when necessary, i.e., when (1.5) is not satisfied. This idea is realized, e.g., if (2.5) is replaced by the formula $d = -\bar{H} g$ with

$$\bar{H} = H + \sigma \|Hg\| I \quad \text{or } \bar{H} = H + \sigma \|Hg\| \frac{gg^{\mathrm{T}}}{g^{\mathrm{T}} g}, \tag{2.19}$$

where H is a matrix obtained by update (2.11) and $\sigma > 0$ is a small number. Theoretical investigation of such modifications of variable metric methods is given in [76].

An important property of variable metric methods belonging to the Broyden class is their superlinear rate of convergence. Very general results concerning superlinear rate of convergence are given in [14]. We summarize them in the following theorem.

Theorem 2.3. *Let the assumptions of Theorem 2.2 be satisfied with $\rho_i = 1$ and $\gamma_i = 1$ and the line search be implemented in such a way that it always tries the steplength $\alpha_i = 1$ first. Let $x_i \to x_*$ and $G(x)$ be Lipschitz continuous at x_* (i.e. $\|G(x) - G(x_*)\| \leqslant \bar{L}\|x - x_*\|$ for all x from some neighborhood of x_*). Then a value $\underline{\beta} < 0$ exists such that if $\beta_i \geqslant \underline{\beta}$ $\forall i \in \mathcal{N}$, then* $\lim_{i\to\infty} \|x_{i+1} - x^*\|/\|x_i - x^*\| = 0$.

This theorem generalizes results proposed in [49], where a superlinear rate of convergence was proved for the restricted Broyden subclass corresponding to the values $0 \leqslant \beta \leqslant 1$ in (2.15) (it also generalizes results given in [26], where only DFP and BFGS symmetric updates are considered). The fact that a superlinear rate of convergence can be obtained for suitable negative values of the parameter β is very useful, since negative values positively influence the global convergence of variable metric methods, see [14].

The statement of Theorem 2.3 is true only if $\rho_i = 1$ and $\gamma_i = 1$. The influence of nonunit values of these parameters on the superlinear rate of convergence of the BFGS method was studied in [68], where it was shown that scaling applied in every iteration eventually requires nonunit values of the stepsize α_i (unless ρ_i and γ_i tend to one). This effect again increases the number of function evaluations.

2.3. Selected variable metric updates

Now we focus our attention to the choice of the value η (or β). Motivated by the above theoretical results, we will assume that $\beta^c < \beta \leqslant 1$ (or $0 \leqslant \eta$), defining the *perfect Broyden subclass*. Among all classic updates (2.12)–(2.14), only the BFGS method can be used in the basic unscaled form. The DFP method requires either accurate line search or scaling in every iteration, otherwise it need not converge. The problem of the unscaled SR1 formula consists in the fact that it does not guarantee positive definiteness of the generated matrices, so that the line search can fail. Therefore, either suitable scaling or a trust-region realization are necessary. Another simple choice

$$\eta = \frac{\rho/\gamma}{\rho/\gamma + a/b} \tag{2.20}$$

is proposed in [52]. This value is self-dual, lies in the restricted Broyden subclass and interpolates properties of both the DFP and the BFGS methods.

Particular variable metric methods are usually obtained by minimizing some potential functions. The most popular, used first in [82], see also [70], is a condition number

$$\kappa(H^{-1}H_+) = \bar{\lambda}(H^{-1}H_+)/\underline{\lambda}(H^{-1}H_+),$$

where H_+ is given by (2.7) and $\bar{\lambda}$ and $\underline{\lambda}$ are the maximum and the minimum eigenvalues, respectively. Writing $\tilde{\eta} = 1 - \eta/\eta^c$ and $\tilde{\omega} = (\rho/\gamma)(c/b)$, we can see that the matrix $H^{-1}H_+$ has $n-2$ unit eigenvalues and the remaining two eigenvalues $0 < \lambda_1 \leqslant \lambda_2$ are solutions of the quadratic equation $\lambda^2 - (\tilde{\eta} + \tilde{\omega})\lambda + \tilde{\eta}\tilde{\omega}b^2/(ac) = 0$. This fact implies that the ratio λ_2/λ_1 reaches its minimum if $\tilde{\eta} = \tilde{\omega}$ or

$$\eta(ac - b^2) = b^2\left(\frac{\rho}{\gamma}\frac{c}{b} - 1\right). \tag{2.21}$$

Taking into account the unit eigenvalues, we can see that the optimal value of η is given by

$$\eta = \frac{bc(\rho/\gamma - b/c)}{ac - b^2} \quad \text{if } b \leqslant \frac{2(\rho/\gamma)ac}{a + (\rho/\gamma)^2 c}, \tag{2.22}$$

$$\eta = \frac{\rho/\gamma}{\rho/\gamma - a/b} \quad \text{if } b > \frac{2(\rho/\gamma)ac}{a + (\rho/\gamma)^2 c} \tag{2.23}$$

(notice that (2.23) corresponds to the SR1 update). This optimally conditioned update was introduced in [20], although formula (2.22) was independently derived in [70].

Formula (2.21) can also be used for deriving the optimal ratio γ/ρ for a given value η, since we can write

$$\frac{\gamma}{\rho} = \frac{bc}{\eta(ac - b^2) + b^2}. \tag{2.24}$$

For $\eta = 1$ (BFGS) we obtain (2.8). For $\eta = 0$ (DFP) we obtain (2.9). For η given by (2.20) (Hoshino) we obtain (2.10). Substituting (2.10) back into (2.20) (or into (2.22)) we get the Oren–Spedicato update $\eta = b/(b + \sqrt{ac})$. Both the Hoshino and the Oren–Spedicato updates lie in the restricted Broyden subclass and, therefore, they are usually less efficient than the BFGS method in the unscaled case. The last case shows us a simple way for obtaining new variable metric updates. By finding the optimal ratio γ/ρ for a given value of η and substituting it back into the expression for η, we get a new update which differs from the original one if γ/ρ is not optimal.

This approach can also be used for the SR1 update. The analysis of update (2.14) shows that the matrix H keeps positive definiteness for b positive if and only if the ratio γ/ρ lies in the union of two disjoint open intervals $0 < \gamma/\rho < b/a$ and $c/b < \gamma/\rho < \infty$, see [85]. Inside each of these two intervals, exactly one value of the ratio γ/ρ exists which satisfies the Oren–Spedicato criterion. We consider only the interval $0 < \gamma/\rho < b/a$, since ratios $c/b < \gamma/\rho < \infty$ lead to unsuitable values $\eta < 0$. The optimal ratio $0 < \gamma/\rho < b/a$ for the SR1 update, derived from (2.23) to (2.24), can be expressed in the form

$$\frac{\gamma}{\rho} = \frac{c}{b}\left(1 - \sqrt{1 - b^2/(ac)}\right) = \frac{b}{a}\Big/\left(1 + \sqrt{1 - b^2/(ac)}\right), \tag{2.25}$$

which is the value proposed in [71]. The important property of this optimally scaled SR1 update is the fact that it generates positive-definite matrices. Unfortunately, this update leads to scaling applied in every iteration, which has a negative influence on the superlinear rate of convergence, as mentioned above. Substituting (2.25) in (2.23) (or (2.22)) we get

$$\eta = 1 + 1/\sqrt{1 - b^2/(ac)}. \tag{2.26}$$

This choice lies outside the restricted Broyden subclass and usually gives better results than the BFGS update in the unscaled case (see [56]). Another very efficient modification of the SR1 method is proposed in [3,56]. This is a combination of the SR1 and the BFGS updates which can be written in the form

$$\eta = 1 \quad \text{if } \rho/\gamma \leqslant a/b, \tag{2.27}$$

$$\eta = \frac{\rho/\gamma}{\rho/\gamma - a/b} \quad \text{if } \rho/\gamma > a/b, \tag{2.28}$$

i.e., $\eta = \max(1, (\rho/\gamma)/(\rho/\gamma - a/b))$. In other words, the SR1 update is chosen if and only if it lies in the perfect Broyden subclass.

Another potential function, which has frequently been used for deriving variable metric updates, is the weighted Frobenius norm $\|W^{-1}(\gamma B_+ - B)\|$ with W symmmetric and positive-definite. It was proved, see [42], that this Frobenius norm reaches its minimum on the set of matrices satisfying the generalized quasi-Newton condition (2.6), if and only if

$$\gamma B_+ = B + \frac{wv^{\mathrm{T}} + vw^{\mathrm{T}}}{s^{\mathrm{T}}v} - \frac{w^{\mathrm{T}}s}{s^{\mathrm{T}}v}\frac{vv^{\mathrm{T}}}{s^{\mathrm{T}}v}, \tag{2.29}$$

where $w = (\gamma/\rho)y - Bs$ and $v = Ws$. If the matrix W is chosen so that $v = Ws$ lies in the subspace generated by the vectors y and Bs (i.e., if $v = y + \lambda Bs$, say) we obtain a portion of the scaled Broyden class (2.15). This portion contains variable metric methods for which $p \geqslant 0$, where p is defined by (2.35) below. The relation between λ and β is given by

$$\beta = \frac{b(b - \lambda^2(\gamma/\rho)c)}{(b - \lambda c)^2}. \tag{2.30}$$

For $\beta = 0$ (BFGS) we get $\lambda = \sqrt{(\rho/\gamma)(b/c)}$. For $\beta = 1$ (DFP) we get $\lambda = 0$. For $\beta = (\gamma/\rho)/(\gamma/\rho - c/b)$ (SR1) we get $\lambda = \rho/\gamma$.

If we set $W = I$ in (2.29), we get the Powell symmetric Broyden (PSB) update

$$\gamma B_+ = B + \frac{sw^{\mathrm{T}} + ws^{\mathrm{T}}}{s^{\mathrm{T}}s} - \frac{w^{\mathrm{T}}s}{s^{\mathrm{T}}s}\frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s}. \tag{2.31}$$

The PSB method does not guarantee positive definiteness of the generated matrices, so that the line search can fail. Therefore, a trust-region realization is necessary. Generally, this method is highly inefficient even if it is superlinearly convergent (and the proof of its superlinear rate of convergence, cf. Theorem 3.1, is much easier than the proof of Theorem 2.3).

Other potential functions have been used for deriving variable metric methods. If $X = H^{-1}H_+$ then [34] shows that the DFP update minimizes the function

$$\psi(X) = \mathrm{trace}(X) - \log(\det(X)), \tag{2.32}$$

on the set of positive-definite matrices H_+ satisfying the quasi-Newton condition $H_+y = d$. Similarly, the BFGS method minimizes (2.32), where $X = B^{-1}B_+$, on the set of positive-definite matrices B_+ satisfying the quasi-Newton condition $B_+d = y$. The functions

$$\sigma(X) = \bar{\lambda}(X)/\sqrt{\det(X)},$$

$$\tau(X) = \mathrm{trace}(X)/(n\underline{\lambda}(X))$$

are both minimized (either for $X = H^{-1}H_+$ or for $X = B^{-1}B_+$) by the optimally scaled SR1 updates, see [95,96] ($\bar{\lambda}$ and $\underline{\lambda}$ are maximum and minimum eigenvalues).

