VARIABLE METRIC METHODS FOR UNCONSTRAINED OPTIMIZATION [1]

Ladislav Lukšan [2] and Emilio Spedicato [3]

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 medium size dense problems, their modifications based either on sparse or partitioned or limited storage updates are very efficient in the case of large-scale sparse problems. Very significant applications of these methods also appear in nonlinear least squares approximation an nonsmooth optimization. In this contribution, we give an extensive review of variable metric methods and their use in various optimization fields.

Key words. Quasi-Newton methods, variable metric methods, unconstrained optimization, nonlinear least squares, nonsmooth optimization, sparse problems, limited storage methods.

1. Introduction. This contribution reviews the efficient class of methods known as variable metric methods or quasi-Newton methods for the local unconstrained optimization, i.e. for finding a point $x_* \in \mathcal{R}^n$ such that $F(x_*) = \min_{x \in \mathcal{R}^n} F(x)$ (we consider only local minima). Here $F : \mathcal{R}^n \to \mathcal{R}$ is a twice continuously differentiable objective function and $\mathcal{R}^n$ is an n-dimensional vector space. Basic methods for solving this problem are iterative. Starting with an initial point $x_1 \in \mathcal{R}^n$, they generate a sequence $x_i \in \mathcal{R}^n$, $i \in \mathcal{N}$ by the simple process

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

where $d_i \in \mathcal{R}^n$ is a direction vector and $\alpha_i \geq 0$ is a scalar, the stepsize ($\mathcal{N}$ is the set of natural numbers). The most efficient optimization methods of the form (1.1)) belong to three basic classes, which are the modified Newton methods, the variable metric methods and the conjugate gradient methods.

The modified Newton methods are based on a local quadratic model

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

where $G_i = G(x_i)$ and $g_i = g(x_i)$ are respectively the Hessian matrix and the gradient of the objective function $F : \mathcal{R}^N \to \mathcal{R}$ at the point $x_i \in \mathcal{R}^n$. The direction vector $d_i \in \mathcal{R}^n$, $i \in \mathcal{N}$, is chosen to minimize $Q_i(d)$ on $\mathcal{R}^n$ or on some subset of $\mathcal{R}^n$. Modified Newton methods converge fast, if they converge, but they have some disadvantages. Minimization of $Q_i(d)$ requires $O(n^3)$ operations and computation of second derivatives can be difficult and time consuming. Moreover, if the Hessian matrices are not positive definite, then a simple implementation of the modified Newton method may not be globally convergent.

Variable metric methods are based on the local quadratic model

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

[1] Work partly supported by grant GA ČR No. 201/96/0918, by CNR–GNIM visiting professorship program and by CNR CT11 1995 grant

[2] Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic (luksan@uivt.cas.cz)

[3] Department of Mathematics, University of Bergamo, Bergamo 24129, Italy (emilio@ibguniv.unibg.it)

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, satisfies the quasi-Newton condition $B_{i+1}(x_{i+1} - x_i) = g_{i+1} - g_i$ and is as close as possible to B_i. Variable metric methods have some advantages against modified Newton methods. Matrices B_i are positive definite so that variable metric methods can be forced to be globally convergent. Moreover, we can update the inverse $H_i = B_i^{-1}$ or the Choleski decomposition $L_i D_i L_i^T = B_i$, instead of B_i itself, using only $O(n^2)$ operations per iteration. Even if variable metric methods ussually converge slower then modified Newton methods, they are the most efficient known methods for small and moderate size dense problems.

Conjugate gradient methods, see [67], [52], [104], use only n-dimensional vectors. Direction vectors $d_i \in \mathcal{R}^n$, $i \in \mathcal{N}$ 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 $F : \mathcal{R}^N \to \mathcal{R}$ at the point x_{i+1} and β_i is a properly defined scalar parameter. Conjugate gradient methods require only $O(n)$ storage elements and $O(n)$ operations per iteration, but they converge much slower than variable metric methods. Conjugate gradient methods are intended for large scale problems, but they can be less efficient, in the sparse case, than special versions of the modified Newton or the variable metric methods.

Basic optimization methods for unconstrained optimization can be realized in various ways which differ in direction determination and stepsize selection. Monotone line search and trust region realizations are the most popular ones, especially in connection with variable metric methods.

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

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

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

$$F_{i+1} - F_i \le \epsilon_1 \alpha_i g_i^T d_i, \tag{1.6}$$

$$g_{i+1}^T d_i \ge \epsilon_2 g_i^T d_i, \tag{1.7}$$

where $0 < \epsilon_1 < 1/2$ and $\epsilon_1 < \epsilon_2 < 1$ (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 [48], characterizes the global convergence of monotone line search methods.

THEOREM 1.1. *Let the objective function $F : \mathcal{R}^N \to \mathcal{R}$ 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 of (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 a spectral condition number of the matrix B_i. Note that (1.8) (or (1.9)) is satisfied if a constant $\underline{c} > 0$ (or $\overline{c} > 0$) and an infinite set $\mathcal{M} \subset \mathcal{N}$ exist so that $c_i \geq \underline{c}$ (or $\kappa_i < \overline{c}$) $\forall i \in \mathcal{M}$.

Variable metric methods in a line search realization require the direction vectors to satisfy conditions (1.5) and $\|B_i d_i + g_i\| \leq \omega_i \|g_i\|$, where $0 \leq \omega_i \leq \overline{\omega} < 1$ is a prescribed precision (the later condition is important 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 $d_i = -H_i g_i$ or use the back substitution $L_i D_i L_i^T d_i = -g_i$ after Choleski decomposition. Otherwise, an iterative method may be preferable. The preconditioned conjugate gradient method is specially 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

$$
\begin{aligned}
q_j &= B_i p_j, \\
\alpha_j &= r_j^T C_i^{-1} r_j / p_j^T q_j, \\
s_{j+1} &= s_j + \alpha_j p_j, \\
r_{j+1} &= r_j - \alpha_j q_j, \\
\beta_j &= r_{j+1}^T C_i^{-1} r_{j+1} / r_j^T C_i^{-1} r_j, \\
p_{j+1} &= C_i^{-1} r_{j+1} + \beta_j p_j
\end{aligned}
\tag{1.10}
$$

for $j \in \mathcal{N}$. This process is terminated if either $\|r_j\| \leq \omega_i \|g_i\|$ (sufficient precision) or $p_j^T q_j \leq 0$ (nonpositive curvature). In both cases we set $d_i = s_j$. The matrix C_i is a preconditioner which should be chosen in such a way to make $B_i C_i$ as well-conditioned as possible. Very efficient preconditioners can be obtained using the incomplete Choleski decomposition, see [11].

Monotone line search is suitable for conjugate gradient and variable metric methods. In [64], [65], [141], a nonmonotone line search is studied, which allows the stepsize α_i to take the unit value in more cases. Notice that choosing $\alpha_i = 1$ is important for obtaining a superlinear rate of convergence with variable metric and modified Newton methods.

Monotone trust region methods use direction vectors $d_i \in \mathcal{R}^n$, $i \in \mathcal{N}$, which satisfy the following conditions

$$\|d_i\| \leq \Delta_i, \tag{1.11}$$

$$\|d_i\| < \Delta_i \Rightarrow \|B_i d_i + g_i\| \leq \omega_i \|g_i\|, \tag{1.12}$$

$$-Q_i(d_i) \geq \underline{\sigma} \|g_i\| \min(\|d_i\|, \|g_i\|/\|B_i\|), \tag{1.13}$$

where $0 \leq \omega_i \leq \overline{\omega} < 1$ and $0 < \underline{\sigma} < 1$. Steplengths $\alpha_i \geq 0$, $i \in \mathcal{N}$, are chosen so that

$$\rho_i(d_i) \leq 0 \Rightarrow \alpha_i = 0, \tag{1.14}$$

$$\rho_i(d_i) > 0 \Rightarrow \alpha_i = 1, \tag{1.15}$$

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

$$\rho_i(d_i) < \underline{\rho} \Rightarrow \underline{\beta} \|d_i\| \leq \Delta_{i+1} \leq \overline{\beta} \|d_i\|, \tag{1.16}$$

$$\rho_i(d_i) \geq \underline{\rho} \Rightarrow \Delta_i \leq \Delta_{i+1} \leq \overline{\Delta}, \tag{1.17}$$

where $0 < \underline{\beta} \leq \overline{\beta} < 1$ and $0 < \underline{\rho} < 1$. The following theorem, see [108], characterizes global convergence of monotone trust region methods.

THEOREM 1.2. *Let the objective function $F : \mathcal{R}^N \to \mathcal{R}$ be bounded from below and have bounded second order derivatives. Consider the trust region method (1.11)-(1.17) and denote $M_i = \max(\|B_1\|, \ldots, \|B_i\|)$, $i \in \mathcal{N}$. If*

$$\sum_{i\in\mathcal{N}} \frac{1}{M_i} = \infty, \tag{1.18}$$

then $\liminf_{i\to\infty} \|g_i\| = 0$.
Note that (1.18) is satisfied if a constant $\overline{B}$ and an infinite set $\mathcal{M} \subset \mathcal{N}$ exist, so that $\|B_i\| \leq \overline{B}$, $\forall i \in \mathcal{M}$.

Monotone trust region methods require the direction vectors to satisfy conditions (1.11)-(1.13). Such vectors can be obtained in three basic ways. The most sophisticated way consists in solving the constrained minimization subproblem

$$d_i = \arg\min_{\|d\|\leq\Delta_i} Q_i(d), \tag{1.19}$$

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 [94], is time consuming since it requires, on average, 2 or 3 Choleski decompositions per iteration. Moreover, an additional matrix has to be used. Therefore, easier approaches have been looked for.

One such approach consists in replacing the complicated subproblem (1.19) by the two-dimensional subproblem

$$d_i = \arg\min_{\|d(\alpha,\beta)\|\leq\Delta_i} Q_i(d(\alpha,\beta)), \tag{1.20}$$

where $d(\alpha,\beta) = \alpha g_i + \beta B_i^{-1} g_i$. The subproblem (1.20) is usually solved approximately by the so-called dog-leg methods [40], [105].

If the original problem is of a large-size, then the inexact trust region method, proposed in [136], can be advantageously used. This method is based on the fact that the vectors s_j, $j \in \mathcal{N}$, determined by the preconditioned conjugate gradient method (1.10), satisfy the recurrence inequalities

$$\begin{aligned} s_{j+1}^T C s_{j+1}^T &> s_j^T C s_j, \\ Q(s_{j+1}) &< Q(s_j), \end{aligned}$$

where Q is the quadratic function (1.2) or (1.3). Thus a suitable path in the trust region is generated. If $\|s_j\| \leq \Delta_i$ and $\|r_j\| \leq \omega_i \|g_i\|$, then we set $d_i = s_j$. If $\|s_j\| \leq \Delta_i$ and $p_j^T q_j \leq 0$, 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 $\|d_j\| \leq \Delta$ 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 by the conjugate gradient process.

Besides monotone trust region methods, nonmonotone methods have been developed, giving more possibilities for the stepsize α_i to be unit. One such realization is studied in [36]. Trust region methods can also be combined with line search methods as it is shown in [99] and [140] .

2. Variable metric methods for dense problems. Variable metric methods have been originally developed for unconstrained minimization of objective functions with dense Hessian matrices. As was mentioned above, these methods use positive definite matrices B_i, $i \in \mathcal{N}$, which are generally constructed iteratively using the 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

$$\tilde{G}_i = \int_0^1 G(x_i + \lambda s_i) d\lambda \tag{2.1}$$

which is usually 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 due to the fact that it is satisfied by the 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 [149]).