Besides the above potential functions, other principles have been used for the derivation of free parameters in the Broyden class of variable metric methods. Byrd et al. [14] recommend a theoretical value $\beta = \beta^{\mathrm{c}} + (1/c)(1/v^{\mathrm{T}}G^{-1}v)$, where $v = (1/b)y - (1/c)Bd$ and G is the exact Hessian matrix. Unfortunately, the exact Hessian matrix is usually unknown and so must be approximated. In [57], a simple approximation $G \approx (1/\gamma)B$ is used with γ given by (2.10) (with $\rho = 1$). Using the expression

Table 1

η	NS with $\rho=1$	PS with $\rho=1$	IS with $\rho=1$
BFGS	7042–10 409	7182–8008	4162–5059
DFP	26 failures	36 failures	6301–7642
(2.20)	8288–10 701	9538–10 118	4316–4892
(2.22)–(2.23)	7038–9290	6821–7557	4522–5052
(2.26)	5940–9979	5358–6543	4065–5340
(2.27)–(2.28)	5888–9596	5022– 6085	4173–5095
(2.33)	6044–9047	5663–6538	4152–4913
η	NS with (2.4)	PS with (2.4)	IS with (2.4)
BFGS	6800–10 120	6742–7430	4127–5049
DFP	24 failures	36 failures	5027–6102
(2.20)	8648–11 003	8720–9356	4218–4883
(2.22)–(2.23)	7444–9542	6130–6684	4324–4821
(2.26)	6112–10 203	5402–6559	3962–5230
(2.27)–(2.28)	5882–9645	4881–6075	4106–5066
(2.33)	5787–8538	5315–6042	3927–4589

for v, we can write $v^{\mathrm{T}}G^{-1}v \approx \gamma v^{\mathrm{T}}Hv = (\gamma/c)(ac/b^2-1)$, which together with (2.18) and (2.16) gives $\eta = (ac\sqrt{c/a}-b^2)/(ac-b^2)$. Keeping the numerator nonnegative, we obtain the formula

$$\eta = \frac{\max(0, \sqrt{c/a}-b^2/(ac))}{1-b^2/(ac)}. \tag{2.33}$$

Note that the denominator in (2.33) and the same expression in (2.26) are usually replaced by $\max(\varepsilon, 1-b^2/(ac))$ with ε a small number (10^{-60}, say). This is a safeguard against division by zero caused by round-off errors.

Finally, we notice that the rank-two update classes we have considered so far, namely updates (2.11) and (2.29), are only special cases of the set of solutions of the quasi-Newton equation. Since the quasi-Newton equation can be viewed as a set of n linear systems, each consisting of a single equation and all differing only in the right hand side, the general solution can easily be obtained using the techniques offered by the ABS class of algorithms for linear equations, see [1]. The general formula obtained contains two parameter matrices, see [87], and is equivalent to a formula previously obtained in [2], using the theory of generalized inverses. No new updates in this general class have yet been developed.

Table 1 compares several variable metric methods of the form (2.11) with standard line-search. They are either unscaled (NS) or use preliminary scaling (PS) or interval scaling (IS). Both the value $\rho=1$ and the nonquadratic correction (2.4) were used. Values of scaling parameter γ have been selected from (2.8) to (2.10) to give the best results for individual methods — i.e., (2.9) for the DFP update and (2.10) for all other updates. Total numbers of iterations and function evaluations for 74 problems (22 from TEST14, 22 from TEST15, 30 from TEST18, [62]) with 20 variables are presented.

Table 1 implies recommendations for the choice of suitable variable metric methods. First, a reasonable scaling strategy, e.g., IS, should be used since it improves efficiency of all investigated updates. Furthermore, if interval scaling is used, then the easily implementable Hoshino method (2.20) is very efficient. Also, update (2.33) is excellent but more complicated (it must be safeguarded against division by zero as shown above). The nonquadratic correction (2.4) improves this update significantly.

An interesting realization of variable metric methods is based on product-form updates. Suppose that $H = ZZ^{\mathrm{T}}$, where Z is a nonsingular square matrix. Then the direction vector $d = -Hg$ can be obtained using three substitutions

$$d = Z\tilde{d}, \quad \tilde{d} = -\tilde{g}, \quad \tilde{g} = Z^{\mathrm{T}}g. \tag{2.34}$$

We write $\tilde{s} = Z^{-1}s = \alpha\tilde{d}$ and $\tilde{y} = Z^{\mathrm{T}}y$ so that $a = \tilde{y}^{\mathrm{T}}\tilde{y}$, $b = \tilde{y}^{\mathrm{T}}\tilde{s}$, $c = \tilde{s}^{\mathrm{T}}\tilde{s}$. If

$$p = \frac{1}{ab}\left(\eta\left(\frac{a}{b} - \frac{\rho}{\gamma}\right) + \frac{\rho}{\gamma}\right) \geqslant 0, \tag{2.35}$$

$$q = \frac{\rho}{\gamma}\frac{1}{ab}(\eta(ac - b^2) + b^2) \geqslant 0, \tag{2.36}$$

then the matrix H_+ can be expressed in the form $H_+ = Z_+Z_+^{\mathrm{T}}$, where Z_+ is obtained from Z by a rank one formula. The general update, derived in [20], is rather complicated, but it contains special cases, which have acceptable complexity. Setting $\eta = 0$ (DFP), we get $p = \rho/(\gamma ab)$ and $q = \rho b/(\gamma a)$, so that

$$\frac{1}{\sqrt{\gamma}}Z_+ = Z + \frac{1}{a}Z\left(\sqrt{\frac{\rho a}{\gamma b}}\tilde{s} - \tilde{y}\right)\tilde{y}^{\mathrm{T}}. \tag{2.37}$$

Setting $\eta = 1$ (BFGS), we get $p = 1/b^2$ and $q = \rho c/(\gamma b)$, so that

$$\frac{1}{\sqrt{\gamma}}Z_+ = Z + \frac{1}{b}Z\tilde{s}\left(\sqrt{\frac{\rho b}{\gamma c}}\tilde{s} - \tilde{y}\right)^{\mathrm{T}}. \tag{2.38}$$

Setting $\eta = (\rho/\gamma)(\rho/\gamma - a/b)$ (SR1), we get $p = 0$ and $q = ((\rho/\gamma)c - b)/(b - (\gamma/\rho)a)$, so that

$$\frac{1}{\sqrt{\gamma}}Z_+ = Z + \frac{\sqrt{q} - 1}{(\rho/\gamma)^2c - 2(\rho/\gamma)b + a}Z\left(\frac{\rho}{\gamma}\tilde{s} - \tilde{y}\right)\left(\frac{\rho}{\gamma}\tilde{s} - \tilde{y}\right)^{\mathrm{T}}. \tag{2.39}$$

Theoretically, it would be possible to invert the above formulas to obtain similar expressions for the matrix $A_+ = Z_+^{-1}$. Unfortunately, the vector $\tilde{y} = Z^{\mathrm{T}}y = (A^{\mathrm{T}})^{-1}y$, required in that case, cannot be determined without inversion of the matrix A. The BFGS update, obtained by inversion of (2.38), is the only one that allows us to overcome this difficulty by using the following transformation:

$$\sqrt{\gamma}A_+ = A + \frac{1}{c}\tilde{s}\left(\sqrt{\frac{\gamma c}{\rho b}}\tilde{y} - \tilde{s}\right)^{\mathrm{T}}A = A + \frac{1}{c}As\left(\sqrt{\frac{\gamma c}{\rho b}}y - A^{\mathrm{T}}As\right)^{\mathrm{T}}. \tag{2.40}$$

Formulae (2.37)–(2.39) are very advantageous for seeking minima on linear manifolds, when the matrix H is singular and the matrix Z is rectangular. Formula (2.40) is useful for nonlinear least squares.

3. Variable metric methods for large-scale problems

Basic variable metric methods cannot be used for large-scale optimization, since they utilize dense matrices. Therefore, new principles have to be found, which take into account the sparsity pattern of the Hessian matrix. There are three basic approaches: preserving the sparsity pattern by special updates; using classic updates applied to submatrices of lower dimension; and reconstruction of matrices from vectors by limited memory methods. The first approach was initiated in [91], the second was proposed in [48] and the third was introduced in [67].

3.1. Sparse variable metric updates

Preserving a sparsity pattern is a strong restriction, which eliminates some important properties of variable metric methods. In general, updates cannot have a low rank. For instance, a diagonal update of a diagonal matrix, which changes it to satisfy the quasi-Newton condition, can have rank n. Moreover, positive definiteness of the updated matrix can be lost for an arbitrary sparse update, which can again be demonstrated on a diagonal matrix. From this point of view, it is interesting that a superlinear rate of convergence can be obtained even if the quadratic termination property does not hold.

Sparse variable metric updates should satisfy the quasi-Newton condition, not violate symmetry and preserve sparsity. Let us write

$$\mathscr{V}_Q = \{B \in \mathbb{R}^{n\times n}: Bs = y\},$$
$$\mathscr{V}_S = \{B \in \mathbb{R}^{n\times n}: B^{\mathrm{T}} = B\},$$
$$\mathscr{V}_G = \{B \in \mathbb{R}^{n\times n}: G_{ij} = 0 \Rightarrow B_{ij} = 0\}$$

(we assume, that $G_{ii} \neq 0 \;\forall 1 \leqslant i \leqslant n$). Clearly, $\mathscr{V}_Q$, $\mathscr{V}_S$, $\mathscr{V}_G$ are linear manifolds ($\mathscr{V}_S$ and $\mathscr{V}_G$ are subspaces) in $\mathbb{R}^{n\times n}$. We can define orthogonal projections $\mathscr{P}_Q$, $\mathscr{P}_S$, $\mathscr{P}_G$ into $\mathscr{V}_Q$, $\mathscr{V}_S$, $\mathscr{V}_G$ as matrices B_+ minimizing the Frobenius norm $\|B_+ - B\|_F$ on $\mathscr{V}_Q$, $\mathscr{V}_S$, $\mathscr{V}_G$, respectively. Similarly, we can define orthogonal projections $\mathscr{P}_{QS}$, $\mathscr{P}_{QG}$, $\mathscr{P}_{SG}$ and $\mathscr{P}_{QSG}$ into $\mathscr{V}_Q \cap \mathscr{V}_S$, $\mathscr{V}_Q \cap \mathscr{V}_G$, $\mathscr{V}_S \cap \mathscr{V}_G$ and $\mathscr{V}_Q \cap \mathscr{V}_S \cap \mathscr{V}_G$, respectively. It is clear that the requirements laid down on a sparse update are satisfed by the matrix $B^+ = \mathscr{P}_{QSG}B$.

To eliminate the zero elements from the quasi-Newton condition, we define vectors $\mathscr{P}_i s \in \mathbb{R}^n$, $1 \leqslant i \leqslant n$, in such a way that

$$e_j^{\mathrm{T}}\mathscr{P}_i s = e_j^{\mathrm{T}} s, \quad G_{ij} \neq 0,$$
$$e_j^{\mathrm{T}}\mathscr{P}_i s = 0, \quad G_{ij} = 0$$

and we rewrite the quasi-Newton condition in the form

$$e_i^{\mathrm{T}}(B_+ - B)\mathscr{P}_i s = e_i^{\mathrm{T}}(y - Bs), \quad 1 \leqslant i \leqslant n.$$

It can be proved, [27], that the orthogonal projections considered can be expressed as

$$\mathscr{P}_Q B = B + \frac{(y - Bs)s^{\mathrm{T}}}{s^{\mathrm{T}}s},$$
$$\mathscr{P}_S B = \tfrac{1}{2}(B + B^{\mathrm{T}}),$$
$$(\mathscr{P}_G B)_{ij} = B_{ij}, \quad G_{ij} \neq 0,$$

$$
\begin{aligned}
&(\mathscr{P}_G B)_{ij} = 0, \quad G_{ij} = 0,\\
&\mathscr{P}_{QS}B = B + \frac{(y-Bs)s^{\mathrm{T}} + s(y-Bs)^{\mathrm{T}}}{s^{\mathrm{T}}s} - \frac{(y-Bs)^{\mathrm{T}}s}{s^{\mathrm{T}}s}\frac{ss^{\mathrm{T}}}{s^{\mathrm{T}}s},\\
&\mathscr{P}_{QG}B = B + \mathscr{P}_G(us^{\mathrm{T}}),\\
&\mathscr{P}_{SG}B = \mathscr{P}_S\mathscr{P}_G B = \mathscr{P}_G\mathscr{P}_S B,\\
&\mathscr{P}_{QSG}B = B + \mathscr{P}_G(vs^{\mathrm{T}} + sv^{\mathrm{T}}),
\end{aligned}
$$

where $u \in \mathbb{R}^n$ solves the linear system $Du = y - Bs$ with positive-semidefinite diagonal matrix

$$D = \sum_{i=1}^{n} \|\mathscr{P}_i s\|^2 e_i e_i^{\mathrm{T}}$$

and $v \in \mathbb{R}^n$ solves the linear system $Qv = y - Bs$ with positive-semidefinite matrix

$$Q = \mathscr{P}_G(ss^{\mathrm{T}}) + \sum_{i=1}^{n} \|\mathscr{P}_i s\|^2 e_i e_i^{\mathrm{T}},$$

which has the same sparsity pattern as the matrix B.

The variable metric method which uses the update

$$B_+ = \mathscr{P}_{QSG}B, \tag{3.1}$$

was proposed in [91]. Realization of this method is time consuming, since an additional linear system has to be solved. Moreover, its convergence properties are not very good, since its variational derivation is similar to the derivation of the inefficient PSB method. Therefore, easier methods with better convergence properties have been looked for. Steihaug [89] has shown that the updates based on the composite projections

$$B^+ = \mathscr{P}_S\mathscr{P}_{QG}B, \tag{3.2}$$

$$B^+ = \mathscr{P}_G\mathscr{P}_{QS}B, \tag{3.3}$$

$$B^+ = \mathscr{P}_{SG}\mathscr{P}_Q B \tag{3.4}$$

and realized in the trust-region framework, lead to methods which are globally and superlinearly convergent. We summarize his results in the following theorem.

Theorem 3.1. *Consider the trust-region method* (1.13)–(1.19), *where* $B_{i+1} = B_i$, *if* (1.16) *holds, or updates of the form* (3.1)–(3.4) *are used, if* (1.17) *holds. Let the objective function* $F : \mathbb{R}^N \to \mathbb{R}$ *be bounded from below and have bounded and Lipschitz continuous second-order derivatives. Then* $\liminf_{i\to\infty} \|g_i\| = 0$. *If, in addition,* $x_i \to x_*$ *and* $\omega_i \to 0$, *see* (1.14), *then* $\lim_{i\to\infty} \|x_{i+1} - x^*\|/\|x_i - x^*\| = 0$.

Unfortunately, a similar result cannot be obtained for a line-search realization, since the hereditary positive definiteness of generated matrices is not guaranteed. Nevertheless, our unpublished experiments indicate that a line-search realization usually outperforms a trust-region implementation. These experiments also imply that update (3.2) is the most efficient one among all composite projections. This fact is also mentioned in [29,94].

Table 2

Method	Iterations	f. eval.	g. eval.	CG steps	CPU time	Failures
LVVM	26 739	27 901	27 901	—	1 : 23	—
LMVM	27 282	31 723	31 723	—	1 : 35	—
LRVM	28 027	30 061	30 061	—	1 : 32	—
SCVM	13 145	27 292	27 292	51 0773	4 : 10	1
SFVM	5308	16 543	41 732	—	1 : 54	1
SPVM	3769	5190	5190	—	0 : 30	—
SDNM	1958	2000	10 238	—	0 : 34	—
STNM	2203	2980	60 420	57 195	1 : 14	—
NCGM	19 974	39 854	39 854	—	1 : 29	—

To eliminate difficulties arising in connection with update (3.1), Tůma has proposed sparse fractioned updates [94]. Let $\mathscr{G} = (V, E)$, $V = \{v_1, \dots, v_n\}$, $E \in V \times V$, be the adjacency graph of the matrix G so that $(v_i, v_j) \in E$ if and only if $G_{ij} \neq 0$ (structurally). Let c: $V \to \{1, \dots, r\}$, $r \leqslant n$ be a colouring of the graph $\mathscr{G}$ so that $c(v_i) \neq c(v_j)$ if and only if $(v_i, v_j) \in E$ (the minimum possible r is the so-called chromatic number of the graph $\mathscr{G}$). This colouring induces a partition $V = \bigcup_{i=1}^{r} C_i$ where $C_i = \{v \in V\colon c(v) = i\}$. Assume now that $s = \sum_{i=1}^{r} s^i$ where $s^i = \sum_{j \in C_i} e_j e_j^{\mathrm{T}} s$ and set

$$B_+ = B^r, \tag{3.5}$$

where

$$x^0 = x, \quad g^0 = g, \quad B^0 = B$$

and

$$x^i = x^{i-1} + s^i, \quad g^i = g(x^i), \quad y^i = g^i - g^{i-1},$$

$$B^i = \mathscr{P}_{Q^i SG} B^{i-1}, \quad \mathscr{V}_{Q^i} = \{B \in \mathbb{R}^{n \times n} : Bs^i = y^i\}$$

for $1 \leqslant i \leqslant r$. As has been already shown, $\mathscr{P}_{Q^i SG} B^{i-1} = B^{i-1} + \mathscr{P}_G(v^i (s^i)^{\mathrm{T}} + s^i (v^i)^{\mathrm{T}})$, where $Q^i v^i = y^i - B^{i-1} s^i$ and where

$$Q^i = \mathscr{P}_G(s^i (s^i)^{\mathrm{T}}) + \sum_{j=1}^{n} \|\mathscr{P}_j s^i\|^2 e_j e_j^{\mathrm{T}} = \sum_{j \in C_i} e_j e_j^{\mathrm{T}} s s^{\mathrm{T}} e_j e_j^{\mathrm{T}} + \sum_{j=1}^{n} \|\mathscr{P}_j s^i\|^2 e_j e_j^{\mathrm{T}}$$

is now a diagonal matrix. Since the matrices Q^i, $1 \leqslant i \leqslant r$ are diagonal, the partial updates $B^i = \mathscr{P}_{Q^i SG} B^{i-1}$, are very simple and can be realized in an efficient way. Notice that this simplicity is compensated by evaluation of intermediate gradients $g^1, \dots, g^{r-1}$. This is a common feature with the method of approximating sparse Hessian matrices proposed by Coleman and Moré [17]. However, the number of groups induced by colouring c given above can be much smaller than the number of groups induced by the symmetric or lower triangular colouring used by Coleman and Moré. Computational experiments confirm that sparse fractioned updates are more efficient than update (3.1) and than composite projections (3.2)–(3.4) (see Table 2).

Another way of obtaining sparse quasi-Newton updates is described in [35]. This method is based on the minimization of the potential function (2.32), where $X = HB_+$, on the linear manifold $\mathscr{V}_Q \cap \mathscr{V}_S \cap \mathscr{V}_G$. Function (2.32) has two advantages. First, its minimization leads to the efficient BFGS formula in the dense case and, secondly, it serves as a barrier function against losing positive definiteness. Fletcher [35] proved that if the minimum of (2.32) on $\mathscr{V}_Q \cap \mathscr{V}_S \cap \mathscr{V}_G$ exists, it is characterized by the existence of $\lambda \in \mathbb{R}^n$ such that

$$\mathscr{P}_G H_+ = \mathscr{P}_G(H + \lambda s^{\mathrm{T}} + s\lambda^{\mathrm{T}}). \tag{3.6}$$

The vector λ cannot be obtained explicitly in the sparse case. Instead, the nonlinear system of equations $B_+(\lambda)s - y = 0$ must be solved using the Newton method, where $B_+(\lambda)$ is a matrix determined from (3.6). This approach has two difficulties. Firstly, the determination of $B_+(\lambda)$ from $\mathscr{P}_G H_+(\lambda)$ is rather complicated and it requires a sparsity pattern which is not changed during the Cholesky decomposition. Secondly, the nonlinear equations have to be solved with the Jacobian matrix M, say, which has the same pattern as B in general. Therefore, the whole process is time consuming and moreover three sparse matrices B, $\mathscr{P}_G H$ and M are necessary. Nevertheless, numerical experiments in [35] indicate robustness and good convergence properties of this method.

Finally, we observe that the approach based upon use of the ABS algorithm can also provide the general solution of the quasi-Newton equation with sparsity and symmetry conditions, since they are just additional linear equations, see [85,88]. The sparse symmetric update is given in explicit form, while in the approach of, e.g., [91], a sparse linear system has to be solved. By requiring that the diagonal element be sufficiently large, extra linear conditions are given which in general allow us to obtain symmetric sparse quasi-positive-definite updates (i.e., updates where the $(n-1)$th principal submatrix is SPD) and quasi-diagonally dominant updates, see [88,86]. The last result can be used to produce full SPD sparse updates by imbedding the minimization of the function $F(x)$ in a suitable equivalent $(n+1)$-dimensional problem. No particular algorithms or numerical experiments are yet available based upon this approach.

3.2. Partitioned variable metric updates

A quite different approach to large-scale optimization, leading to partitioned updating methods, is proposed in [48]. It is based on properties of partially separable functions of the form

$$F(x) = \sum_{k=1}^{m} f_k(x), \tag{3.7}$$

where each of the element function $f_k(x)$ depends only on n_k variables and n_k is much less than n, the size of the original problem. In this case, we can define packed element-gradients $\hat{g}_k(x) \in \mathbb{R}^{n_k}$ and packed element-Hessian matrices $\hat{G}_k(x) \in \mathbb{R}^{n_k \times n_k}$, $1 \leqslant k \leqslant m$, as dense but small-size vectors and matrices. Such a formulation is highly practical since, e.g., sparse nonlinear least-square problems (see (4.1) below) have this structure.

Partitioned updating methods consider each element function separately and update approximations $\hat{B}_k$, $1 \leqslant k \leqslant m$, of the packed element-Hessian matrices $\hat{G}_k(x)$ using the quasi-Newton conditions $\hat{B}_k^+ \hat{s}_k = \hat{y}_k$, where $\hat{s}_k$ is a part of the vector s consisting of components corresponding to variables of f_k and $\hat{y}_k = \hat{g}_k^+ - \hat{g}_k$ (we use $+$ as the upper index in the partitioned case). Therefore, a variable metric update of the form (2.15) can be used for each of the element functions. However, there are some differences between the classic and the partitioned approach. First, the main reason for partitioned update is an approximation of the element Hessian matrix, so that scaling and nonquadratic corrections do not usually improve efficiency. Secondly, denoting $\hat{b}_k = \hat{y}_k^{\mathrm{T}} \hat{s}_k$, $\hat{c}_k = \hat{s}_k^{\mathrm{T}} \hat{B}_k \hat{s}_k$, we can observe that $\hat{b}_k \geqslant 0$ does not have to be guaranteed for all $1 \leqslant k \leqslant m$. This difficulty is unavoidable and an efficient algorithm has to handle this situation. Therefore, the following partitioned BFGS method is recommended:

$$\begin{aligned} \hat{B}_k^+ &= \hat{B}_k + \frac{1}{\hat{b}_k}\hat{y}_k\hat{y}_k^{\mathrm{T}} - \frac{1}{\hat{c}_k}\hat{B}_k\hat{s}_k(\hat{B}_k\hat{s}_k)^{\mathrm{T}}, \quad \hat{b}_k > 0, \\ \hat{B}_k^+ &= \hat{B}_k, \quad \hat{b}_k \leqslant 0. \end{aligned} \tag{3.8}$$

Another possibility is the partitioned rank one method

$$\begin{aligned} \hat{B}_k^+ &= \hat{B}_k + \frac{1}{\hat{b}_k - \hat{c}_k}(\hat{y}_k - \hat{B}_k\hat{s}_k)(\hat{y}_k - \hat{B}_k\hat{s}_k)^{\mathrm{T}}, \quad |\hat{b}_k - \hat{c}_k| \neq 0, \\ \hat{B}_k^+ &= \hat{B}_k, \quad |\hat{b}_k - \hat{c}_k| = 0. \end{aligned} \tag{3.9}$$

which can be used for indefinite matrices. Usually, the latter method works worse but can be useful in some pathological cases. Therefore, combined methods are welcome. One such combination is proposed in [50]. It starts with the partitioned BFGS update (3.8). When a negative curvature $\hat{b}_k < 0$ appears in some iteration then (3.8) is switched to (3.9) for $\hat{B}_k$ and is kept in all subsequent iterations. We suggest another strategy, which was used in our experiments reported in Table 2. This is based on the observation that (3.8) usually fails in the case when too many elements have indefinite Hessian matrices. Therefore, we start with the partitioned BFGS update (3.8). If $m_{\mathrm{neg}} \geqslant \theta m$, where m_{neg} is a number of elements with a negative curvature and θ is a threshold value, then (3.9) is used for all elements in all subsequent iterations (we recommend $\theta = \frac{1}{2}$).