Instead of matrices B_i, $i \in \mathcal{N}$, we can alternatively construct matrices $H_i = B_i^{-1}$, since the equation $B_i d_i + g_i$ can be easily solved in this case by setting

$$d_i = -H_i g_i \tag{2.2}$$

We can slightly generalize the above approach to include a nonquadratic correction and scaling. For simplifying the notation, we omit the index i and replace the index $i+1$ by $+$. Moreover, we denote

$$a = y^T H y, \quad b = y^T s, \quad c = s^T H^{-1} s. \tag{2.3}$$

A nonquadratic correction was first introduced in [13]. Consider the function $\phi(\lambda) = F(x + \lambda s)$. Using the backward Taylor expansion, we can write $\phi(0) = \phi(1) - \phi'(1) + (1/2)\phi''(\lambda)$, where $0 \le \lambda \le 1$. On the other hand, if we assume the generalized quasi-Newton condition

$$H_+ y = \rho s, \tag{2.4}$$

then $s^T B_+ s = s^T y / \rho$. Approximating $s^T B_+ s$ by $\phi''(\lambda)$ obtained from the backward Taylor expansion, we get

$$\rho = \frac{s^T y}{2(F - F^+ + s^T g^+)}. \tag{2.5}$$

Formula (2.5) was derived in [128] and also used in [148].

Another modification of the quasi-Newton equation, based upon a special approximation of G_{i+1} that is usually better than (2.1), has been proposed in [150]. This modification amounts to add to the vector y a term of the form $\gamma s / \| s \|$, where $\gamma = 3(g^+ + g)^T s - 6(f^+ - f)$. Rather extensive numerical experiments indicate that an implementation of the BFGS method with this definition of y is better than the implementation of the standard formula.

Linear scaling of the matrix was first introduced in [100]. A simple reason for scaling is the replacement of the matrix H by the scaled matrix γH before update to make the difference $H_+ - \gamma H$ as small as possible. One possibility is to derive γ from

(2.4) after premultiplying it by a vector and replacing H_+ by γH. Using the vector y, we obtain

$$\gamma/\rho = b/a. \tag{2.6}$$

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

$$\gamma/\rho = c/b. \tag{2.7}$$

Another useful value is the geometrical mean

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

It is interesting that these simple values often considerably improve the efficiency of variable metric methods, while more sophisticated formulas, derived by minimization of certain potential functions, give usually worse results. Scaling applied in every iteration is inefficient in general, see [114], but can be very useful on very difficult functions, see [122]. Therefore, some selective scaling strategies have been developed. The simplest possibility, scaling only in the first iteration (or preliminary scaling, PS), is proposed in [114] (but already in [100] it was shown that scaling at the first iteration was enough to satisfy a number of properties on quadratic functions). In [27], it is recommended to use the scaling parameter $\gamma = \max(1, \min(\tilde{\gamma}, \overline{\gamma}))$ in every iteration, where $\tilde{\gamma}$ is a theoretically computed value and $\overline{\gamma}$ 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 [86]. Here the value $\gamma = \tilde{\gamma}$ is used, if $\underline{\gamma} \le \tilde{\gamma} \le \overline{\gamma}$. Otherwise we set $\gamma = 1$. Recommended values $0 < \underline{\gamma} < 1 < \overline{\gamma}$, corresponding to individual formulas (2.6)-(2.8), are also given in [86].

Now we are in a position to derive a class of scaled variable metric methods satisfying the generalized quasi-Newton condition (2.4). 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 small. Since two vectors s and Hy appear in the generalized quasi-Newton condition (2.4), we restrict our attention to rank two updates of the form $\Delta H = \gamma U M U^T$, 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

$$\frac{1}{\gamma} H_+ = H + \frac{\rho}{\gamma} \frac{1}{b} s s^T - \frac{1}{a} H y (H y)^T + \frac{\eta}{a} \left(\frac{a}{b} s - H y \right) \left(\frac{a}{b} s - H y \right)^T. \tag{2.9}$$

Formula (2.9) defines a three-parameter class, the so-called Huang-Oren class of variable metric updates, see [128]. If we assume ρ and γ to be constants, given by (2.5) and (2.6)-(2.8), 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 (Davidon [28], Fletcher and Powell [51]) update

$$\frac{1}{\gamma} H_+ = H + \frac{\rho}{\gamma} \frac{1}{b} s s^T - \frac{1}{a} H y (H y)^T. \tag{2.10}$$

Setting $\eta = 1$, we get the scaled BFGS (Broyden [17], Fletcher [47], Goldfarb [58], Shanno [113]) update

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

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

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

The formula (2.9) gives another reason for scaling. It can be proved, see [100], that if $0 \le \eta \le 1$ and $b/c \le \rho/\gamma \le a/b$, then $\kappa(\tilde{G} H_+) \le \kappa(\tilde{G} H)$, where $\tilde{G}$ is the matrix defined by (2.1) (κ denotes the spectral condition number). It is clear that for (2.6)-(2.8) the inequality $b/c \le \rho/\gamma \le a/b$ holds ($b/c \le a/b$ 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 (2.26)).

Denoting $\Delta B = B_+ - (1/\gamma) B$, we can write (2.4) in the form $\Delta B s = (1/\rho) y - (1/\gamma) B s$. Proceeding as above, we obtain

$$\gamma B_+ = B + \frac{\gamma}{\rho} \frac{1}{b} y y^T - \frac{1}{c} B s (B s)^T + \frac{\beta}{c} \left(\frac{c}{b} y - B s \right) \left(\frac{c}{b} y - B s \right)^T, \tag{2.13}$$

see (2.9), 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

$$\beta \eta (ac - b^2) + (\beta + \eta) b^2 = b^2. \tag{2.14}$$

Therefore, setting $\beta = 1$, we get the scaled DFP update

$$\gamma B_+ = B + \left(\frac{\gamma}{\rho} + \frac{c}{b} \right) \frac{1}{b} y y^T - \frac{1}{b} \left(B s y^T + y (B s)^T \right). \tag{2.15}$$

Setting $\beta = 0$, we get the scaled BFGS update

$$\gamma B_+ = B + \frac{\gamma}{\rho} \frac{1}{b} y y^T - \frac{1}{c} B s (B s)^T. \tag{2.16}$$

Setting $\beta = (\gamma/\rho)/(\gamma/\rho - c/b)$, we get the scaled SR1 update

$$\gamma B_+ = B + \left(\frac{\gamma}{\rho} - \frac{c}{b} \right)^{-1} \frac{1}{b} \left(\frac{\gamma}{\rho} y - B s \right) \left(\frac{\gamma}{\rho} y - B s \right)^T. \tag{2.17}$$

Comparing (2.15)-(2.17) with (2.10)-(2.12), we observe a kind of duality. The DFP and BFGS formulas are dual to each other and the SR1 formula is self-dual.

The scaled Broyden class contains very effective variable metric methods, which have excellent theoretical properties.

In subsequent considerations, we will assume that the vectors s and $H y$ are linearly independent. Otherwise, the generalized quasi-Newton condition (2.4) 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). Denoting

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

we can deduce, from (2.14), that $\eta < \eta^c$, $\eta^c < \eta < 0$, $0 \le \eta \le 1$, $1 < \eta$, if and only if $\beta < \beta^c$, $1 < \beta$, $0 \le \beta \le 1$, $\beta^c < \beta < 0$, respectively. Moreover, one can prove, see [120], [121], that the matrix H_+ (or B_+) is positive definite if and only if $b > 0$ and $\eta > \eta^c$ (or $\beta > \beta^c$). The 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 relation $0 \le \eta \le 1$ (or $0 \le \beta \le 1$) defines the so called *restricted Broyden class*, 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 [17].

THEOREM 2.1. *Let the objective function $F : \mathcal{R}^N \to \mathcal{R}$ be quadratic with positive definite Hessian matrix G. Consider the variable metric method (1.1) with stepsizes chosen so that $g_{i+1}^T d_i = 0$ (perfect line search) and direction vectors determined by (2.2) and (2.9). Then there exists an index i, $1 \le i \le n$, such that the direction vectors d_j, $1 \le j \le i$, are mutually G-conjugate (i.e. $d_j^T G d_k = 0$ whenever $j \ne k$ and $1 \le j \le i$, $1 \le k \le i$) and, moreover, $g_{i+1} = 0$ and $x_{i+1} = x_*$.*

Quadratic termination property requires perfect line searches in general. Since this property seemed essential in the past, many authors proposed variable metric methods keeping this property even without perfect line searches. Such property characterizes the SR1 update with $\rho = 1$ and $\gamma = 1$, but it does not guarantee positive definiteness of the generated matrices. An interesting possibility was proposed in [29], where the Broyden class (2.9) was replaced by updates of the form $H_+ = H + VNV^T$, where $V = [u, v]$, N is a symmetric 2×2 matrix, $u = s - Hy$ and v is a vector generated recursively by the formula

$$v_+ = u^T y v - v^T y u. \tag{2.19}$$

Determining N from the quasi-Newton condition, we obtain a one-parameter class

$$H_+ = H + \frac{uu^T}{u^T y} - \phi \frac{v_+ v_+^T}{v^T y}. \tag{2.20}$$

Since the determination of a suitable value ϕ in (2.20) is difficult, Davidon recommended the use of projections Ps and PHy, instead of s and Hy, where $P = V(V^T H^{-1} V)^{-1} V^T H^{-1}$. These projections are substituted into (2.9)-(2.12) (with $\rho = 1$ and $\gamma = 1$) for obtaining the matrix H_+ defined by (2.19)-(2.20). Unfortunately, condition (1.7) does not guarantee positivity of $y^T Ps$, so that positive definiteness of H_+ can be lost. Formulas (2.19)-(2.20) are not used presently since quadratic termination was shown to be unnecessary for obtaining superlinear rate of convergence (cf. Theorem 2.3), hence the expensive projections Ps and PHy may not be used.

The most general global convergence results can be found in [23]. We summarize them in the following theorem.

THEOREM 2.2. *Consider the variable metric method (1.1) with $B_i d_i = -g_i$, (1.6), (1.7) and (2.13) with $1 \le \gamma_i \le \overline{\gamma}$, $0 < \underline{\rho} \le \rho_i \le \overline{\rho}$ and $(1 - \underline{\delta})\beta_i^c \le \beta_i \le 1 - \underline{\delta}$, where $0 < \underline{\delta} < 1$. Let the initial point $x_1 \in \mathcal{R}^n$ be chosen so that the objective function $F : \mathcal{R}^n \to \mathcal{R}$ is uniformly convex and has bounded second order derivatives on the convex hull of the level set $\mathcal{L}_1 = \{x \in \mathcal{R}^n : F(x) \le F(x_1)\}$. Then $\liminf_{i\to\infty} \|g_i\| = 0$.*

The above theorem has some important consequences. First, it cannot be proved when $\beta \ge 1$, which is a reason explaining the bad properties of the DFP method.

Secondly, it confirms that values $\beta^c < \beta < 0$ (or $1 < \eta$) are permissible (computational experiments have shown that some particular methods from this subclass are very efficient in practice). Third, the restriction $\gamma \geq 1$ has also a practical consequence and it was used in [27] 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 $\cos\theta_i$, 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 restart eventually does not occur if the objective function is uniformly convex in a neighborhood of the minimizer.

Another way which guarantees global convergence of the line search method consists in turn the search direction towards the negative gradient. This idea is realized e.g. if (2.2) is replaced by the formula $d = -\bar{H}g$ with

$$\bar{H} = H + \sigma\|Hg\|I \quad \text{or} \quad \bar{H} = H + \sigma\|Hg\|\frac{gg^T}{g^Tg} \tag{2.21}$$

where H is a matrix obtained by the variable metric update (2.9) and $\sigma > 0$ is a small number. Theoretical investigation of such modifications of variable metric methods is given in [110].

An important property of variable metric methods belonging to the Broyden class is their superlinear rate of convergence. The most general results concerning superlinear rate of convergence are given in [21]. 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.* $\|G(x) - G(x_*)\| \leq \overline{L}\|x - x_*\|$ *for all* x *from some neighborhood of* x_**). Then a value* $\underline{\beta} < 0$ *exists such that if* $\beta_i \geq \underline{\beta}$ $\forall i \in \mathcal{N}$, *then* $\lim_{i\to\infty} \|x_{i+1} - x^*\|/\|x_i - x^*\| = 0$.