Partitioned variable metric methods are very efficient for solving real-world problems, but their convergence properties have not yet been satisfactorily investigated. Griewank and Toint [49] have proved a superlinear rate of convergence of partitioned variable metric methods belonging to the restricted Broyden class. Unfortunately, a general global-convergence theory, which would include the most efficient algorithms, e.g., the partitioned BFGS method given above, is not known. Some partial results are given in [92], where global convergence is proved under complicated and restrictive conditions. Some globally convergent modifications of partitioned variable metric methods are also given in [47]. Unfortunately, we have experimentally found that these modifications are computationally less efficient and cannot be competitive with the best strategies given above.

A disadvantage of partitioned variable metric methods is that approximations of packed element-Hessian matrices have to be stored. Therefore, the number of stored elements can be much greater than the number of nonzero elements in the standard sparse pattern. For this reason, it is suitable to construct the standard sparse Hessian approximation before solving a linear system, since a multiplication by a sparse matrix is more efficient than the use of the partitioned structure.

3.3. Variable metric methods with limited memory

Variable metric methods with limited memory are based on the application of a limited number of BFGS updates, which are computed recursively using previous differences s_j, y_j, $i-n \leqslant j \leqslant i-1$ (i is the iteration number). Their development started by the observation that an application of the BFGS update is equivalent to a conjugate gradient step in the case of perfect line search, see [72], and is more efficient in other cases. In [12,13] a limited number of BFGS steps was used for construction of a suitable preconditioner to the conjugate gradient method; and a similar approach has been used for the approximation of the Hessian matrix, see [67,55,43]. Such applications have been made possible by a special form of the BFGS update

$$H_+ = \gamma V^{\mathrm{T}} H V + \frac{\rho}{b} s s^{\mathrm{T}},$$
$$V = I - \frac{1}{b} y s^{\mathrm{T}}$$

We define the m-step BFGS method with limited memory as the iterative process (1.1) and (2.5), where $H_i = H_i^i$ and the matrix H_i^i is generated by the recurrence formula

$$H_{j+1}^i = \gamma_j^i V_j^{\mathrm{T}} H_j^i V_j + \frac{\rho_j}{b_j} s_j s_j^{\mathrm{T}} \tag{3.10}$$

for $i-m \leqslant j \leqslant i-1$, where $H_{i-m}^i = I$. At the same time $\gamma_{i-m}^i = b_{i-1}/a_{i-1}$ and $\gamma_j^i = 1$ for $i-m < j \leqslant i-1$. Using induction, we can rewrite (3.10) in the form

$$H_{j+1}^i = \frac{b_{i-1}}{a_{i-1}} \left(\prod_{k=i-m}^{j} V_k \right)^{\mathrm{T}} \left(\prod_{k=i-m}^{j} V_k \right) + \sum_{l=i-m}^{j} \frac{\rho_l}{b_l} \left(\prod_{k=l+1}^{j} V_k \right)^{\mathrm{T}} s_l s_l^{\mathrm{T}} \left(\prod_{k=l+1}^{j} V_k \right) \tag{3.11}$$

for $i-m \leqslant j \leqslant i-1$. From (3.11), we can deduce that the matrix H_i^i can be determined using $2m$ vectors $s_j, y_j, i-m \leqslant j \leqslant i-1$, without storing the matrices H_j^i, $i-m \leqslant j \leqslant i-1$. This matrix need not be constructed explicitly since we need only the vector $s_i = -H_i^i g_i$, which can be computed using two recurrences (the Strang formula [67]). First, the vectors

$$u_j = -\left(\prod_{k=j}^{i-1} V_k \right) g_i,$$

where $i-1 \geqslant j \geqslant i-m$, are computed using the backward recurrence

$$\sigma_j = s_j^{\mathrm{T}} u_{j+1} / b_j,$$
$$u_j = u_{j+1} - \sigma_j y_j$$

for $i-1 \geqslant j \geqslant i-m$, where $u_i = -g_i$. Then the vectors

$$v_{j+1} = \frac{b_{i-1}}{a_{i-1}} \left(\prod_{k=i-m}^{j} V_k \right)^{\mathrm{T}} u_{i-m} + \sum_{l=i-m}^{j} \frac{\rho_l}{b_l} \left(\prod_{k=l+1}^{j} V_k \right)^{\mathrm{T}} s_l s_l^{\mathrm{T}} u_{l+1},$$

where $i-m \leqslant j \leqslant i-1$, are computed using the forward recurrence

$$v_{i-m} = (b_{i-1}/a_{i-1}) u_{i-m},$$
$$v_{j+1} = v_j + (\rho_j \sigma_j - y_j^{\mathrm{T}} v_j) s_j$$

for $i-m \leqslant j \leqslant i-1$, where $v_{i-m} = (b_{i-1}/a_{i-1}) u_{i-m}$. Finally we set $s_i = v_i$.

Recently, a new approach to variable metric methods with limited memory, based on explicit expression of the matrix $H_i = H_i^i$ using low-order matrices, was proposed in [15]. Let $H_i = H_i^i$ be the matrix obtained after m steps of the form

$$H_{j+1}^i = H_j^i + [s_j, H_j^i y_j] M_j [s_j, H_j^i y_j]^{\mathrm{T}},$$

$i-m\leqslant j\leqslant i-1$, where M_j, $i-m\leqslant j\leqslant i-1$, are symmetric 2×2 matrices which realize a particular variable metric method (2.11) with $\rho_j=\gamma_j=1$. We need an expression

$$H_i = H^i_{i-m} - [S_i, H^i_{i-m}Y_i]N_i^{-1}[S_i, H^i_{i-m}Y_i]^{\mathrm{T}}, \tag{3.12}$$

where $S_i=[s_{i-m},\dots,s_{i-1}]$, $Y_i=[y_{i-m},\dots,y_{i-1}]$ and N_i is a symmetric matrix of order $2m$. Formula (3.12) was obtained for classical variable metric methods (DFP, BFGS, SR1), since the matrices M_j^{-1}, $i-m\leqslant j\leqslant i-1$, have a relatively simple form in these cases. Derivations, which can be found in [15], are formally rather complicated. Therefore, we introduce only the final results. For this purpose, we denote by R_i the upper triangular matrix of order m, such that $(R_i)_{kl}=s_k^{\mathrm{T}}y_l$, for $k\leqslant l$, and $(R_i)_{kl}=0$, otherwise. Furthermore, we denote by C_i the diagonal matrix of order m, such that $(C_i)_{kk}=s_k^{\mathrm{T}}y_k$. Taking

$$N_i = \begin{bmatrix} -C_i & R_i - C_i \\ (R_i-C_i)^{\mathrm{T}} & Y_i^{\mathrm{T}}H^i_{i-m}Y_i \end{bmatrix} \tag{3.13}$$

in (3.12), we get the m-step DFP update. Taking

$$N_i = \begin{bmatrix} 0 & R_i \\ R_i^{\mathrm{T}} & C_i + Y_i^{\mathrm{T}}H^i_{i-m}Y_i \end{bmatrix} \tag{3.14}$$

in (3.12), we get the m-step BFGS update. The m-step SR1 update can be written in the following slightly simpler form:

$$H_i = H^i_{i-m} + (S_i - H^i_{i-m}Y_i)(R_i + R_i^{\mathrm{T}} - C_i - Y_i^{\mathrm{T}}H^i_{i-m}Y_i)^{-1}(S_i - H^i_{i-m}Y_i)^{\mathrm{T}}. \tag{3.15}$$

In the sequel, we restrict our attention to the BFGS method. If we choose $H^i_{i-m}=\gamma^i_{i-m}I$, where $\gamma^i_{i-m}=b_{i-1}/a_{i-1}$, and if we explicitly invert matrix (3.14), we can write

$$H_i = \gamma_{i-m}I + [S_i, \gamma_{i-m}Y_i]\begin{bmatrix} (R_i^{-1})^{\mathrm{T}}(C_i+\gamma_{i-m}Y_i^{\mathrm{T}}Y_i)R_i^{-1} & -(R_i^{-1})^{\mathrm{T}} \\ -R_i^{-1} & 0 \end{bmatrix}[S_i, \gamma_{i-m}Y_i]^{\mathrm{T}}. \tag{3.16}$$

This formula has the advantage that no inversion or matrix decomposition is used.

Similar explicit expressions can be obtained for the matrices $B_i=H_i^{-1}$ using duality relations. Since we replace S_i and Y_i by Y_i and S_i, respectively, we have to replace the upper part of $S_i^{\mathrm{T}}Y_i$ by the upper part of $Y_i^{\mathrm{T}}S_i$ (or by the transposed lower part of $S_i^{\mathrm{T}}Y_i$). Therefore, we define the lower triangular matrix L_i, such that $(L_i)_{kl}=s_k^{\mathrm{T}}y_l$, $k\geqslant l$ and $(L_i)_{kl}=0$, otherwise. Then the m-step BFGS update can be written in the form

$$B_i = B^i_{i-m} - [Y_i, B^i_{i-m}S_i]\begin{bmatrix} -C_i & (L_i-C_i)^{\mathrm{T}} \\ L_i - C_i & S_i^{\mathrm{T}}B^i_{i-m}S_i \end{bmatrix}^{-1}[Y_i, B^i_{i-m}S_i]^{\mathrm{T}}. \tag{3.17}$$

The limited-memory variable metric methods described above require a double set of difference vectors. Fletcher [33] has proposed a method that requires only a single set of these vectors. The same property is possessed by the limited-memory reduced-Hessian variable metric methods introduced in [44] and based on product form updates investigated in [79]. Consider variable metric methods of the form (2.15) with $B_1=I$ (the unit matrix). Let $\mathscr{G}_i$ and $\mathscr{D}_i$ be linear subspaces spanned by the columns of matrices $G_i=[g_1,\dots,g_i]$ and $D_i=[d_1,\dots,d_i]$, respectively. In [79] it is proved that $\mathscr{D}_i=\mathscr{G}_i$ and that $B_iv\in\mathscr{G}_i$ and $B_iw=\lambda_iw$, whenever $v\in\mathscr{G}_i$ and $w\in\mathscr{G}_i^{\perp}$ (a possible nonunit value λ_i is a consequence of nonquadratic correction and scaling). Let Z_i be a matrix whose columns form an orthonormal basis in $\mathscr{G}_i$ and let $Q_i = [Z_i, W_i]$ be a square orthogonal matrix. Then the above result implies

$$Q_i^{\mathrm{T}} B_i Q_i = \begin{bmatrix} Z_i^{\mathrm{T}} B_i Z_i & 0 \\ 0 & \lambda_i I \end{bmatrix}, \quad Q_i^{\mathrm{T}} g_i = \begin{bmatrix} Z_i^{\mathrm{T}} g_i \\ 0 \end{bmatrix}, \tag{3.18}$$

so that

$$d_i = Z_i \tilde{d}_i, \quad Z_i^{\mathrm{T}} B_i Z_i \tilde{d}_i = -\tilde{g}_i, \quad \tilde{g}_i = Z_i^{\mathrm{T}} g_i. \tag{3.19}$$

In other words, all information concerning variable metric updates is contained in the reduced Hessian approximation $Z_i^{\mathrm{T}} B_i Z_i$ so that the reduced system (3.19) is sufficient for obtaining the direction vector.

This idea can be used for developing limited-memory reduced Hessian variable metric methods. These methods use limited-dimension subspaces $\mathscr{G}_i = \mathrm{span}[g_{i-m+1}, \ldots, g_i]$ and $\mathscr{D}_i = \mathrm{span}[d_{i-m+1}, \ldots, d_i]$ which change on every iteration. Now $\mathscr{D}_i = \mathscr{G}_i$ does not hold in the limited-dimension case, but the quadratic termination property requires columns of Z_i to form a basis in $\mathscr{D}_i$ instead of $\mathscr{G}_i$. Hence the above process has to be slightly reformulated. Instead of Z_i we use an upper triangular matrix T_i such that $D_i = Z_i T_i$ and the reduced Hessian approximation is given in the factorized form $Z_i^{\mathrm{T}} B_i Z_i = R_i^{\mathrm{T}} R_i$ with R_i again upper triangular. Using a scaling parameter γ_1, we can set

$$D_1 = [g_1], \quad T_1 = [\|g_1\|], \quad R_1 = [\sqrt{1/\gamma_1}], \quad \tilde{g}_1 = [\|g_1\|].$$

On every iteration, we first solve two equations $R_i^{\mathrm{T}} R_i \tilde{d}_i = -\tilde{g}_i$, $T_i v_i = \tilde{d}_i$ and set $d_i = D_i v_i$. After determining the direction vector d_i, the line search is performed to obtain a new point $x_{i+1} = x_i + \alpha_i d_i$. Moreover, the matrices D_i and T_i have to be changed to correspond to the subspace $\mathscr{D}_i$. For this purpose, we replace the last column of D_i by d_i and the last column of T_i by $\tilde{d}_i$. Now a representation of the subspace $\mathscr{D}_{i+1}$ has to be formed. First, we project the new gradient $g_{i+1} = g(x_{i+1})$ into the subspace $\mathscr{D}_i$ by solving the equation $T_i^{\mathrm{T}} r_{i+1} = D_i^{\mathrm{T}} g_{i+1}$. Then we determine the quantity $\rho_{i+1} = \|g_{i+1}\| - \|r_{i+1}\|$, set $D_{i+1} = [D_i, g_{i+1}]$ and

$$T_{i+1} = \begin{bmatrix} T_i & r_{i+1} \\ 0 & \rho_{i+1} \end{bmatrix}, \quad \tilde{g}_{i+1} = \begin{bmatrix} r_{i+1} \\ \rho_{i+1} \end{bmatrix}.$$

Using the scaling parameter γ_{i+1}, we obtain a temporary representation of the reduced Hessian approximation in the form $Z_{i+1}^{\mathrm{T}} B_i Z_{i+1} = R_{i+1}^{\mathrm{T}} R_{i+1}$, where

$$R_{i+1} = \begin{bmatrix} R_i & 0 \\ 0 & \sqrt{1/\gamma_{i+1}} \end{bmatrix}, \quad \tilde{g}_{i+1} = \begin{bmatrix} r_{i+1} \\ \rho_{i+1} \end{bmatrix}.$$

This factorization is updated to satisfy the quasi-Newton condition $R_{i+1}^{\mathrm{T}} R_{i+1} \tilde{s}_i = \tilde{y}_i$, where

$$\tilde{s}_i = \alpha_i \begin{bmatrix} \tilde{d}_i \\ 0 \end{bmatrix}, \quad \tilde{y}_i = \tilde{g}_{i+1} - \begin{bmatrix} \tilde{g}_i \\ 0 \end{bmatrix}.$$

Numerically stable methods described in [45] can be used for this purpose. If the subspace $\mathscr{D}_{i+1}$ has dimension $m+1$, then it must be reduced before the new iteration is started. Denote the matrices after such reduction by $\bar{D}_{i+1}$, $\bar{T}_{i+1}$, $\bar{R}_{i+1}$. Then $\bar{D}_{i+1}$ is obtained from D_{i+1} by deleting its first column and matrices $\bar{T}_{i+1}$, $\bar{R}_{i+1}$ are constructed using elementary Givens rotations (see [44]). Notice that the scaling parameters used above have a similar meaning to those in (2.15). Values $\gamma_1 = 1$ and $\gamma_{i+1} = \tilde{s}_i^{\mathrm{T}} \tilde{y}_i / \tilde{y}_i^{\mathrm{T}} \tilde{y}_i$ are recommended.

3.4. Computational experiments

Now, we can present computational experiments with various variable metric methods for large-scale unconstrained optimization. Table 2 compares the sparse-composite update SCVM (3.2), the sparse-fractioned update SFVM (3.5), the sparse-partitioned BFGS update SPVM (3.8), the limited-memory BFGS update in vector form LVVM (3.11), the limited-memory BFGS update in matrix form LMVM (3.16) and the limited memory BFGS update in reduced-Hessian form LRVM (3.19). The limited-memory updates LVVM and LMVM were constructed from 5 previous steps ($m = 5$) and LRVM was constructed from 10 previous steps ($m = 10$) . For further comparison, we introduce results for the sparse discrete Newton method SDNM [17], the truncated Newton method STNM [21] and the nonlinear conjugate gradient method NCGM [37]. Most of the tested methods were implemented in a line-search framework with direct computation of direction vectors (limited-memory methods in the form (1.10), SFVM and SPVM using sparse Cholesky decomposition (1.11)). The sparse composite method SCVM and the truncated Newton method STNM were implemented by using the unpreconditioned inexact conjugate gradient method (1.12) (again with standard line search). The sparse discrete Newton method SDNM was implemented in a trust-region framework by using the optimal procedure (1.21). We have chosen the most suitable implementations for individual methods. Computational experiments were performed on a DIGITAL UNIX workstation using 22 sparse test problems from TEST14 [62] with 1000 variables. The CPU times in Table 2 represent total time for all 22 test problems and are measured in minutes.

From Table 2, it appears that only the SPVM and SDNM methods are worth considering and other variable metric methods are unsuitable for large-scale problems. Indeed, SPVM and SDNM are excellent for general partially separable problems or general problems with sufficiently sparse Hessian matrices (they can be inefficient for ill-conditioned sum of squares as shown in Table 4 below). On the other hand, variable metric methods with limited memory LVVM, LMVM, LRVM, the truncated Newton method STNM and the nonlinear conjugate gradient method NCGM also work well for problems with dense Hessian matrices. Such problems frequently appear in practice. For instance, a product of functions or a squared sum of functions have the same complexity as a sum of functions (3.7) but their Hessian matrices can be completely full. The sparse composite update SCVM is not robust in general. It sometimes fails for difficult problems and generates matrices which are not suitable for sparse Cholesky decomposition (an iterative solution is then required). We review SCVM here, since it gives an excellent tool for improving methods for large sparse sum of squares as demonstrated in Section 4.

4. Variable metric methods for nonlinear least squares

4.1. Basic ideas for using variable metric updates

Suppose that the objective function $F : \mathbb{R}^N \to \mathbb{R}$ has the form

$$F(x) = \tfrac{1}{2} f^{\mathrm{T}}(x) f(x) = \frac{1}{2} \sum_{k=1}^{m} f_k^2(x), \tag{4.1}$$

where $f_k : \mathbb{R}^n \to \mathbb{R}$, $1 \leqslant k \leqslant m$, are twice continuously differentiable functions. This objective function is frequently used for nonlinear regression and for solving systems of nonlinear equations. We can express the gradient and Hessian matrix of (4.1) in the form

$$g(x) = J^{\mathrm{T}}(x)f(x) = \sum_{k=1}^{m} f_k(x)g_k(x), \tag{4.2}$$

$$G(x) = J^{\mathrm{T}}(x)J(x) + C(x) = \sum_{k=1}^{m} g_k(x)g_k^{\mathrm{T}}(x) + \sum_{k=1}^{m} f_k(x)G_k(x), \tag{4.3}$$

where $g_k(x)$ and $G_k(x)$ are the gradients and the Hessian matrices of the functions $f_k : \mathbb{R}^n \to R$, $1 \leqslant k \leqslant m$ and $f^{\mathrm{T}}(x) = [f_1(x), \dots f_m(x)]$, $J^{\mathrm{T}}(x) = [g_1(x), \dots g_m(x)]$. $J(x)$ is the Jacobian matrix of the mapping f at the point x.

The most popular method for nonlinear least squares is the Gauss–Newton method, which uses the first part of (4.3) as an approximation of the Hessian matrix, i.e., $B_i = J_i^{\mathrm{T}}J_i$, $\forall i \in \mathcal{N}$. This method is very efficient for zero-residual problems with $F(x_*) = 0$. In this case, $x_i \to x_*$ implies $F(x_i) \to F(x_*) = 0$ and, therefore, $f_k(x_i) \to 0$ $\forall k, 1 \leqslant k \leqslant m$. If $\|G_k(x)\| \leqslant \bar{G}$, $\forall k, 1 \leqslant k \leqslant m$, then also

$$\|C(x_i)\| = \left\| \sum_{k=1}^{m} f_k(x_i)G_k(x_i) \right\| \leqslant \bar{G} \sum_{k=1}^{m} |f_k(x_i)| \to 0$$

and, therefore, $\|G(x_i) - B_i\| = \|C(x_i)\| \to 0$, which implies a superlinear rate of convergence, see [26]. Since the Jacobian matrices J_i, $i \in \mathcal{N}$, are usually ill-conditioned, even singular, the Gauss–Newton method is most frequently implemented in a trust-region framework.

The Gauss–Newton method is very efficient when applied to a zero-residual problem. It usually outperforms variable metric methods in this case. On the other hand, the rapid convergence can be lost if $F(x_*)$ is large, since $B_i = J_i^{\mathrm{T}}J_i$ can be a bad approximation of G_i. For these reasons, combinations of the Gauss–Newton method with special variable metric updates may be advantageous. Such combined methods exist and can be very efficient, but three problems have to be carefully solved. Firstly, suitable variable metric updates have to be found, together with corresponding quasi-Newton conditions. Secondly, a way for combining these updates with the Gauss–Newton method has to be chosen. Thirdly, a strategy for suppressing the influence of variable metric updates, in case the Gauss–Newton method converges rapidly, has to be proposed. We will investigate these problems in reverse order.

The main idea for suppressing the influence of variable metric updates consists in using the Gauss–Newton method, if it converges rapidly, and variable metric corrections otherwise. The choice of a suitable switching criterion is very important. The most general and, at the same time, most efficient strategy is proposed in [38]. It uses the condition

$$F - F_+ \leqslant \bar{\theta}_1 F, \tag{4.4}$$

where $0 < \bar{\theta}_1 < 1$. If (4.4) holds, then a variable metric correction is applied in the subsequent iteration. Otherwise, the Gauss–Newton method is used. This strategy is based on the fact that $F_{i+1}/F_i \to 0$, if $F_i \to F_* = 0$ superlinearly, and $F_{i+1}/F_i \to 1$, if $F_i \to F_* > 0$.

Now, we describe techniques for combining variable metric updates with the Gauss–Newton method. We consider the following techniques: simple correction, cumulative correction and successive approximation of the second-order term in (4.3). We shall use A to denote a matrix such that $A^{\mathrm{T}}A$ approximates the Hessian matrix $J^{\mathrm{T}}J + C$ (see (2.40)).

A simple correction is useful in the sparse case, when a cumulative correction cannot be realized. On non-Gauss–Newton iterations we compute the matrix B_+ (or A_+) from $J_+^{\mathrm{T}}J_+$ (or J_+) using a variable metric update. Otherwise, we set $B_+ = J_+^{\mathrm{T}}J_+$ (or $A_+ = J_+$).

A cumulative correction is proposed in [38]. On non-Gauss-Newton iterations we compute the matrix B_+ (or A_+) from B (or A) using a variable metric update. Otherwise, we set $B_+ = J_+^{\mathrm{T}}J_+$ (or $A_+ = J_+$).