This theorem generalizes results given in [63], where superlinear rate of convergence was proved for the restricted Broyden subclass corresponding to the values $0 \leq \beta \leq 1$ in (2.13). 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 global convergence of variable metric methods. Roughly speaking, global convergence is usually improved if the value of η is increased, while superlinear rate of convergence can deteriorate for a large value of η, which increases the number of function evaluations. This effect is demonstrated in the following table. Here the total numbers of iterations and function evaluations are shown, in a case when the unscaled variable metric methods are used for solving 80 problems with 20 variables.

η	0.2	0.5	1.0	2.0	5.0	10.0	20.0	50.0
main iter.	13100	9650	8224	7202	6916	6532	6349	6589
f. eval.	16618	13337	12113	11579	13431	12662	13279	15394
CPU time	38.72	31.25	26.48	26.36	25.71	27.30	27.47	34.49

Table 1

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 [98], 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.

Now we focus our attention to the choice of the value η (or β). Motivated by the above theoretical results, we will assume that $\beta^c < \beta \leq 1$ (or $0 \leq \eta$), defining the perfect Broyden subclass. Among all classic updates (2.10)-(2.12), 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 may do 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

$$\eta = \frac{\rho/\gamma}{\rho/\gamma + a/b} \tag{2.22}$$

is proposed by Hoshino in [70]. 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 potential function, used first in [123], see also [101], is a condition number

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

where H_+ is given by (2.3) and $\overline{\lambda}$ and $\underline{\lambda}$ are the maximum and the minimum eigenvalues respectively. Denoting $\tilde{\eta} = 1 - \eta/\eta^c$ 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 $0 < \lambda_1 \leq \lambda_2$ 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

$$\eta(ac - b^2) = b^2 \left(\frac{\rho}{\gamma}\frac{c}{b} - 1 \right). \tag{2.23}$$

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

$$\eta = \frac{bc(\rho/\gamma - b/c)}{ac - b^2} \quad \text{if} \quad b \leq \frac{2(\rho/\gamma)ac}{a + (\rho/\gamma)^2 c}, \tag{2.24}$$

$$\eta = \frac{\rho/\gamma}{\rho/\gamma - a/b} \quad \text{if} \quad b > \frac{2(\rho/\gamma)ac}{a + (\rho/\gamma)^2 c} \tag{2.25}$$

(notice that (2.25) corresponds to the SR1 update). This optimally conditioned update was introduced in [29], the formula (2.24) was independently derived in [101] (see [125] for the equivalence of the Davidon's criterion with the Oren-Spedicato criterion).

Formula (2.23) can also be used for deriving the optimal ratio γ/ρ for a given value η, since we can write

$$\frac{\gamma}{\rho} = \frac{bc}{\eta(ac - b^2) + b^2}. \tag{2.26}$$

For $\eta = 1$ (BFGS) we obtain (2.6). For $\eta = 0$ (DFP) we obtain (2.7). For η given by (2.22) (Hoshino) we obtain (2.8). Substituting (2.8) back into (2.22) (or into (2.24)) we get the Oren-Spedicato update $\eta = b/(b + \sqrt{ac})$ (this value was also derived in [145] using a different variational principle). 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: Finding 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 may be used also for the SR1 update. The analysis of update (2.9) shows that the matrix H keeps positive definiteness for b positive if and only if the ratio γ/ρ lies inside the two disjoint open intervals $0 < \gamma/\rho < b/a$ and $c/b < \gamma/\rho < \infty$, see [128] (and [75], who derived only the second of the two intervals). Inside each of these two intervals there exists exactly one value of the ratio γ/ρ, which satisfies the Oren-Spedicato criterion.

The optimal ratio γ/ρ for the SR1 update, derived from (2.25)-(2.26), can be expressed in the form

$$\frac{\gamma}{\rho} = \frac{c}{b}\left(1 - \sqrt{1 - b^2/(ac)}\right) \text{``} = \frac{b}{a}/\left(1 + \sqrt{1 - b^2/(ac)}\right)\text{``}, \tag{2.27}$$

which is the value proposed in [102]. The important property of the optimally scaled SR1 update is the fact that it generates positive definite matrices since $0 < \gamma/\rho < b/a$. Unfortunately, this update leads to scaling applied in every iteration, which has a negative influence on the superlinear rate of convergence as it was mentioned above. Substituting (2.27) back into (2.25) (or into (2.24)) we get

$$\eta = 1 + 1/\sqrt{1 - b^2/(ac)}. \tag{2.28}$$

This choice lies outside the restricted Broyden subclass and gives usually better results than the BFGS update in the unscaled case (see [82]). Another very efficient modification of the SR1 method is proposed in [3] and [82]. This is a combination of the SR1 and the BFGS updates which can be written in the form

$$\eta = 1 \quad \text{if} \quad \rho/\gamma \le a/b, \tag{2.29}$$

$$\eta = \frac{\rho/\gamma}{\rho/\gamma - a/b} \quad \text{if} \quad \rho/\gamma > a/b, \tag{2.30}$$

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 been frequently 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 [54], that this Frobenius norm reaches its minimum, on the set of matrices satisfying the generalized quasi-Newton condition (2.4), if and only if

$$\gamma B_+ = B + \frac{wv^T + vw^T}{s^T v} - \frac{w^T s}{s^T v}\frac{vv^T}{s^T v}, \tag{2.31}$$

where $w = (\gamma/\rho)y - Bs$ and $v = Ws$. If the matrix W is chosen in such a way that $v = Ws$ lies in the subspace generated by the vectors y and Bs, i.e. if $z = y + \lambda Bs$, say, we obtain a portion of the scaled Broyden class (2.13). This portion contains variable metric methods for which $p \geq 0$, where p is defined by (2.41) below. The relation between λ and β is given by the formula

$$\beta = \frac{b(b - \lambda^2(\gamma/\rho)c)}{(b - \lambda c)^2}. \tag{2.32}$$

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 $\lambda = \rho/\gamma$.

Similar results can be obtained for the weighted Frobenius norm $\|W(H_+ - \gamma H)\|$. We get the update

$$\frac{1}{\gamma}H_+ = H + \frac{uv^T + vu^T}{y^T v} - \frac{u^T y}{y^T v}\frac{vv^T}{y^T v}, \tag{2.33}$$

where $u = (\rho/\gamma)s - Hy$ and $v = W^{-1}y$. If we set $v = s$, we get the BFGS update. It means that the BFGS method minimizes the weighted Frobenius norm $\|W(H_+ - \gamma H)\|$ for all SPD matrices W satisfying the condition $Ws = y$. Formula (2.33) defines a more general class of rank-two variable metric updates. This class contains a part of the Broyden class and is contained in the general class (2.20) (with v an arbitrary and v_+ given by (2.19)).

If we set $W = I$ in (2.31), we get the Powell symmetric Broyden (PSB) update

$$\gamma B_+ = B + \frac{sw^T + ws^T}{s^T s} - \frac{w^T s}{s^T s}\frac{ss^T}{s^T s}. \tag{2.34}$$

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).

A sophisticated choice of the weight W in terms of Bayesian estimation theory for the unknown Hessian was given by Thomas, see [137]. Thomas formula requires more overhead and memory since H must be updated by an additional rank-one correction. Limited experiments provided by Thomas indicated that his update is competitive with the BFGS formula. There is also a version of the update suitable for nonlinear equations, whose excellent behaviour compared with other quasi-Newton formulas has been confirmed by [128].

Other potential functions have been used for deriving variable metric methods. In [49], it was shown that the DFP update minimizes the function

$$\psi(X) = \mathrm{trace}(X) - \log(\det(X)), \tag{2.35}$$

where $X = H^{-1}H_+$, on the set of positive definite matrices H_+ satisfying the quasi-Newton condition $H_+ y = d$. Similarly, the BFGS method minimizes (2.35), where $X = B^{-1}B_+$, on the set of positive definite matrices B_+ satisfying the quasi-Newton condition $B_+ d = y$. Functions

$$\begin{aligned} \sigma(X) &= \overline{\lambda}(X)/\sqrt[n]{\det(X)}, \\ \tau(X) &= \mathrm{trace}(X)/\underline{\lambda}(X) \end{aligned}$$

are both minimized (either for $X = H^{-1}H_+$ and for $X = B^{-1}B_+$) by the optimally scaled SR1 update, see [143] and [144] ($\overline{\lambda}$ and $\underline{\lambda}$ are the maximum and the minimum eigenvalues respectively).

Besides the above potential functions, other principles have been used for the derivation of the free parameters in Broyden class of variable metric methods. Zhang and Tewarson [151] derived a least change update of the matrix L in the factorization $B = LL^T$. This update corresponds to the negative value

$$\beta = -\frac{1}{(\lambda+1)^2(c/b)}, \tag{2.36}$$

where λ is the unique positive root of the nonlinear equation

$$(\lambda+1)^2\left(\frac{c}{b}\lambda^2 - 1\right) = \lambda^2\left(\frac{a}{c} - 1\right). \tag{2.37}$$

Another approach is used in [152] and [84], where the angle θ_+, see (1.5), and the norm $\|s_+\|$ are minimized, respectively. The resulting value β is again negative. Mifflin and Nazareth [93] proposed an update based on minimization of the last prior deviation measured by the norm $\|(H_+y_- - s_-)\tilde{G}(H_+y_- - s_-)\|$, where $s_- = x - x_-$, $y_- = g - g_-$ are the most recent differences and $\tilde{G}$ is a symmetric positive definite matrix such that $\tilde{G}s = y$. The last prior deviation reaches its minimum if $\eta = (y^Ts_-/s^Ty_-)/(y^Ts_-/s^Ty_- - a/b)$. This update has properties very similar to the SR1 update (if $y^Ts_- = s^Ty_-$, then it coincides with the unscaled SR1 update). Therefore, safeguard like (2.29)-(2.30) is useful. In this case

$$\eta = 1 \quad \text{if} \quad s^Ty_-/y^Ts_- \geq b/a, \tag{2.38}$$

$$\eta = \frac{y^Ts_-/s^Ty_-}{y^Ts_-/s^Ty_- - a/b} \quad \text{if} \quad s^Ty_-/y^Ts_- < b/a, \tag{2.39}$$

Variable metric methods are more frequently realized in the form (2.9), but form (2.13) is also possible. In the second case, the Choleski decomposition LDL^T of the matrix B is updated using $O(n^2)$ operations by the numerically stable method described in [56]. This possibility is very attractive, since positive definiteness can be controlled. However, numerical experiments indicate that the form (2.9) is more efficient, measured by computational time, since cheaper operations are used and stability is not lost. The table given below compares several variable metric methods of the form (2.9) (with standard line search) with either preliminary scaling PS or interval scaling IS and also with nonquadratic correction (2.5). Values of the scaling parameter γ have been selected from (2.6)-(2.8) in such a way to give the best results for individual methods. Again total numbers of iterations and function evaluations for 80 problems with 20 variables are presented.