A successive approximation of the second-order term is based on the model $B = J^{\mathrm{T}}J + C$. The matrix C_+ is computed from the matrix C using variable metric updates. If the Gauss–Newton method should not be used, we set $B_+ = J_+^{\mathrm{T}}J_+ + C_+$. Otherwise, we set $B_+ = J_+^{\mathrm{T}}J_+$. While simple and cumulative corrections use the standard updates described in previous sections, the successive approximation of the second-order term requires special updates (known as structured updates) which we now describe. We will suppose that $\rho = 1$ and $\gamma = 1$ in (2.15). Later we will consider a special scaling technique.

4.2. Structured variable metric updates

There are two possibilities for construction of structured variable metric updates. The first method is based on the transformed quasi-Newton condition $C_+s = z = J_+^{\mathrm{T}}f_+ - J^{\mathrm{T}}f - J_+^{\mathrm{T}}J_+s$. Therefore, the general update has the form (2.15) with B and y replaced by C and z, respectively. The SR1 update, derived in this way, can be written in the form

$$C_+ = C + \frac{(z - Cs)(z - Cs)^{\mathrm{T}}}{s^T(z - Cs)}. \tag{4.5}$$

This SR1 update is very efficient and usually outperforms other structured variable metric updates. Notice that the BFGS method cannot be realized in this approach since positivity of $s^{\mathrm{T}}z$ is not guaranteed.

The second possibility involves updating $\bar{B} = J_+^{\mathrm{T}}J_+ + C$ to obtain $B_+ = J_+^{\mathrm{T}}J_+ + C_+$ satisfying the quasi-Newton condition $B_+s = y = J_+^{\mathrm{T}}f_+ - J^{\mathrm{T}}f$. The resulting general update has the form (2.15) with B replaced by $\bar{B}$. Since $y - \bar{B}s = z - Cs$, it is advantageous to use formula (2.29). Then

$$\begin{aligned} C_+ &= C + \frac{(y - \bar{B}s)v^{\mathrm{T}} + v(y - \bar{B}s)^{\mathrm{T}}}{s^{\mathrm{T}}v} - \frac{(y - \bar{B}s)^{\mathrm{T}}s}{s^{\mathrm{T}}v}\frac{vv^{\mathrm{T}}}{s^{\mathrm{T}}v} \\ &= C + \frac{(z - Cs)v^{\mathrm{T}} + v(z - Cs)^{\mathrm{T}}}{s^{\mathrm{T}}v} - \frac{(z - Cs)^{\mathrm{T}}s}{s^{\mathrm{T}}v}\frac{vv^{\mathrm{T}}}{s^{\mathrm{T}}v} \end{aligned} \tag{4.6}$$

with $v=s$ for the structured PSB update, $v=y$ for the structured DFP update and $v=y+(y^{\mathrm{T}}s/s^{\mathrm{T}}\bar{B}s)^{1/2}\bar{B}s$ for the structured BFGS update. Methods based on formula (4.6) have been investigated in [24], where superlinear convergence of the structured BFGS method was proved.

The vectors y and z, used in formulae (4.5)–(4.6), can be defined in various ways, but always based on $z = y - J_+^{\mathrm{T}}J_+s$. The standard choice

$$z = J_+^{\mathrm{T}}f_+ - J^{\mathrm{T}}f - J_+^{\mathrm{T}}J_+s, \tag{4.7}$$

corresponding to the quasi-Newton condition $(J_+^{\mathrm{T}}J_+ + C_+)s = J_+^{\mathrm{T}}f_+ - J^{\mathrm{T}}f$, is introduced in [22]. In [6], a similar choice

$$z = J_+^{\mathrm{T}}f_+ - J^{\mathrm{T}}f - J^{\mathrm{T}}Js \tag{4.8}$$

corresponding to the quasi-Newton condition $(J^{\mathrm{T}}J + C_+)s = J_+^{\mathrm{T}}f_+ - J^{\mathrm{T}}f$, is given. Another choice [4,83]) is based on the objective function $\tilde{F}(x) = (1/2)(f^{\mathrm{T}}(x)f(x) - x^{\mathrm{T}}J^{\mathrm{T}}Jx)$, whose Hessian matrix is just the matrix $J^{\mathrm{T}}J$ that we want to approximate. Applying the standard variable metric method to the function $\tilde{F}$, we obtain the quasi-Newton condition $C_+s = \tilde{g}_+ - \tilde{g} = z$, where

$$z = J_+^{\mathrm{T}}f_+ - J_+^{\mathrm{T}}J_+x_+ - J^{\mathrm{T}}f + J^{\mathrm{T}}Jx. \tag{4.9}$$

A popular choice, proposed in [9], is based on the explicit form of the second-order term in (4.3). Suppose that the approximations B_k^+ of the Hessian matrices G_k satisfy the quasi-Newton conditions $B_k^+s = g_k^+ - g_k$, $1 \leqslant k \leqslant m$. Then we can write

$$z = C_+s \triangleq \sum_{k=1}^{m} f_k^+ B_k^+ s = \sum_{k=1}^{m} f_k^+(g_k^+ - g_k) = (J_+ - J)^{\mathrm{T}}f_+. \tag{4.10}$$

Interesting methods for nonlinear least squares have been obtained from the product-form BFGS update (2.40) (other product-form updates are less suitable since they require the inversion of the matrix $A^{\mathrm{T}}A$). A generalization of (2.40) (with $\rho = 1$ and $\gamma = 1$), related to structured update (4.6), is described in [97]. Here A is replaced by the matrix $J + L$, where J is the Jacobian matrix and L plays a similar role to C in (4.6). Thus $B = (J + L)^{\mathrm{T}}(J + L)$, $B_+ = (J_+ + L_+)^{\mathrm{T}}(J_+ + L_+)$ and if we set $\bar{B} = (J_+ + L)^{\mathrm{T}}(J_+ + L)$, we can express (4.6) as

$$L_+ = L + \frac{(J_+ + L)s}{s^{\mathrm{T}}\bar{B}s}\left(\sqrt{\frac{s^{\mathrm{T}}\bar{B}s}{s^{\mathrm{T}}y}}\,y - \bar{B}s\right)^{\mathrm{T}}, \tag{4.11}$$

which is similar to (2.40).

Structured variable metric updates can be improved by a suitable scaling technique. The main reason for scaling is controlling the size of the matrix C. Therefore, the quasi-Newton condition $C_+s = z$ is preferred. The scaling parameter γ is chosen in such a way that $(1/\gamma)Cs$ is close to z in some sense. In analogy with (2.9), we can choose $\gamma = s^{\mathrm{T}}Cs/s^{\mathrm{T}}z$ or $\gamma = \max(s^{\mathrm{T}}Cs/s^{\mathrm{T}}z, 1)$, which is the value proposed in [23]. Biggs [9] recommends the value $\gamma = f^{\mathrm{T}}f/f_+^{\mathrm{T}}f$ based on a quadratic model. If we choose the scaling parameter γ, then we replace C by $(1/\gamma)C$ in (4.5)–(4.6) to obtain a scaled structured update. A more complicated process, described in [97], is used in connection with product form update (4.11). All the above methods can be realized efficiently using switching strategy (4.4). Structured variable metric updates can also be used permanently (without switching), as follows from the theory given in [24], but such a realization is usually less efficient.

Interesting variable metric updates are based on an approximation of the term

$$T(x) = \sum_{k=1}^{m} \frac{f_k(x)}{\|f(x)\|} G_k(x).$$

Table 3

Line-search realization	Iterations	f. eval.	g. eval.	CPU time	Failures
Scaled BFGS	4229	5301	5301	1.43	1
Standard GN	4809	8748	13 555	3.46	7
GN with (4.5) and (4.4)	1447	2546	3993	1.37	—
GN with (4.13) and (4.4)	1594	2807	4400	1.32	—
GN with (2.17) and (4.4)	1658	2805	4461	1.15	—
Trust-region realization	Iterations	f. eval.	g. eval.	CPU time	Failures
Standard GN	2114	2512	2194	1.31	—
GN with (4.5) and (4.4)	1497	1777	1579	1.05	—
GN with (4.13) and (4.4)	1480	1753	1562	1.04	—
GN with (2.17) and (4.4)	1476	1846	1555	0.99	—

Thus we have the model $B = J^TJ + \|f\|T$. By analogy with structured variable metric methods, Huschens [54] proposed totally structured variable metric methods which consist in updating the matrix $\bar{B} = J_+^TJ_+ + \|f\|T$ to get the matrix $\tilde{B}_+ = J_+^TJ_+ + \|f\|T_+$, satisfying the quasi-Newton condition $\tilde{B}_+s = y$. Finally, the matrix $B_+ = J_+^TJ_+ + \|f_+\|T_+$ is chosen. Using expression (2.27), we can write

$$T_+ = T + \frac{1}{\|f\|}\left(\frac{(y-\bar{B}s)v^T + v(y-\bar{B}s)^T}{s^Tv} - \frac{(y-\bar{B}s)^Ts}{s^Tv}\frac{vv^T}{s^Tv}\right)$$

$$= T + \frac{(\tilde{z}-Ts)v^T + v(\tilde{z}-Ts)^T}{s^Tv} - \frac{(\tilde{z}-Ts)^Ts}{s^Tv}\frac{vv^T}{s^Tv}, \tag{4.12}$$

where $\tilde{z} = z/\|f\| = (y - J_+^TJ_+s)/\|f\|$. Setting $v = s$, we get the totally structured PSB method. Setting $v = y$, we get the totally structured DFP method. Setting $v = y + (y^Ts/s^T\bar{B}s)^{1/2}\bar{B}s$, we get the totally structured BFGS method. The totally structured SR1 method has the form

$$T_+ = T + \frac{(\tilde{z}-Ts)(\tilde{z}-Ts)^T}{s^T(\tilde{z}-Ts)}. \tag{4.13}$$

The use of $\|f\|$ instead of $\|f_+\|$ in the quasi-Newton condition $(J_+^TJ_+ + \|f\|T_+)s = y$ leads to methods which have a quadratic rate of convergence in the case of zero-residual problems and a superlinear rate of convergence otherwise, see [54]. This is the most significant theoretical result concerning permanent realization of structured variable metric updates.

We now present numerical experiments with various methods for nonlinear least squares. Table 3 compares the BFGS method with interval scaling (2.10) and nonquadratic correction (2.4), the standard Gauss–Newton method, the Gauss–Newton method with structured SR1 update (4.5) and switching strategy (4.4), the Gauss–Newton method with totally structured SR1 update (4.13) and switching strategy (4.4) and the Gauss–Newton method with the cumulative BFGS correction (2.17) and switching strategy (4.4). The first part of Table 3 refers to the standard line-search implementation and the second part refers to the trust-region implementation (1.22). Structured updates (4.5) and (4.13) were scaled in each iteration as in [23]. The cumulative BFGS update was scaled only on the first iteration. Computational experiments have been performed on a PENTIUM PC computer using 82 test problems (30 from [65], 22 from TEST15, 30 from TEST18, [62]) with 20 variables (62 of them have zero residual at the solution). The CPU times in Table 3 represent total time for all 82 test problems and are measured in seconds.

Results in Table 3 suggest that trust region realizations are preferable whenever the matrix $B_i = J_i^{\mathrm{T}} J_i$ is used (this matrix is usually ill-conditioned). Furthermore, they show the efficiency of switching strategy (4.4). Structured updates were also tested without switching but results obtained were much worse. The efficiency of scaled BFGS method with line-search confirms its robustness for nonlinear least squares (CPU time is low since $\mathrm{O}(n^2)$ operations per iteration are used).

4.3. Variable metric updates for sparse least squares

The Gauss–Newton method can also be combined with variable metric updates in the sparse case. We will now describe some such possibilities. One is a combination of the Gauss–Newton method with the composite update (3.2), so that

$$B_+ = \begin{cases} \mathscr{P}_S \mathscr{P}_{QG}(J_+^{\mathrm{T}} J_+) & \text{if } F - F_+ \leqslant \bar{\theta}_1 F, \\ J_+^{\mathrm{T}} J_+ & \text{if } F - F_+ > \bar{\theta}_1 F. \end{cases} \tag{4.14}$$

Computational efficiency of this hybrid method was studied in [59].