η	value of γ	PS with $\rho = 1$	IS with $\rho = 1$	IS with (2.5)
BFGS	(2.8)	8523 - 9641	4483 - 5451	4383 - 5311
DFP	(2.7)	29 failures	6746 - 8156	5525 - 6762
(2.22)	(2.8)	11027-11875	4686 - 5347	4631 - 5455
(2.24)-(2.25)	(2.8)	7878 - 8911	4925 - 5503	4689 - 5314
(2.28)	(2.8)	7175 - 8561	4522 - 5526	4456 - 5594
(2.29)-(2.30)	(2.8)	6049 - 7469	4541 - 5618	4375 - 5316
(2.36)-(2.37)	(2.6)	7757 - 9047	4692 - 5618	4337 - 5277
(2.38)-(2.39)	(2.6)	6162 - 7737	5376 - 6707	4171 - 5227

Table 2

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

$$d = Z\tilde{d}, \quad \tilde{d} = -\tilde{g}, \quad \tilde{g} = Z^T g. \tag{2.40}$$

Denote $\tilde{s} = Z^{-1}s = \alpha\tilde{d}$, $\tilde{y} = Z^T y$ and $\tilde{U} = [\tilde{s}, \tilde{y}]$, so that $a = \tilde{y}^T\tilde{y}$, $b = \tilde{y}^T\tilde{s}$, $c = \tilde{s}^T\tilde{s}$. If $p \geq 0$ and $q \geq 0$, where

$$p = \frac{1}{ab}\left(\eta\left(\frac{a}{b} - \frac{\rho}{\gamma}\right) + \frac{\rho}{\gamma}\right), \tag{2.41}$$

$$q = \frac{\rho}{\gamma}\frac{1}{ab}\left(\eta(ac - b^2) + b^2\right), \tag{2.42}$$

then the matrix (2.9) can be expressed in the form $H_+ = Z_+ Z_+^T$, where

$$\frac{1}{\sqrt{\gamma}} Z_+ = Z(I + \tilde{U}uv^T\tilde{U}^T), \tag{2.43}$$

$$uv^T = \frac{1}{\delta}\begin{bmatrix} \rho/\gamma + a\sqrt{p} \\ -1 - b\sqrt{p} \end{bmatrix}\begin{bmatrix} \sqrt{q} - b\sqrt{p} \\ -\sqrt{q}\gamma/\rho + c\sqrt{p} \end{bmatrix}^T,$$

$$\delta = \left(\frac{\rho}{\gamma}c - b\right) + \left(b - \frac{\gamma}{\rho}a\right)\sqrt{q} + \left(ac - b^2\right)\sqrt{p}.$$

The general formula (2.43), derived in [29], 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

$$\frac{1}{\sqrt{\gamma}} Z_+ = Z + \frac{1}{a} Z\left(\sqrt{\frac{\rho a}{\gamma b}}\tilde{s} - \tilde{y}\right)\tilde{y}^T. \tag{2.44}$$

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

$$\frac{1}{\sqrt{\gamma}} Z_+ = Z + \frac{1}{b} Z\tilde{s}\left(\sqrt{\frac{\rho b}{\gamma c}}\tilde{s} - \tilde{y}\right)^T. \tag{2.45}$$

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

$$\frac{1}{\sqrt{\gamma}} Z_+ = Z + \frac{\sqrt{q} - 1}{(\rho/\gamma)^2 c - 2(\rho/\gamma)b + a} Z\left(\frac{\rho}{\gamma}\tilde{s} - \tilde{y}\right)\left(\frac{\rho}{\gamma}\tilde{s} - \tilde{y}\right)^T. \tag{2.46}$$

Theoretically, it would be possible to invert (2.43), using the Sherman-Morrison formula, obtaining a similar expression for the matrix $A_+ = Z_+^{-1}$. Unfortunately, the vector $\tilde{y} = Z^T y = (A^T)^{-1}y$, required in that case, cannot be determined without inversion of the matrix A. The BFGS update, obtained by inversion of (2.45), is only one that allows us to overcome this difficulty by using the following transformation

$$\sqrt{\gamma} A_+ = A + \frac{1}{c}\tilde{s}\left(\sqrt{\frac{\gamma c}{\rho b}}\tilde{y} - \tilde{s}\right)^T A = A + \frac{1}{c}As\left(\sqrt{\frac{\gamma c}{\rho b}}y - A^T As\right)^T. \tag{2.47}$$

Formulas (2.43)-(2.46) are very advantageous for seeking minima on linear manifolds, when the matrix H is singular and the matrix Z is rectangular. Formula (2.47) is useful for nonlinear least squares. Another application of product form updates is proposed in [109]. We will demonstrate it on the BFGS method first. Let $\bar{Z} = ZQ$, where Q is an orthogonal matrix. Then $\bar{Z}\bar{Z}^T = ZZ^T = H$ and, therefore, $\bar{Z}^T B\bar{Z} = I$ (B-conjugacy condition). If we choose an orthogonal matrix Q in such a way that the first column of $\bar{Z}$ is a multiple of s, we obtain, from the B-conjugacy condition, $s^T B\bar{Z}e_j = 0$, for $2 \le j \le n$. Since $s = Z\tilde{s} = \bar{Z}Q^T\tilde{s}$, the orthogonal matrix Q has to be chosen in such a way that $Q^T\tilde{s} = \|\tilde{s}\|e_1$. Formula (2.45) (with $\rho = 1$ and $\gamma = 1$) can be rewritten in the form

$$Z_+ = Z + \frac{1}{b}Z\tilde{s}\left(\sqrt{\frac{b}{c}}\tilde{s} - \tilde{y}\right)^T = Z + \frac{1}{b}s\left(\sqrt{\frac{b}{c}}Bs - y\right)^T Z. \tag{2.48}$$

Applying (2.48) on the matrix $\bar{Z}$ we get

$$\begin{aligned} Z_+e_1 &= \frac{1}{\sqrt{b}}s \\ Z_+e_j &= \bar{Z}e_j - \frac{y^T\bar{Z}e_j}{b}s,\ 2 \le j \le n \end{aligned}$$

since $Z_+e_1 = \lambda s$ by (2.48) and both the quasi-Newton and the B_+-conjugacy conditions imply $\lambda^2 b = \lambda^2 s^T B_+ s = 1$.

Further simplification can be obtained, if the matrix Q is chosen in such a way that $Q^T\tilde{y}$ is a linear combination of e_1 and e_2 (this requires additional elementary Givens rotations which sucessively replace the last $n-2$ elements of $\tilde{y}$ by zeros). Then

$$\begin{aligned} Z^+e_1 &= \frac{1}{\sqrt{b}}s, \\ Z^+e_2 &= \phi(\bar{Z}e_2 - \frac{y^T\bar{Z}e_2}{b}s), \\ Z^+e_j &= \bar{Z}e_j,\ 3 \le j \le n, \end{aligned}$$

where $\phi = 1$ so that only the first and the second columns of the matrix $\bar{Z}$ are updated. Therefore, most of operations applied on Z are stable Givens rotations. Moreover, see [117], the second order information obtained by the updating technique successively propagates from the first to the last column of the updated matrix, so that remaining columns can be scaled without loss of this information. The above formulas can be generalized to include a one-parameter subclass of the Broyden class of variable metric updates (see [116]). It suffices to choose the parameter ϕ by the formula

$$\phi = \sqrt{\frac{\eta(y^T\bar{Z}e_2)^2 + (y^T\bar{Z}e_1)^2}{(y^T\bar{Z}e_2)^2 + (y^T\bar{Z}e_1)^2}},$$

where η is the parameter in (2.9) (notice that $(y^T\bar{Z}e_2)^2 + (y^T\bar{Z}e_1)^2 = a$).

At the end of this section, we have to mention generalizations of variable metric methods having a finite termination property not only for convex quadratic functions but also for more general objective functions. One such a class is based on the generalized quadratic functions

$$F(x) = \phi(\bar{F}(x)), \tag{2.49}$$

where $\bar{F} : \mathcal{R}^N \to \mathcal{R}$ is a convex quadratic function and $\phi : \mathcal{R} \to \mathcal{R}$ is a continuously differentiable function such that $\sigma(x) = \partial\phi(\bar{F}(x))/\partial\bar{F} > 0$. It is clear that global minima of F and $\bar{F}$ coincide. Differentiating (2.49), we get $g(x) = \sigma(x)\bar{g}(x)$. Therefore, knowing $\sigma(x)$, we can derive $\bar{g}(x)$ from $g(x)$ and then substitute $\bar{g}(x)$ into formulas defining variable metric methods to obtain a minimizer of $\bar{F}$ which coincides with a minimizer of F. This way, first utilized in [124], is characterized by using $d = -H(g/\sigma)$ and $y = g_+/\sigma_+ - g/\sigma$ instead of $d = -Hg$ and $y = g_+ - g$ in all required expressions. Thus the main problem, which has to be solved, is the determination of the ratio σ_+/σ (the value σ_1, which realizes initial scaling, can be chosen arbitrarily). The ratio σ_+/σ can be easily determined for particular generalized quadratic functions. The general solution, independent on the particular model, was found in [126] for the perfect line search case and in [118] for the general case. Let $x_- = x + \alpha_- d$ and $x_+ = x + \alpha_+ d$ be two distinct points such that the gradients $g_- = g(x_-)$ and $g_+ = g(x_+)$ are linearly independent. Then we can write

$$\frac{\sigma}{\sigma_+} = \frac{\|g_-\|^2 g^T g_+ - g_-^T g_+ g^T g_-}{\|g_+\|^2 \|g_-\|^2 - (g_-^T g_+)^2} \, \frac{\alpha_- - \alpha_+}{\alpha_-}. \tag{2.50}$$

This approach is not quite suitable for construction of variable metric methods, since two function values per iteration have to be computed (classic variable metric methods usually require only one function value per iteration). Therefore, this approach is more suitable for conjugate gradient methods where more accurate line searches are used.

Other generalizations of variable metric methods are based on the conic function

$$F(x) = \frac{\bar{F}(x)}{l^2(x)} = F(0) + \frac{g^T(0)x}{1 + c^T x} + \frac{1}{2} \frac{x^T(\bar{G} - g(0)c^T - cg^T(0) - 2F(0)cc^T)x}{(1 + c^T x)^2}, \tag{2.51}$$

where $\bar{F} : \mathcal{R}^N \to \mathcal{R}$ is a convex quadratic function and $l : \mathcal{R}^N \to \mathcal{R}$ is a linear function with a constant gradient c (without loss of generality we may suppose that $l(0) = 1$). There are three basic approaches to use a conic model. The first and most sophisticated approach is based on the collinear scaling

$$\tilde{x} = \frac{x}{l(x)} = \frac{x}{1 + c^T x} \tag{2.52}$$

Using (2.51)-(2.52) and assuming that the matrix $\tilde{G} = \bar{G} - g(0)c^T - cg^T(0) - 2F(0)cc^T$ is positive definite, we can see that $F(x) = F(0) + g^T(0)\tilde{x} + (1/2)\tilde{x}^T\tilde{G}\tilde{x} = \tilde{F}(\tilde{x})$, where $\tilde{F}(\tilde{x})$ is a convex quadratic function. It can be shown, see [30], that a perfect line search in the space of x corresponds to a perfect line search in the space of $\tilde{x}$. Thus we can construct variable metric methods reaching a minimum of the quadratic function $\tilde{F}(\tilde{x})$ after a finite number of steps. Then, using the inverse transformation to (2.52), we obtain methods which find a minimum of the conic function (2.51) after the same number of steps. This approach, used in [31] and [81], is rather complicated and again two function values per iteration are required. Therefore, such methods are not competitive with standard variable metric methods.

The second approach is based on the local conic model

$$F(x + s) \approx C(s) = F(x) + \frac{g^T(x)s}{1 + c^T s} + \frac{1}{2} \frac{s^T B s}{(1 + c^T s)^2},$$

which reaches its minimum for $s = d = B^{-1}g/(1+c^TB^{-1}g)$. The matrix B is updated in such a way to the new conic model $C_+(s)$ (with the matrix B_+ and the vector c_+) satisfies the following interpolation (generalized quasi-Newton) conditions

$$C_+(0) = F, \quad \nabla C_+(0) = g, \quad C_+(s) = F_+, \quad \nabla C_+(s) = g_+.$$

It can be shown, see [119] and [8], that these interpolation conditions imply a simple quasi-Newton condition of the form $B_+s = \beta(g_+ - \beta^2 g)$ and that $c_+ = (1-\beta)g/g^Ts$, where

$$\beta = \frac{F_+ - F - \sqrt{(F_+ - F)^2 - s^Tg_+s^Tg}}{s^Tg}.$$

Notice that some safeguards, e.g. choosing $\beta = 1$, which corresponds to the quadratic model with $c = 0$, have to be used if the argument of the square root is negative. This approach does not lead to finite conic termination. Nevertheless, superlinear rate of convergence and global convergence are proved in [9], [35] and [34]. Unfortunately, computational experiments indicate that such methods are less efficient than standard variable metric methods.

The third approach is more general and leads to methods reaching a minimum of the generalized conic functions

$$F(x) = \phi(\bar{F}(x), l(x)), \tag{2.53}$$

where $\bar{F} : \mathcal{R}^N \to \mathcal{R}$ is a convex quadratic function, $l : \mathcal{R}^N \to \mathcal{R}$ is a linear function with a constant gradient c and $\phi : \mathcal{R} \to \mathcal{R}^2$ is a continuously differentiable function such that $\sigma(x) = \partial\phi(\bar{F}(x), l(x))/\partial\bar{F} > 0$, after a finite number of steps. It is based on the fact that if x is restricted to the linear manifold orthogonal to c, then (2.53) is a generalized quadratic function. Therefore, knowing the vector c, we can find, after at most $n-1$ steps, a minimizer x^A of (2.53) subject to the linear constraint $c^T(x - x_1^A) = 0$, where x_1^A is the starting point. Having two such minimizers x^A and x^B satisfying $c^T(x^A - x^B) \neq 0$, we can determine the global minimizer of (2.53) which lies on the line connecting x^A and x^B. This approach, used in [80], has a theoretical significance, since it shows that a broad class of objective functions can be minimized after a finite number of steps. However, the practical significance of such methods is not great, since the determination of the vector c is rather complicated and again two function values per iteration are required.

Finally, we notice that the rank-two update classes that we have considered till now, namely updates (2.9) and (2.31), are only special cases of the set of solutions of the quasi-Newton equation. Observing that the quasi-Newton equation can be viewed as a set of n linear systems, each one consisting of just a single equation and all differing only in the right hand side, the general solution can be easily obtained using the techniques offered by the so called ABS class of algorithms for linear equations, see Abaffy and Spedicato, [1]. The obtained general formula contains two parameter matrices, see [133], and is equivalent with a formula previously obtained by Adachi, see [2], using the theory of generalized inverses. No new updates in this general class have yet been developed.

3. Variable metric methods for sparse 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 two basic approaches: preserving sparsity pattern by special updates or using classic updates applied to submatrices of lower dimension. The first approach was initiated by the work of Toint [138], the second one was initially proposed in [62].

Preserving sparsity pattern is a strong restriction, which eliminates some important properties of variable metric methods. Updates cannot have a low rank in general. For instance, a diagonal update of a diagonal matrix, which changes it to satisfy the quasi-Newton condition, can have rank n in general. Moreover, positive definiteness of the updated matrix can be lost for an arbitrary sparse update, which can be again 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. Denote

$$\begin{aligned} \mathcal{V}_Q &= \{B \in R^{n\times n} : Bs = y\} \\ \mathcal{V}_S &= \{B \in R^{n\times n} : B^T = B\} \\ \mathcal{V}_G &= \{B \in R^{n\times n} : G_{ij} = 0 \Rightarrow B_{ij} = 0\} \end{aligned}$$

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

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

$$\begin{aligned} s^i_j &= s_j \ , \ G_{ij} \neq 0 \\ s^i_j &= 0 \quad , \ G_{ij} = 0 \end{aligned}$$

and we rewrite the quasi-Newton condition in the form

$$e_i^T(B_+ - B)s^i = e_i^T(y - Bs), \quad 1 \leq i \leq n.$$

It can be proved, see [42], that the considered orthogonal projections can be expressed in the form

$$\begin{aligned} \mathcal{P}_Q B &= B + \frac{(y - Bs)s^T}{s^T s}, \\ \mathcal{P}_S B &= \frac{1}{2}(B + B^T), \\ (\mathcal{P}_G B)_{ij} &= B_{ij} \ , \ G_{ij} \neq 0, \\ (\mathcal{P}_G B)_{ij} &= 0 \quad , \ G_{ij} = 0, \\ \mathcal{P}_{QS} B &= B + \frac{(y - Bs)s^T + s(y - Bs)^T}{s^T s} - \frac{(y - Bs)^T s}{s^T s}\,\frac{ss^T}{s^T s}, \end{aligned}$$

$$
\begin{aligned}
\mathcal{P}_{QG}B &= B + \mathcal{P}_G(us^T), \\
\mathcal{P}_{SG}B &= \mathcal{P}_S\mathcal{P}_G B = \mathcal{P}_G\mathcal{P}_S B, \\
\mathcal{P}_{QSG}B &= B + \mathcal{P}_G(vs^T + sv^T),
\end{aligned}
$$

where the vector $u \in \mathcal{R}^n$ is a solution to the linear system $Du = y - Bs$ with the positive semidefinite diagonal matrix

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

and the vector $v \in \mathcal{R}^n$ is a solution to the linear system $Qv = y - Bs$ with the positive semidefinite matrix

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

which has the same sparsity pattern as the matrix B.

The variable metric method which uses the update

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

was proposed by Toint in [138]. Realization of this method is time consuming, since an additional linear system has to be solved. Convergence properties of this method are also not excellent, 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 [135] has shown that the updates based on the composite projections

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

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

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

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 variable metric method (1.11)-(1.17), where* $B_{i+1} = B_i$, *if (1.14) holds, and updates of the form (3.1)-(3.4) are used, otherwise. Let the objective function* $F : \mathcal{R}^N \to \mathcal{R}$ *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 (1.12), 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, computational experiments show that a line search realization usually outperforms a trust region implementation. Moreover, the update (3.2) is the most efficient among all composite projections. It is also mentioned in [44] and [142].

Another way for obtaining sparse quasi-Newton update is described in [50]. This method is based on the minimization of the potential function (2.35), where $X = HB_+$, on the linear manifold $\mathcal{V}_Q \cap \mathcal{V}_S \cap \mathcal{V}_G$. Function (2.35) 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 loosing positive definiteness. Fletcher [50] proved that if the minimum of (2.35) on $\mathcal{V}_Q \cap \mathcal{V}_S \cap \mathcal{V}_G$ exists, it is characterized by the existence of $\lambda \in \mathcal{R}^n$ such that

$$\mathcal{P}_G H_+ = \mathcal{P}_G(H + \lambda s^T + s\lambda^T). \tag{3.5}$$

The vector λ cannot be obtained explicitly in the sparse case. In such a case the nonlinear system of equations $B_+(\lambda)s - y = 0$ has to be solved using the Newton method, where $B_+(\lambda)$ is a matrix determined from (3.5). This approach has two difficulties. First, the determination of $B_+(\lambda)$ from $\mathcal{P}_G H_+(\lambda)$ is rather complicated and it requires a sparsity pattern which is not changed during the Choleski 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, $\mathcal{P}_G H$ and M are necessary. Nevertheless, numerical experiments in [50] indicate robustness and good convergence properties of this method.

A quite different approach to large-scale optimization was developed by Griewank and Toint in [62]. Their idea leads to partitioned updating methods based on the expression

$$F(x) = \sum_{k=1}^{m} f_k(x), \tag{3.6}$$

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 $\hat{g}_k(x) \in \mathcal{R}^{n_k}$ and packed Hessian matrices $\hat{G}_k(x) \in \mathcal{R}^{n_k \times n_k}$, $1 \le k \le m$, as dense but small-size vectors and matrices. Such a formulation is highly practical since e.g. sparse nonlinear least squares problems (see (5.1) below) have this structure.

Partitioned updating methods consider each element function separately and update approximations $\hat{B}_k$, $1 \le k \le m$, 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 (2.13) 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 play a role. Secondly, denoting $\hat{b}_k = \hat{y}_k^T \hat{d}_k$, $\hat{c}_k = \hat{d}_k^T \hat{B}_k \hat{d}_k$, we can observe that $\hat{b}_k \ge 0$ does not have to be guaranteed for all $1 \le k \le m$. This difficulty is essential and an efficient algorithm has to handle this situation. Therefore, the following partitioned BFGS method is recommended

$$\begin{aligned} \hat{B}_k^+ &= \hat{B}_k + \frac{1}{\hat{b}_k}\hat{y}_k\hat{y}_k^T - \frac{1}{\hat{c}_k}\hat{B}_k\hat{d}_k\left(\hat{B}_k\hat{d}_k\right)^T \quad , \quad b_k > 0, \\ \hat{B}_k^+ &= \hat{B}_k \quad , \quad b_k \le 0. \end{aligned} \tag{3.7}$$

Partitioned variable metric methods are very efficient for solving real world problems, much more efficient than sparse updates, but their convergence properties have not yet been satisfactorily investigated. Griewank and Toint [63] have proven 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 [139], where global convergence is proved under complicated and restrictive conditions. Some globally convergent modifications of partitioned variable metric methods are given in [61].

A disadvantage of partitioned variable metric methods consists in the fact 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 then the use of the partitioned structure. Nevertheless, approximations of packed element Hessian matrices can be used for construction of preconditioners [32].

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 conditions, since they are just additional linear equations, see [133], and also with symmetry conditions, see [129] and [134], since also symmetry is a linear condition. The sparse symmetric update is given in explicit form, while in the approach of, e.g., [138], a sparse linear system has to be solved. By requiring that the diagonal element be sufficiently large extra linear conditions are given which allow in general 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 [134] and [130]. The last result can be used to produce full SPD sparse updates by imbedding the minimization of the function $F(x)$ in the equivalent $(n+1)-$dimensional problem of minimizing $F(x)+x_{n+1}^2$. No particular algorithms or numerical experiments are yet available based upon this approach.

4. Variable metric methods with limited storage. Variable metric methods with limited storage are based on the application of a limited number of BFGS updates, which are computed recursively using previous differences s_j, y_j, $i-n \le j \le i-1$ (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 [103], and it is more efficient in other cases. Later a limited number of BFGS steps have been used for construction of a suitable preconditioner to the conjugate gradient method in [19] and [20] and also for the approximation of the Hessian matrix, see [97], [78], [55]. Such applications have been made possible by a special form of the BFGS update

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

We define the m-step BFGS method with limited storage as the iterative process (1.1) and (2.2), where $H_i = H_i^i$ and the matrix H_i^i is generated by the recurrence formula

$$H_{j+1}^i = \gamma_j^i V_j^T H_j^i V_j + \frac{\rho_j}{b_j} s_j s_j^T \tag{4.1}$$

for $i-m \le j \le i-1$, where $H^i_{i-m} = I$. At the same time $\gamma^i_{i-m} = b_{i-1}/a_{i-1}$ and $\gamma^i_j = 1$ for $i-m < j \le i-1$. Using induction, we can rewrite (4.1) in the form

$$H^i_{j+1} = \frac{b_{i-1}}{a_{i-1}} \left(\prod_{k=i-m}^{j} V_k \right)^T \left(\prod_{k=i-m}^{j} V_k \right) + \sum_{l=i-m}^{j} \frac{\rho_l}{b_l} \left(\prod_{k=l+1}^{j} V_k \right)^T s_l s_l^T \left(\prod_{k=l+1}^{j} V_k \right) \tag{4.2}$$

for $i-m \le j \le i-1$. From (4.2), we can deduce that the matrix H^i_i can be determined using $2m$ vectors s_j, y_j, $i-m \le j \le i-1$, without storing the matrices H^i_j, $i-m \le j \le i-1$. This matrix may 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 [97]). First, the vectors

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

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

$$\begin{aligned} \sigma_j &= s_j^T u_{j+1}/b_j, \\ u_j &= u_{j+1} - \sigma_j y_j, \end{aligned}$$