An interesting approach, based on the partitioned SR1 update, was proposed in [93] and also studied in [59]. The partitioned SR1 update is applied to the approximations $\hat{T}_k$ of the packed element-Hessian matrices $\hat{G}_k(x)$ of the functions $f_k : \mathbb{R}^n \to \mathbb{R}$, $1 \leqslant k \leqslant m$, contained in (4.1). These matrices are updated in such a way that

$$\hat{T}_k^+ = \begin{cases} \hat{T}_k + \dfrac{(\hat{y}_k - \hat{T}_k \hat{s}_k)(\hat{y}_k - \hat{T}_k \hat{s}_k)^{\mathrm{T}}}{\hat{s}_k^{\mathrm{T}}(\hat{y}_k - \hat{T}_k \hat{s}_k)} & \text{if } |\hat{s}_k^{\mathrm{T}}(\hat{y}_k - \hat{T}_k \hat{s}_k)| > \bar{\theta}_0, \\ \hat{T}_k & \text{if } |\hat{s}_k^{\mathrm{T}}(\hat{y}_k - \hat{T}_k \hat{s}_k)| \leqslant \bar{\theta}_0 \end{cases} \tag{4.15}$$

and are used for construction of approximations $\hat{B}_k$ of the packed element-Hessian matrices $\hat{g}_k \hat{g}_k^{\mathrm{T}} + f_k \hat{G}_k$. Using (4.4), we can write

$$\hat{B}_k^+ = \hat{g}_k^+ (\hat{g}_k^+)^{\mathrm{T}} + f_k^+ \hat{T}_k^+ \quad \text{if } F - F_+ \leqslant \bar{\theta}_1 F, \tag{4.16}$$

$$\hat{B}_k^+ = \hat{g}_k^+ (\hat{g}_k^+)^{\mathrm{T}} \quad \text{if } F - F_+ > \bar{\theta}_1 F. \tag{4.17}$$

In the first iteration we set $\hat{T}_k = I$, $1 \leqslant k \leqslant m$. Notice that the matrices $\hat{T}_k^+$, $1 \leqslant k \leqslant m$, have to be stored simultaneously, which is a disadvantage of this method.

Another interesting way for improving the sparse Gauss–Newton method is based on the factorized formula (4.11), which is used as a simple update so that $L = 0$. Taking $L = 0$ in (5.11), we can express $A_+ = J_+ + L_+$ in the form

$$\begin{aligned} A_+ &= J_+ + \frac{J_+ s}{s^{\mathrm{T}} J_+^{\mathrm{T}} J_+ s} \left(\sqrt{\frac{s^{\mathrm{T}} J_+^{\mathrm{T}} J_+ s}{s^{\mathrm{T}} y}}\, y - J_+^{\mathrm{T}} J_+ s \right)^{\mathrm{T}} \\ &= J_+ + \frac{J^+ s}{\|J^+ s\|} \left(\frac{y}{\sqrt{s^{\mathrm{T}} y}} - J_+^{\mathrm{T}} \frac{J^+ s}{\|J^+ s\|} \right)^{\mathrm{T}}. \end{aligned} \tag{4.18}$$

Table 4

Method	Iterations	f. eval.	g. eval.	CPU time	Failures
GN	11 350	11760	11 402	3 : 49	2
GNCVM	7264	7688	7316	2 : 36	—
GNPVM	8562	9588	8614	3 : 48	1
GNDNM	7012	7604	9286	2 : 35	—
SPVM	14 009	29 161	29 161	4 : 59	3
SDNM	12 588	84 484	84 337	8 : 38	4

Then we can use the matrix (4.18) if $F - F_+ \leqslant \bar{\theta}_1 F$, and set $A_+ = J_+$, otherwise (see [59] for more detail).

An interesting sparse hybrid method is based on the SR1 update. Consider the augmented linear least-squares problem $\tilde{J}_+ d_+ \approx -\tilde{f}_+$ where

$$\tilde{J}_+ = \begin{bmatrix} J_+ \\ w \end{bmatrix}, \qquad \tilde{f}_+ = \begin{bmatrix} f_+ \\ 0 \end{bmatrix}. \tag{4.19}$$

The normal equations for this problem have the form $B_+ d_+ = -J_+^{\mathrm{T}} f_+$, where

$$B^+ = \tilde{J}_+^{\mathrm{T}} \tilde{J}_+ = J_+^{\mathrm{T}} J_+ + w w^{\mathrm{T}}. \tag{4.20}$$

If we choose

$$w = (y - J_+^{\mathrm{T}} J_+ s) / \sqrt{s^{\mathrm{T}}(y - J_+^{\mathrm{T}} J_+ s)}, \tag{4.21}$$

then (4.20) gives exactly the SR1 update (with B replaced by $J_+^{\mathrm{T}} J_+$). Note that (4.21) can be used only if $s^{\mathrm{T}}(y - J_+^{\mathrm{T}} J_+ s) > 0$, which slightly restricts the use of update (4.19). We use the augmented linear least-squares problem $\tilde{J}_+ d_+ \approx -\tilde{f}_+$ (with w given by (4.21)), if $F - F_+ \leqslant \bar{\theta}_1 F$ and $s^{\mathrm{T}}(y - J_+^{\mathrm{T}} J_+ s) > \bar{\theta}_0$ hold simultaneously, and the standard linear least-squares problem $J_+ d \approx -f_+$, otherwise.

Table 4 compares the standard Gauss–Newton method GN, the Gauss–Newton method with composite update GNCVM (4.14) and the Gauss–Newton method with partitioned update GNPVM (4.15)–(4.16). For further comparison, we quote results for the combined Gauss–Newton and discrete Newton method GNDNM, utilizing switching strategy (4.4) and also for the partitioned BFGS method SPVM (3.8) and the sparse discrete Newton method SDNM. All these methods have been implemented within a trust-region strategy (1.21), see [66]. Computational experiments were performed on a DIGITAL UNIX workstation using 52 sparse test problems (22 from TEST15, 30 from TEST18, [62]) with 1000 variables (38 of them have zero residual at the solution). The CPU times in Table 4 represent the total for all 52 test problems and are quoted in minutes. Sparse and limited-memory variable metric methods have not been efficient for solving these problems.

Table 4 implies that special methods for least-squares problems are usually more efficient than methods for general problems. This conclusion also holds for other classes of problems. For instance, the last 30 problems used in Table 4 are solutions to systems of nonlinear equations, which can also be solved more efficiently by special methods. Inefficiency of SDNM was mainly caused by four failures (3000 iterations or 5000 function evaluations did not suffice). But SDNM did not outperform combined methods even if difficult problems were excluded.

5. Conclusion

In this paper, we have given a review of variable metric or quasi-Newton methods for unconstrained optimization, paying particular attention to the derivation of formulas and their efficient implementation (we have tried to quote all relevant literature). Quasi-Newton methods can be also used for solving systems of nonlinear equations, see, e.g., [10,28,64], but theoretical investigation and practical realization require a slightly different point of view. Another field for application of variable metric methods is general constrained optimization. Nevertheless, problems connected with potential functions, constraint handling or interior point approach are dominant in this case and go beyond the scope of this contribution.

Numerical experience, partially reported in this paper, gives implications for the choice of a suitable optimization method. We would like to give few recommendations for potential users. Standard variable metric methods described in Section 2 are mostly suitable for dense small or moderate-size general problems (up to 100–200 variables, say). Reasonable scaling and nonquadratic correction can improve the efficiency of these methods.

If we have a large-scale problem, then the choice of method depends on the problem structure. General problems with sparse Hessian matrices are successfully solved by the discrete Newton method. Partially separable problems can be efficiently solved by partitioned variable metric updates. If the Hessian matrix has no structure, then limited memory variable metric methods as well as the truncated Newton method and the nonlinear conjugate gradient method are suitable.

If the objective function is a sum of squares, then special methods for least squares should be used. Trust region realizations are most suitable in this case. We recommend the Gauss–Newton method with variable metric corrections. The switching strategy (4.4) is very efficient. If the problem is dense then the cumulative BFGS update is of a primary interest. The simple composite update (4.14) is suitable in the sparse case.

Variable metric methods can be successfully adapted to solve nondifferentiable problems. An efficient variable metric method for nonsmooth optimization is proposed in [63].

References

[1] J. Abaffy, E. Spedicato, ABS Projection Algorithms, Mathematical Techniques for Linear and Nonlinear Equations, Ellis Horwood, Chichester, 1989.

[2] N. Adachi, On variable metric algorithm, J. Optim. Theory Appl. 7 (1971) 391–409.

[3] M. Al-Baali, Highly efficient Broyden methods of minimization with variable parameter, Optim. Methods Software 1 (1992) 301–310.

[4] M. Al-Baali, R. Fletcher, Variational methods for nonlinear least squares, J. Optim. Theory Appl. 36 (1985) 405–421.

[5] O. Axelsson, Iterative Solution Methods, Cambridge University Press, Cambridge, 1996.

[6] J.T. Betts, Solving the nonlinear least squares problem: application of a general method, J. Optim. Theory Appl. 18 (1976) 469–483.

[7] M.C. Biggs, Minimization algorithms making use of nonquadratic properties of the objective function, J. Inst. Math. Appl. 8 (1971) 315–327.

[8] M.C. Biggs, A note on minimization algorithms which make use of non-quadratic properties of the objective function, J. Inst. Math. Appl. 12 (1973) 337–338.

[9] M.C. Biggs, The estimation of the Hessian matrix in nonlinear least squares problems with nonzero residuals, Math. Programming 12 (1977) 67–80.

[10] C.G. Broyden, A class of methods for solving nonlinear simultaneous equations, Math. Comp. 19 (1965) 577–593.

[11] C.G. Broyden, The convergence of a class of double rank minimization algorithms, Part 1 – general considerations, Part 2 – the new algorithm, J. Inst. Math. Appl. 6 (1970) 76–90, 222–231.

[12] A. Buckley, A combined conjugate-gradient quasi-Newton minimization algorithm, Math. Programming 15 (1978) 200–210.

[13] A. Buckley, A. LeNir, QN-like variable storage conjugate gradients, Math. Programming 27 (1983) 155–175.

[14] R.H. Byrd, D.C. Liu, J. Nocedal, On the behavior of Broyden's class of quasi-Newton methods, Report No. NAM 01, Dept. of Electrical Engn. and Computer Science, Northwestern University, Evanston, 1990.

[15] R.H. Byrd, J. Nocedal, R.B. Schnabel, Representation of quasi-Newton matrices and their use in limited memory methods, Math. Programming 63 (1994) 129–156.

[16] R.H. Byrd, J. Nocedal, Y.X. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, SIAM J. Numer. Anal. 24 (1987) 1171–1190.

[17] M. Coleman, J.J. Moré, Estimation of sparse Hessian matrices and graph coloring problems, Math. Programming 42 (1988) 245–270.

[18] M. Contreras, R.A. Tapia, Sizing the BFGS and DFP updates: a numerical study, J. Optim. Theory Appl. 78 (1993) 93–108.

[19] W.C. Davidon, Variable metric method for minimisation, A.E.C. Research and Development Report ANL-5990, 1959.

[20] W.C. Davidon, Optimally conditioned optimization algorithms without line searches, Math. Programming 9 (1975) 1–30.

[21] R.S. Dembo, T. Steihaug, Truncated-Newton algorithms for large-scale unconstrained minimization, Math. Programming 26 (1983) 190–212.

[22] J.E. Dennis, Some computational techniques for the nonlinear least squares problem, in: G.D. Byrne, C.A. Hall (Eds.), Numerical Solution of Nonlinear Algebraic Equations, Academic Press, London, 1974.