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

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

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

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

for $i-m \le j \le i-1$, 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 storage, based on explicit expression of the matrix $H_i = H^i_i$ using low order matrices, was utilized in [22]. Let $H_i = H^i_i$ be the matrix obtained after m steps of the form

$$H^i_{j+1} = H^i_j + [s_j, H^i_j y_j] M_j [s_j, H^i_j y_j]^T,$$

$i-m \le j \le i-1$, where M_j, $i-m \le j \le i-1$, are symmetric matrices of order 2 which realize a particular variable metric method (2.9) with $\rho_j = 1$ and $\gamma_j = 1$. We wish to find an expression

$$H_i = H^i_{i-m} - [S_i, H^i_{i-m} Y_i] N_i^{-1} [S_i, H^i_{i-m} Y_i]^T, \tag{4.3}$$

where $S_i = [s_{i-m}, \ldots, s_{i-1}]$, $Y_i = [y_{i-m}, \ldots, y_{i-1}]$ and N_i is a symmetric matrix of order $2m$. Formula (4.3) was obtained for classical variable metric methods (DFP, BFGS, SR1), since the matrices M_j^{-1}, $i-m \le j \le i-1$, have a relatively simple form in these cases. Derivations, which can be found in [22], 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 $(R_i)_{kl} = s_k^T y_l$, for

$k \le l$, and $(R_i)_{kl} = 0$, otherwise. Furthermore, we denote by C_i the diagonal matrix of order m, such that $(C_i)_{kk} = s_k^T y_k$. Taking

$$N_i = \begin{bmatrix} -C_i, & R_i - C_i \\ (R_i - C_i)^T, & Y_i^T H_{i-m}^i Y_i \end{bmatrix} \tag{4.4}$$

in (4.3), we get the m-step DFP update. Taking

$$N_i = \begin{bmatrix} 0, & R_i \\ R_i^T, & C_i + Y_i^T H_{i-m}^i Y_i \end{bmatrix} \tag{4.5}$$

in (4.3), we get the m-step BFGS update. The m-step SR1 update can be written in the following slightly simpler form

$$H_i = H_{i-m}^i + (S_i - H_{i-m}^i Y_i)(R_i + R_i^T - C_i - Y_i^T H_{i-m}^i Y_i)^{-1}(S_i - H_{i-m}^i Y_i)^T. \tag{4.6}$$

In the sequel, we restrict our attention to the BFGS method which is the most efficient among all above updates. If we choose $H_{i-m}^i = \gamma_{i-m}^i I$, where $\gamma_{i-m}^i = b_{i-1}/a_{i-1}$, and if we explicitly invert the matrix (4.5), we can write

$$H_i = \gamma_{i-m} I + [S_i, \gamma_{i-m} Y_i] \begin{bmatrix} (R_i^{-1})^T (C_i + \gamma_{i-m} Y_i^T Y_i) R_i^{-1}, & -(R_i^{-1})^T \\ -R_i^{-1}, & 0 \end{bmatrix} [S_i, \gamma_{i-m} Y_i]^T . \tag{4.7}$$

This formula is more advantageous than the general one, since 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 $S_i^T Y_i$ by the upper part of $Y_i^T S_i$ (or by the transposed lower part of $S_i^T Y_i$). Therefore, we define the lower triangular matrix L_i, such that $(L_i)_{kl} = s_k^T y_l$, $k \ge l$ and $(L_i)_{kl} = 0$, otherwise. Then the m-step BFGS update can be written in the form

$$B_i = B_{i-m}^i - \left[Y_i, B_{i-m}^i S_i\right] \begin{bmatrix} -C_i, & (L_i - C_i)^T \\ L_i - C_i, & S_i^T B_{i-m}^i S_i \end{bmatrix}^{-1} \left[Y_i, B_{i-m}^i S_i\right]^T \tag{4.8}$$

If we choose $B_{i-m}^i = (1/\gamma_{i-m}^i) I$, where $\gamma_{i-m}^i = b_{i-1}/a_{i-1}$, and if we explicitly invert the matrix contained in (4.8), we can write

$$\begin{aligned} B_i = \frac{1}{\gamma_{i-m}} I \quad - \quad & \left[Y_i, \frac{1}{\gamma_{i-m}} S_i\right] \begin{bmatrix} -C_i^{-1/2}, & (\bar{L}_i^{-1}(L_i - C_i))^T \\ 0, & (\bar{L}_i^{-1})^T \end{bmatrix} \\ & \begin{bmatrix} C_i^{-1/2}, & 0 \\ -\bar{L}_i^{-1}(L_i - C_i), & \bar{L}_i^{-1} \end{bmatrix} \left[Y_i, \frac{1}{\gamma_{i-m}} S_i\right]^T , \end{aligned} \tag{4.9}$$

where

$$\bar{L}_i \bar{L}_i^T = (L_i - C_i)^T C_i^{-1/2} (L_i - C_i) + \frac{1}{\gamma_{i-m}} S_i T S_i \tag{4.10}$$

Formula (4.9) is more advantageous than (4.8), since the Choleski decomposition (4.10) is applied only on a matrix of order m.

Now we are in a position to present numerical experiments with various variable metric methods for large sparse unconstrained optimization. The table given below compares the sparse composite update SVM (3.2), the partitioned BFGS update PVM (3.7), the limited storage BFGS update LVM (4.2) and the partitioned discrete Newton method PNM described e.g. in [85]. All tested methods have been implemented in such a way that the direction vectors are determined using an unpreconditioned conjugate gradient method and stepsizes are chosen by a standard line search. The PVM method has also been implemented using a conjugate gradient method preconditioned by the incomplete Choleski decomposition (PVMIC). This preconditioning was not successful for any of the other methods. Computational experiments have been performed using 65 sparse test problems with 100 variables. Since the last 50 problems were sums of squares, we also performed composite tests, where the last 50 problems have been solved using a combination of the variable metric method with the Gauss-Newton method. In this case, SVM/GN means the simple sparse update (5.17), PVM/GN means the partitioned update (5.18) and PNM/GN means a combination of the partitioned discrete Newton method with the Gauss-Newton method (see [85] for more details).

method	main iter.	f. eval.	g. eval.	CG iter.	CPU time	failures
LVM	19023	20602	20602	-	3:34.37	2
SVM	9460	17520	17520	258533	4:55.78	-
PVM	4267	6407	6407	111646	2:36.05	1
PVMIC	3814	4686	4686	14570	1:36.72	-
PNM	2586	4222	16638	49899	2:19.79	1
SVM/GN	3286	5879	6751	74292	1:33.26	-
PVM/GN	2078	2676	3911	55788	1:13.71	-
PNM/GN	1786	2948	6529	55004	1:23.70	-

Table 3

5. Variable metric methods for nonlinear least squares. Suppose that the objective function $F : \mathcal{R}^N \to \mathcal{R}$ has the form

$$F(x) = \frac{1}{2} f^T(x) f(x) = \frac{1}{2} \sum_{k=1}^{m} f_k^2(x), \tag{5.1}$$

where $f_k : \mathcal{R}^n \to \mathcal{R}$, $1 \le k \le m$, are twice continuously differentiable functions. This objective function is frequently used for nonlinear regression and for solving systems of nonlinear equations. Utilizing (5.1), we can express the gradient and the Hessian matrix in the form

$$g(x) = J^T(x) f(x) = \sum_{k=1}^{m} f_k(x) g_k(x), \tag{5.2}$$

$$G(x) = J^T(x) J(x) + C(x) = \sum_{k=1}^{m} g_k(x) g_k^T(x) + \sum_{k=1}^{m} f_k(x) G_k(x), \tag{5.3}$$

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

The most popular method for nonlinear least squares problems is the Gauss-Newton method, which uses the first part of (5.3) as an approximation of the Hessian matrix, i.e. $B_i = J_i^T J_i \ \forall i \in \mathcal{N}$. 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, $f_k(x_i) \to 0 \ \forall k, 1 \le k \le m$. If $\|G_k(x)\| \le \overline{G}, \forall k, 1 \le k \le m$, then also

$$\|C(x_i)\| = \|\sum_{k=1}^{m} f_k(x_i)G_k(x_i)\| \le \overline{G} \sum_{k=1}^{m} |f_k(x_i)| \to 0$$

and, therefore, $\|G(x_i) - B_i\| = \|C(x_i)\| \to 0$, which implies a superlinear rate of convergence, see [41]. Since the Jacobian matrices J_i, $i \in \mathcal{N}$, are usually ill-conditioned, even singular, the Gauss-Newton method is most frequently realized in a trust region framework.

The Gauss-Newton method is very efficient, when it is 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 $B_i = J_i^T J_i$ can be a bad approximation of G_i in this case. According to these reasons, combinations of the Gauss-Newton method with special variable metric updates should be advantageous. Such combined methods exist and can be very efficient, but three problems have to be carefully solved. First, 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 the 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 criterion strongly influences the efficiency of the resulting method. The criterion proposed in [45] is based on comparing two predicted reductions with the actual one. A variable metric correction is applied in the subsequent iteration if

$$\left|\frac{F_+ - F}{s^T g + (1/2)s^T B s} - 1\right| \le \bar{\theta}_2 \left|\frac{F_+ - F}{s^T g + (1/2)s^T J^T J s} - 1\right|, \tag{5.4}$$

where B is the corrected matrix, and $0 < \bar{\theta}_2 < 1$. Condition (5.4) has a disadvantage in that an additional matrix vector product is necessary. This condition cannot be used if the matrix $J^T J$ (or J) is overwritten.

Another criterion, given in [111], is based on the observation that the ratio $\|Pf\|/\|P_- f_-\|$, where $P = J(J^T J)^{-1} J^T$, $P_- = J_-(J_-^T J_-)^{-1} J_-^T$ (f_-, J_- are quantities from the previous iteration), is a good estimate of the convergence rate of the Gauss-Newton method, at least in a neighbourhood of the solution. A neighbourhood of the solution can be detected by comparing the values $\|Pf\|$ and $\|f\|$. If we use the Gauss-Newton direction $d = -(J^T J)^{-1} J^T f$, then $Pf = -Jd$ and the resulting condition can be written in the form

$$\|Jd\| \le \min(\bar{\theta}_1 \|f\|, \bar{\theta}_2 \|J_- d_-\|), \tag{5.5}$$

where $0 < \bar{\theta}_1 < \bar{\theta}_2 < 1$. A variable metric correction is applied if (5.5) holds. Condition (5.5) can be used only if $J^T J d = -J^T f$ so that it is irrelevant if the direction vector d is obtained using variable metric corrections.

The most general and, at the same time, most efficient strategy is proposed in [53]. It uses the condition

$$F - F_+ \le \bar{\theta}_1 F, \tag{5.6}$$

where $0 < \bar{\theta}_1 < 1$. If (5.6) 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 briefly describe techniques for combining variable metric updates with the Gauss-Newton method. We will consider the following techniques: simple correction, cumulative correction and a successive approximation of the second order term in (5.3).

A simple correction is useful in the sparse case, when a cumulative correction cannot be realized. If the Gauss-Newton method should not be used, we compute the matrix B_+ (or A_+) from the matrix $J_+^T J_+$ (or J_+) using a variable metric update. Otherwise, we set $B_+ = J_+^T J_+$ (or $A_+ = J_+$).

A cumulative correction is proposed in [53]. If the Gauss-Newton method should not be used, we compute the matrix B_+ (or A_+) from the matrix B (or A) using a variable metric update. Otherwise, we set $B_+ = J_+^T J_+$ (or $A_+ = J_+$).

A successive approximation of the second order term is based on the model $B = J^T J + C$. The matrix C_+ is computed from the matrix C using variable metric updates. If the Gauss-Newton method should not be used, we set $B_+ = J_+^T J_+ + C_+$. Otherwise, we set $B_+ = J_+^T J_+$. While simple and cumulative corrections use the standard updates described in previous sections, the successive approximation of the second order term requires special updates (so-called structured updates) which we now describe. We will suppose that $\rho = 1$ and $\gamma = 1$ in (2.13). Later we will consider a special scaling technique.