[23] J.E. Dennis, D. Gay, R.E. Welsch, An adaptive nonlinear least squares algorithm, ACM Trans. Math. Software 7 (1981) 348–368.

[24] J.E. Dennis, H.J. Martinez, R.A. Tapia, Convergence theory for the structured BFGS secant method with application to nonlinear least squares, J. Optim. Theory Appl. 61 (1989) 161–177.

[25] J.E. Dennis, H.H.W. Mei, An unconstrained optimization algorithm which uses function and gradient values, Report No. TR-75-246, Dept. of Computer Science, Cornell University, 1975.

[26] J.E. Dennis, J.J. Moré, A characterization of superlinear convergence and its application to quasi-Newton methods, Math. Comp. 28 (1974) 549–560.

[27] J.E. Dennis, R.B. Schnabel, Least change secant updates for quasi-Newton methods, Report No. TR78-344, Dept. of Computer Sci., Cornell University, Ithaca, 1978.

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

[29] J.E. Dennis, R.B. Schnabel, A view of unconstrained optimization, in: G.L. Nemhauser, A.H.G. Rinnooy Kan, M.J. Todd (Eds.), Optimization, North-Holland, Amsterdam, 1989.

[30] L.C.W. Dixon, Quasi-Newton algorithms generate identical points, Math. Programming 2 (1972) 383–387.

[31] R. Fletcher, A new approach to variable metric algorithms, Comput. J. 13 (1970) 317–322.

[32] R. Fletcher, Practical Methods of Optimization, Wiley, New York, 1987.

[33] R. Fletcher, Low storage methods for unconstrained optimization, in: E.L. Algower, K. Georg (Eds.), Computational Solution of Nonlinear Systems of Equations, Lectures in Applied Mathematics, Vol. 26, AMS Publications, Providence, RI, 1990.

[34] R. Fletcher, A new variational result for quasi-Newton formulae, SIAM J. Optim. 1 (1991) 18–21.

[35] R. Fletcher, An optimal positive definite update for sparse Hessian matrices, SIAM J. Optim. 5 (1995) 192–218.

[36] R. Fletcher, M.J.D. Powell, A rapidly convergent descent method for minimization, Comput. J. 6 (1963) 163–168.

[37] R. Fletcher, C.M. Reeves, Function minimization by conjugate gradients, Comput. J. 7 (1964) 149–154.

[38] R. Fletcher, C. Xu, Hybrid methods for nonlinear least squares, IMA J. Numer. Anal. 7 (1987) 371–389.

[39] J.A. Ford, I.A. Moghrabi, Alternative parameter choices for multi-step quasi-Newton methods, Optim. Methods Software 2 (1993) 357–370.

[40] J.A. Ford, I.A. Moghrabi, Multi-step quasi-Newton methods for optimization, J. Comput. Appl. Math. 50 (1994) 305–323.

[41] J.A. Ford, I.A. Moghrabi, Minimum curvature multi-step quasi-Newton methods for unconstrained optimization, Report No. CSM-201, Department of Computer Science, University of Essex, Colchester, 1995.

[42] J. Greenstadt, Variations on variable metric methods, Math. Comput. 24 (1970) 1–18.

[43] J.C. Gilbert, C. Lemarechal, Some numerical experiments with variable-storage quasi-Newton algorithms, Math. Programming 45 (1989) 407–435.

[44] P.E. Gill, M.W. Leonard, Limited-memory reduced-Hessian methods for large-scale unconstrained optimization, Report NA 97-1, Dept. of Mathematics, University of California, San Diego, La Jolla, 1997.

[45] P.E. Gill, W. Murray, M.A. Saunders, Methods for computing and modifying LDV factors of a matrix, Math. Comput. 29 (1975) 1051–1077.

[46] D. Goldfarb, A family of variable metric algorithms derived by variational means, Math. Comput. 24 (1970) 23–26.

[47] A. Griewank, The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients, Math. Programming 50 (1991) 141–175.

[48] A. Griewank, P.L. Toint, Partitioned variable metric updates for large-scale structured optimization problems, Numer. Math. 39 (1982) 119–137.

[49] A. Griewank, P.L. Toint, Local convergence analysis for partitioned quasi-Newton updates, Numer. Math. 39 (1982) 429–448.

[50] A. Griewank, P.L. Toint, Numerical experiments with partially separable optimization problems, in: D.F. Griffits (Ed.), Numerical Analysis, Proc. Dundee 1983, Lecture Notes in Mathematics, Vol. 1066, Springer, Berlin, 1984, pp. 203–220.

[51] M.R. Hestenes, C.M. Stiefel, Methods of conjugate gradient for solving linear systems, J. Res. NBS 49 (1964) 409–436.

[52] S. Hoshino, A formulation of variable metric methods, J. Inst. Math. Appl. 10 (1972) 394–403.

[53] H.Y. Huang, Unified approach to quadratically convergent algorithms for function minimization, J. Optim. Theory Appl. 5 (1970) 405–423.

[54] J. Huschens, On the use of product structure in secant methods for nonlinear least squares, SIAM J. Optim. 4 (1994) 108–129.

[55] D.C. Liu, J. Nocedal, On the limited memory BFGS method for large-scale optimization, Math. Programming 45 (1989) 503–528.

[56] L. Lukšan, Computational experience with improved variable metric methods for unconstrained minimization, Kybernetika 26 (1990) 415–431.

[57] L. Lukšan, Computational experience with known variable metric updates, J. Optim. Theory Appl. 83 (1994) 27–47.

[58] L. Lukšan, J. Vlček, Simple scaling for variable metric updates, Report No. 611, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague 1995.

[59] L. Lukšan, Hybrid methods for large sparse nonlinear least squares, J. Optim. Theory Appl. 89 (1996) 575–595.

[60] L. Lukšan, Numerical methods for unconstrained optimization, Report DMSIA 12/97, University of Bergamo, 1997.

[61] L. Lukšan, M. Tůma, M. Šiška, J. Vlček, N. Ramešová, Interactive system for universal functional optimization (UFO) — Version 1998, Report No. 766, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, 1998.

[62] L. Lukšan, J.Vlček, Subroutines for testing large sparse and partially separable unconstrained and equality constrained optimization problems, Report No. 767, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, 1999.

[63] L. Lukšan, J. Vlček, Globally convergent variable metric method for convex nonsmooth unconstrained minimization, J. Optim. Theory Appl. 102 (1999) 593–613.

[64] J.M. Martinez, A quasi-Newton method with modification of one column per iteration, Computing 33 (1984) 353–362.

[65] J.J. Moré, B.S. Garbow, K.E. Hillstrom, Testing unconstrained optimization software, ACM Trans. Math. Software 7 (1981) 17–41.

[66] J.J. Moré, D.C. Sorensen, Computing a trust region step, Report ANL-81-83, Argonne National Laboratory, 1981.

[67] J. Nocedal, Updating quasi-Newton matrices with limited storage, Math. Comp. 35 (1980) 773–782.

[68] J. Nocedal, Y. Yuan, Analysis of a self-scaling quasi-Newton method, Math. Programming 61 (1993) 19–37.

[69] S.S. Oren, D.G. Luenberger, Self scaling variable metric (SSVM) algorithms, Part 1 — criteria and sufficient condition for scaling a class of algorithms, Part 2 — implementation and experiments, Management Sci. 20 (1974) 845–862, 863–874.

[70] S.S. Oren, E. Spedicato, Optimal conditioning of self scaling variable metric algorithms, Math. Programming 10 (1976) 70–90.

[71] M.R. Osborne, L.P. Sun, A new approach to the symmetric rank-one updating algorithm. Report No. NMO/01, School of Mathematics, Australian National University, Canberra, 1988.

[72] A. Perry, A modified conjugate gradient algorithm, Oper. Res. 26 (1978) 1073–1078.

[73] E. Polak, G. Ribiére, Note sur la convergence des méthodes de directions conjugées, Rev. Française Inform. Mech. Oper. 16-R1 (1969) 35–43.

[74] M.J.D. Powell, A new algorithm for unconstrained optimization, in: J.B. Rosen, O.L. Mangasarian, K. Ritter (Eds.), Nonlinear Programming, Academic Press, London, 1970.

[75] M.J.D. Powell, On the global convergence of trust region algorithms for unconstrained minimization, Math. Programming 29 (1984) 297–303.

[76] D.G. Pu, W.W. Tian, A class of modified Broyden algorithms, J. Comput. Math. 12 (1994) 366–379.

[77] D.F. Shanno, Conditioning of quasi-Newton methods for function minimization, Math. Comp. 24 (1970) 647–656.

[78] D.F. Shanno, K.J. Phua, Matrix conditioning and nonlinear optimization, Math. Programming 14 (1978) 144–160.

[79] D. Siegel, Implementing and modifying Broyden class updates for large scale optimization, Report DAMTP NA12, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, 1992.

[80] E. Spedicato, Stability of Huang's update for the conjugate gradient method, J. Optim. Theory Appl. 11 (1973) 469–479.

[81] E. Spedicato, On condition numbers of matrices in rank-two minimization algorithms, in: L.C.W. Dixon, G.P. Szego (Eds.), Towards Global Optimization, North-Holland, Amsterdam, 1975.

[82] E. Spedicato, A bound on the condition number of rank-two corrections and applications to the variable metric method, Calcolo 12 (1975) 185–199.

[83] E. Spedicato, Parameter estimation and least squares, in: F. Archetti, M. Cugiani (Eds.), Numerical Techniques for Stochastic Systems, North-Holland, Amsterdam, 1980.

[84] E. Spedicato, A class of rank-one positive definite quasi-Newton updates for unconstrained minimization, Math. Operationsforsch. Statist. Ser. Optim. 14 (1983) 61–70.

[85] E. Spedicato, A class of sparse symmetric quasi-Newton updates, Ricerca Operativa 22 (1992) 63–70.

[86] E. Spedicato, N.Y. Deng, Z. Li, On sparse quasi-Newton quasi-diagonally dominant updates, Report DMSIA 1/96, University of Bergamo, 1996.

[87] E. Spedicato, Z. Xia, Finding general solutions of the quasi-Newton equation via the ABS approach, Optim. Methods Software 1 (1992) 243–252.

[88] E. Spedicato, J. Zhao, Explicit general solution of the Quasi-Newton equation with sparsity and symmetry, Optim. Methods Software 2 (1993) 311–319.

[89] T. Steihaug, Local and superlinear convergence for truncated iterated projections methods, Math. Programming 27 (1983) 176–190.

[90] T. Steihaug, The conjugate gradient method and trust regions in large-scale optimization, SIAM J. Numer. Anal. 20 (1983) 626–637.

[91] P.L. Toint, On sparse and symmetric matrix updating subject to a linear equation, Math. Comp. 31 (1977) 954–961.

[92] P.L. Toint, Global convergence of the partitioned BFGS algorithm for convex partially separable optimization, Math. Programming 36 (1986) 290–306.

[93] P.L. Toint, On large scale nonlinear least squares calculations, SIAM J. Sci. Statis. Comput. 8 (1987) 416–435.

[94] M. Tůma, Sparse fractioned variable metric updates, Report No. 497, Institute of Computer and Information Sciences, Czechoslovak Academy of Sciences, Prague 1991.

[95] H. Wolkowicz, Measures for rank-one updates, Math. Oper. Res. 19 (1994) 815–830.

[96] H. Wolkowicz, An all-inclusive efficient region of updates for least change secant methods, SIAM J. Optim. 5 (1996) 172–191.

[97] H. Yabe, T. Takahashi, Factorized quasi-Newton methods for nonlinear least squares problems, Math. Programming 51 (1991) 75–100.

[98] Y. Yuan, Non-quasi-Newton updates for unconstrained optimization, J. Comput. Math. 13 (1995) 95–107.

[99] J. Zhang, N.Y. Deng, L. Chen, A new quasi-Newton equation and related methods for unconstrained optimization, Report MA-96-05, City University of Hong Kong, 1996.