There are two possibilities for construction of structured variable metric updates. The first method is based on the transformed quasi-Newton condition $C_+ s = z = J_+^T f_+ - J^T f - J_+^T J_+ s$. Therefore, the general update has the form (2.13) with B and y replaced by C and z, respectively. The SR1 update, derived in this way, can be written in the form

$$C_+ = C + \frac{(z - Cs)(z - Cs)^T}{s^T (z - Cs)}. \tag{5.7}$$

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 $s^T z$ is not guaranteed.

The second possibility consists in updating the matrix $\bar{B} = J_+^T J_+ + C$ to obtain the matrix $B_+ = J_+^T J_+ + C_+$ satisfying the quasi-Newton condition $B_+ s = y = J_+^T f_+ - J^T f$. The resulting general update has the form (2.13) with B replaced by $\bar{B}$. Since $y - \bar{B}s = z - Cs$, it is advantageous to use the formula (2.31). Then

$$\begin{aligned} C_+ &= C + \frac{(y - \bar{B}s)v^T + v(y - \bar{B}s)^T}{s^T v} - \frac{(y - \bar{B}s)^T s}{s^T v} \frac{v v^T}{s^T v} \\ &= C + \frac{(z - Cs)v^T + v(z - Cs)^T}{s^T v} - \frac{(z - Cs)^T s}{s^T v} \frac{v v^T}{s^T v} \end{aligned} \tag{5.8}$$

with $v = s$ for the structured PSB update, $v = y$ for the structured DFP update and $v = y + (y^T s/s^T \bar{B}s)^{1/2}\bar{B}s$ for the structured BFGS update. Methods based on the formula (5.8) have been investigated in [39], where a superlinear rate of convergence of the structured BFGS method was proved.

The vectors y and z, used in formulas (5.7)-(5.8), can be defined in various ways based on various ideas, but always $z = y - J_+^T J_+ s$. The standard choice

$$z = J_+^T f_+ - J^T f - J_+^T J_+ s, \tag{5.9}$$

corresponding to the quasi-Newton condition $(J_+^T J_+ + C_+)s = J_+^T f_+ - J^T f$, is introduced in [37]. In [12], a similar choice

$$z = J_+^T f_+ - J^T f - J^T J s, \tag{5.10}$$

corresponding to the quasi-Newton condition $(J^T J + C_+)s = J_+^T f_+ - J^T f$, is given. Another choice, studied in [7] and [127], is based on the objective function $\tilde{F}(x) = (1/2)(f^T(x)f(x) - x^T J^T J x)$, whose Hessian matrix is just the matrix $J^T J$ 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

$$z = J_+^T f_+ - J_+^T J_+ x_+ - J^T f + J^T J x. \tag{5.11}$$

A popular choice, proposed in [14], is based on the explicit form of the second order term in (5.3). Suppose that the approximations B_k^+ of the Hessian matrices G_k satisfy the quasi-Newton conditions $B_k^+ s = g_k^+ - g_k$, $1 \le k \le m$. Then we can write

$$z = C_+ s \triangleq \sum_{k=1}^{m} f_k^+ B_k^+ s = \sum_{k=1}^{m} f_k^+ (g_k^+ - g_k) = (J_+ - J)^T f_+. \tag{5.12}$$

Interesting methods for nonlinear least squares have been obtained from the product form BFGS update (2.47) (other product form updates are less suitable since they require the inversion of the matrix $A^T A$). A generalization of (2.47) (with $\rho = 1$ and $\gamma = 1$), related to structured update (5.8), is described in [147]. Here the matrix A is replaced by the matrix $J + L$, where J is the Jacobian matrix and L plays a similar role as C in (5.8). Thus $B = (J + L)^T (J + L)$, $B_+ = (J_+ + L_+)^T (J_+ + L_+)$ and if we denote $\bar{B} = (J_+ + L)^T (J_+ + L)$, we can express (5.8) in the form

$$L_+ = L + \frac{(J_+ + L)s}{s^T \bar{B} s} \left(\sqrt{\frac{s^T \bar{B} s}{s^T y}}\, y - \bar{B} s \right)^T, \tag{5.13}$$

which is similar to (2.47). Formula (5.13) can be derived in the following way. If we define $h = (J_+ + L_+)s$, then the quasi-Newton condition $B_+ s = y$ can be written in the form $(J_+ + L_+)^T h = y$ and we can formulate the problem of minimizing $\|L_+ - L\|_F$ subject to $(J_+ + L_+)^T h = y$. This variational problem has the solution

$$L_+ = L + \frac{h(s - \bar{B}^T h)^T}{h^T h},$$

which together with the definition of h and the equality $h^T h = s^T y$ gives (5.13).

Formula (5.13) has a disadvantage in that the least squares solution of the overdetermined system $(J+L)d=-f$ does not coincide with the solution of the normal equation $(J+L)^T(J+L)d=-g=-J^Tf$ (the later solution is necessary for construction of variable metric methods). Therefore, some efficient methods for direction determination (QR decomposition or the LSQR method) cannot be used in this case. This difficulty can be overcome if the matrix L satisfies the condition $L^Tf=0$, so that $(J+L)^Tf=J^Tf$. Therefore, it is advantageous to add the constraint $L_+^Tf_+=0$ to the variational problem determinimg L_+. The solution to this modified variational problem is derived in [115] and [146]. It is based on the substitution $L_+=P\bar{L}$, where $P=I-f_+f_+^T/f_+^Tf_+$ is the orthogonal projection matrix forcing L_+ to satisfy $L_+^Tf_+=0$. If we denote

$$\tilde{B}=(J_++L)^TP(J_++L),$$

$$\tilde{y}=y-\frac{J_+^Tf_+(J_+^Tf_+)^T}{f_+^Tf_+}s,$$

then we obtain

$$L_+=PL+\frac{P(J_++L)s}{s^T\tilde{B}s}\left(\sqrt{\frac{s^T\tilde{B}s}{s^T\tilde{y}}}\tilde{y}-\tilde{B}s\right)^T. \tag{5.14}$$

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 (2.7), we can choose $\gamma=s^TCs/s^Tz$ or $\gamma=\max(s^TCs/s^Tz,1)$, which is the value proposed in [38]. Biggs [14] recommends the value $\gamma=f^Tf/f_+^Tf$ based on a quadratic model. If we choose the scaling parameter γ, then we replace C by $(1/\gamma)C$ in (5.7)-(5.8) to obtain a scaled structured update. A more complicated process, described in [147], is used in connection with product form updates (5.13)-(5.14).

All above methods can be efficiently realized using switching strategies (5.4)-(5.6). Structured variable metric updates can be also used permanently (without switching), as follows from the theory given in [39], but such a realization is usually less efficient.

Interesting variable metric updates, intended especially for the permanent realization, are based on an approximation of the term

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

Thus we have the model $B=J^TJ+\|f\|T$. In analogy with structured variable metric methods, Huschens [72] 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 the expression (2.31), we can write

$$\begin{aligned} T_+ &= T+\frac{1}{\|f\|}\left(\frac{(y-\bar{B}s)v^T+v(y-\bar{B}s)^T}{s^Tv}-\frac{(y-\bar{B}s)^Ts}{s^Tv}\frac{vv^T}{s^Tv}\right) \\ &= T+\frac{(\tilde{z}-Ts)v^T+v(\tilde{z}-Ts)^T}{s^Tv}-\frac{(\tilde{z}-Ts)^Ts}{s^Tv}\frac{vv^T}{s^Tv} \end{aligned} \tag{5.15}$$

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

$$T_+ = T + \frac{(\tilde{z} - Ts)(\tilde{z} - Ts)^T}{s^T(\tilde{z} - Ts)}. \tag{5.16}$$

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

Now we are in a position to present numerical experiments with various methods for nonlinear least squares problems (see also the results in [132]). The table below shows a computational comparison of the standard BFGS method, the standard Gauss-Newton method, the Gauss-Newton method with the cumulative BFGS correction and the switching strategy (5.6), the Gauss-Newton method with the structured SR1 update (5.7) and either the switching strategy (5.6) or the permanent realization and the Gauss-Newton method with the totally structured SR1 update (5.16) and either the switching strategy (5.6) or the permanent realization. The first part of Table 4 correponds to the line search implementation and the second part corresponds to the trust region implementation. Computational experiments have been obtained using 115 test problems with $2 \sim 20$ variables.

line search realization	main iter.	f. eval.	g. eval.	CPU time	failures
standard BFGS	6197	8459	8459	20.71	4
standard GN	6327	11616	17942	46.36	9
GN/(2.16) and (5.6)	2231	3764	5991	20.32	1
GN/(5.7) and (5.6)	2155	3850	6003	25.76	-
GN/(5.16) and (5.6)	2201	3954	6153	24.33	-
GN/(5.7)	6005	8997	15001	43.66	5
GN/(5.16)	5238	7850	13086	37.68	3
trust region realization	main iter.	f. eval.	g. eval.	CPU time	failures
standard BFGS	6957	8123	7070	23.62	3
standard GN	3052	3513	3165	22.68	1
GN/(2.16) and (5.6)	2164	2553	2728	16.70	1
GN/(5.7) and (5.6)	2127	2570	2240	17.91	1
GN/(5.16) and (5.6)	2166	2593	2278	17.24	1
GN/(5.7)	5618	6293	5731	39.27	3
GN/(5.16)	4344	5100	4458	27.90	2

Table 4

The Gauss-Newton method can be combined with variable metric updates also in the sparse case. In the sequel, we will describe some of such possibilities. One of them is a combination of the Gauss-Newton method with the update (3.2), so that

$$\begin{aligned} B_+ &= \mathcal{P}_S \mathcal{P}_{QG}(J_+^T J_+) \quad \text{if} \quad F - F_+ \le \bar{\theta}_1 F. \\ B_+ &= J_+^T J_+ \quad \text{if} \quad F - F_+ > \bar{\theta}_1 F, \end{aligned} \tag{5.17}$$

Computational efficiency of this hybrid method was studied in [85].

An interesting approach, based on the partitioned SR1 update, was proposed in [140] and also studied in [85]. The partitioned SR1 update is applied to the approximations $\hat{T}_k$ of the packed Hessian matrices $\hat{G}_k(x)$ of the functions $f_k : \mathcal{R}^n \to \mathcal{R}$, $1 \le k \le m$, contained in (5.1). These matrices are updated in such a way that

$$\hat{T}_k^+ = \hat{T}_k + \frac{(\hat{y}_k - \hat{T}_k\hat{s}_k)(\hat{y}_k - \hat{T}_k\hat{s}_k)^T}{\hat{s}_k^T(\hat{y}_k - \hat{T}_k\hat{s}_k)} \quad \text{if} \quad |\hat{s}_k^T(\hat{y}_k - \hat{T}_k\hat{s}_k)| > \bar{\theta}_0, \tag{5.18}$$
$$\hat{T}_k^+ = \hat{T}_k \quad \text{if} \quad |\hat{s}_k^T(\hat{y}_k - \hat{T}_k\hat{s}_k)| \le \bar{\theta}_0.$$

and are used for construction of approximations $\hat{B}_k$ of the packed element Hessian matrices $\hat{g}_k\hat{g}_k^T + f_k\hat{G}_k$. Using (5.6), we can write

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

In the first iteration we set $\hat{T}_k = I$, $1 \le k \le m$. Notice that the matrices $\hat{T}_k^+$, $1 \le k \le m$, 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 (5.13), which is used as a simple update. Taking $L = 0$, we can express $A_+ = J_+ + L_+$ in the form

$$\begin{aligned} A_+ &= J_+ + \frac{J_+ s}{s^T J_+^T J_+ s}\left(\sqrt{\frac{s^T J_+^T J_+ s}{s^T y}}\, y - J_+^T J_+ s\right)^T \\ &= J_+ + \frac{J^+ s}{\|J^+ s\|}\left(\frac{y}{\sqrt{s^T y}} - J_+^T \frac{J^+ s}{\|J^+ s\|}\right)^T, \end{aligned} \tag{5.19}$$

Then we can use the matrix (5.19) for the direction determination, if $F - F_+ \le \bar{\theta}_1 F$, and the matrix $A_+ = J_+$, otherwise. The expression (5.19) is not appropriate, when we want to utilize the least squares solution of the overdetermined system $Ad = -f$. In this case, the formula

$$\begin{aligned} A_+ &= J_+ + \frac{P J_+ s}{s^T J_+^T P J_+ s}\left(\sqrt{\frac{s^T J_+^T P J_+ s}{s^T \tilde{y}}}\, \tilde{y} - J_+^T P J_+ s\right)^T \\ &= J_+ + \frac{\tilde{v}}{\|\tilde{v}\|}\left(\frac{\tilde{y}}{\sqrt{s^T \tilde{y}}} - J_+^T \frac{\tilde{v}}{\|\tilde{v}\|}\right)^T, \end{aligned} \tag{5.20}$$

where

$$\tilde{v} = P J_+ s = J_+ s - \frac{f_+^T J_+ s}{f_+^T f_+} f_+,$$

$$\tilde{y} = y - \frac{f_+^T J_+ s}{f_+^T f_+} J_+^T f_+,$$

is satisfactory.

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

$$\tilde{J}_+ = \begin{bmatrix} J_+ \\ w \end{bmatrix}, \; \tilde{f}_+ = \begin{bmatrix} f_+ \\ 0 \end{bmatrix}. \tag{5.21}$$

The normal equation of this problem has the form $B_+ d_+ = -J_+^T f_+$, where

$$B^+ = \tilde{J}_+^T \tilde{J}_+ = J_+^T J_+ + w w^T,$$

which together with the choice

$$w = (y - J_+^T J_+ s) / \sqrt{s^T (y - J_+^T J_+ s)} \tag{5.22}$$

gives exactly the SR1 update. Note that (5.22) can be used only if $s^T(y - J_+^T J_+ s) > \bar{\theta}_0$, which slightly restricts the use of the update (5.21). We use the augmented linear least squares problem $\tilde{J}_+ d_+ \approx -\tilde{f}_+$, if $F - F_+ \le \bar{\theta}_1 F$ and $s^T(y - J_+^T J_+ s) > \bar{\theta}_0$ hold simultaneously, and the standard linear least squares problem $J_+ d \approx -f_+$, otherwise.

High efficiency of the sparse hybrid methods (5.17) and (5.18) is demonstrated in Table 3 (SVM/GN and PVM/GN).

6. Variable metric methods for nonsmooth problems. Meanwhile, we have assumed that the objective function is continuously differentiable. Nevertheless, many problems lead to objective functions that do not satisfy this assumption. Analysis and solution of these problems require a special theory, nonsmooth analysis, which is an extensive and fast developed part of mathematics. General theory, including all possible cases, is considerably complicated and it overcomes scope of this contribution. Therefore, we restrict our attention on two special but quite representative problems namely on minimax and convex optimization.

Consider first the minimax optimization. In this case, the objective function $F : \mathcal{R}^n \to \mathcal{R}$ has the form

$$F(x) = \max_{1 \le k \le m} f_k(x), \tag{6.1}$$

where $f_k : \mathcal{R}^n \to \mathcal{R}$, $1 \le k \le m$, are twice continuously differentiable functions. This objective function is not differentiable at the point $x \in \mathcal{R}^n$ if $f_k(x) = f_l(x) = F(x)$ but $g_k(x) \ne g_l(x)$ for some $k \ne l$. The mimimax problem (6.1) can be replaced by an equivalent nonlinear programming problem

$$(\bar{x}, \bar{\xi}) = \arg \min_{(x,\xi) \in \mathcal{F}^{n+1}} \xi, \tag{6.2}$$

where

$$\mathcal{F}^{n+1} = \{(x, \xi) \in \mathcal{R}^{n+1} : f_k(x) \le \xi, \quad 1 \le k \le m\}. \tag{6.3}$$

Although almost all nonlinear programming methods can be used for solving this problem, its special structure brings new possibilities. We restrict our attention on recursive quadratic programming methods. These methods have the basic form (1.1), where the direction vector $d_i \in \mathcal{R}^n$ is obtained by solving the quadratic programming subproblem

$$(6.4) \qquad (d_i, \xi_i) = \arg \min_{(d,\xi)\in\mathcal{F}_i^{n+1}} \left\{ \frac{1}{2} d^T B_i d + \xi \right\},$$

where B_i is a symmetric positive definite matrix and

$$(6.5) \qquad \mathcal{F}_i^{n+1} = \{(d,\xi) \in \mathcal{R}^{n+1} : f_k(x_i) + g_k^T(x_i)d \le \xi, \quad 1 \le k \le m\}$$

This QP subproblem has a special form. First, the function (6.4) is convex and the set (6.5) is nonempty so that the solution (d_i, ξ_i) always exists. Secondly, the Hessian matrix of the function (6.4) (that depends on $(d,\xi) \in \mathcal{R}^{n+1}$) is singular even if B_i is positive definite. Therefore, standard quadratic programming methods can meet difficulties and methods based on a special structure of the QP subproblem are welcome. The dual method proposed in [79] was shown to be very efficient in this case.

After determinimg the direction vector $d_i \in \mathcal{R}^n$, we have to use a line search strategy for stepsize selection. Unfortunately, the standard conditions (1.6)-(1.7) are not suitable for the minimax optimization (the gradients $g(x_i)$ and $g(x_{i+1})$ are not defined in general). Instead of (1.6)-(1.7), we can use the Armijo approach (see [10]). Then $\alpha_i > 0$ is the first member of the sequence $\underline{\beta}^j$, $j \in \mathcal{N}$, $0 < \underline{\beta} < 1$, such that

$$(6.6) \qquad F(x_i + \alpha_i d_i) - F(x_i) \le -\epsilon_1 \alpha^i d_i^T B_i d_i$$

where $0 < \epsilon_1 < 1/2$. It was proved, see e.g. [66], that the above recursive quadratic programming method is globally convergent provided the second order derivatives of $f_k : \mathcal{R}^n \to \mathcal{R}$, $1 \le k \le m$, are bounded on the convex hull of the level set $\mathcal{L}_1 = \{x \in \mathcal{R}^n : F(x) \le F(x_1)\}$ and the condition numbers $\kappa(B_i) = \|B_i\| \, \|B_i^{-1}\|$, $i \in \mathcal{N}$, grows at most linearly (i.e. $\kappa(B_i) \le iC$ for some $C > 0$).

The matrices B_i, $i \in \mathcal{N}$, can be advantageously generated using variable metric updates. In order to assure superlinear rate of convergence, we require these matrices to approximate the Hessian matrices of the Lagrangian function as well as possible. The Lagrangian function at the solution of the original problem has the form $L(\bar{x}, \bar{u}) = \bar{u}^T f(\bar{x})$, where

$$\begin{aligned} F(\bar{x})e - f(\bar{x}) &\ge 0, & \bar{u} &\ge 0, \\ \bar{u}^T(F(\bar{x})e - f(\bar{x})) &= 0, & \bar{u}^T e &= 1 \end{aligned}$$

and $J^T(\bar{x})\bar{u} = 0$. ($\bar{u} \in \mathcal{R}^m$ is the Lagrange multiplier vector, $e \in \mathcal{R}^m$ is the vector of unit elements and $J(\bar{x})$ is the Jacobian matrix introduced in the previous section). Let $u_i \in \mathcal{R}^m$ be the Lagrange multiplier vector at the solution of i-th QP subproblem so that

$$\begin{aligned} \xi_i e - f_i - J_i d_i &\ge 0, & u_i &\ge 0, \\ u_i^T(\xi_i e - f_i - J_i d_i) &= 0, & u_i^T e &= 1, \end{aligned}$$

where $B_i d_i = J_i^T u_i$ (here $f_i = f(x_i)$ and $J_i = J(x_i)$). If $x_i \to \bar{x}$, then $u_i \to \bar{u}$ so that u_i can be used as a good approximation of $\bar{u}$. For this reason, we can utilize the quasi-Newton condition

$$(6.7) \qquad B_i(x_{i+1} - x_i) = J^T(x_{i+1})u_i - J^T(x_i)u_i,$$

which is satisfied by any update of the form (2.13) with $s_i = x_{i+1} - x_i$ and $y_i = J^T(x_{i+1})u_i - J^T(x_i)u_i$. However, one difficulty arises, namely that the condition $s_i^T y_i > 0$ cannot be insured by a suitable line search termination criterion. This is a consequence of the fact that the line search is related to the objective function $F(x)$ while the quasi-Newton condition contains the gradients of the Lagrangian function $L(x,u)$. There are two possibilities for overcoming this difficulty. First, the variable metric update may not be used if the condition $s_i^T y_i > 0$ is not satisfied. The second way is based on the correction proposed in [107]. In this case, we choose λ, $0 < \lambda < c/(c-b)$, and replace y and b by $\lambda y + (1-\lambda)Bs$ and $\lambda b + (1-\lambda)c$ in (2.13), respectively.

The above approach can be generalized on more complicated so-called nonsmooth composite functions, which have the form $F(x) = \bar{F}(f(x))$, where $\bar{F} : \mathcal{R}^m \to \mathcal{R}$ is a convex polyhedral function and $f : \mathcal{R}^n \to \mathcal{R}^m$ is a twice continuously differentiable mapping (see [48]). Then the special QP subproblem (6.4)-(6.5) has to be replaced by another QP subproblem taking into account structure of the polyhedral function.

Consider now that the objective function $F : \mathcal{R}^N \to \mathcal{R}$ is convex on the convex hull $\hat{\mathcal{L}}_1$ of the level set $\mathcal{L}_1 = \{x \in \mathcal{R}^n : F(x) \le F(x_1)\}$, i.e. $F(\lambda x + (1-\lambda)y) \le \lambda F(x) + (1-\lambda)F(y)$ for all $x \in \hat{\mathcal{L}}_1$, $y \in \hat{\mathcal{L}}_1$ and $0 \le \lambda \le 1$. It is well known (see e.g. [68]) that the convex function F is continuous and differentiable almost everywhere in the interior of $\hat{\mathcal{L}}_1$. Moreover, the nonempty compact convex set $\partial F(x) = \{g \in \mathcal{R}^n : F(y) - F(x) \ge g^T(y-x)\ \forall y \in \hat{\mathcal{L}}_1\}$, the so-called subdifferential, exists at an arbitrary point x, lying in the interior of $\hat{\mathcal{L}}_1$, together with the directional derivative

$$\lim_{t \to 0} \frac{F(x+td) - F(x)}{t} = \max_{g \in \partial F(x)} g^T d$$

(d is also arbitrary). The last equality implies that $x \in \hat{\mathcal{L}}_1$ is a minimizer of F if and only if $0 \in \partial F(x)$.

The most sophisticated methods for nonsmooth convex optimization are based on the so-called bundle principle. Instead of the singleton $F_i = F(x_i)$, $g_i \in \partial F(x_i)$, the bundle $F_j^i = F(z_j) + g_j^T(x_i - z_j)$, $g_j^i = g_j \in \partial F(z_j)$, $j \in \mathcal{J}_i \subset \{1, \dots i\}$, $\mathcal{J}_i \cap \{i\} \ne \emptyset$, is used in every iteration. The piecewise linear function

$$\check{F}(x) = \max_{j \in \mathcal{J}_i} \{F(z_j) + g_j^T(x - z_j)\} = \max_{j \in \mathcal{J}_i} \{F(x_i) + (g_j^i)^T(x - x_i) - \beta_j^i\}, \tag{6.8}$$

where $\beta_j^i = F(x_i) - F_j^i$, $j \in \mathcal{J}_i$, are linearization errors, is constructed and the direction vector

$$d_i = \arg\min_{d \in \mathcal{R}^n} \left\{ \check{F}(x_i + d) + \frac{1}{2} d^T B_i d \right\} \tag{6.9}$$

is determined (the additional quadratic term in (6.9) defines the Moreau-Yoshida regularization of $\check{F}(x_i + d)$, see [77], and plays a similar role as a trust region approach). Like the minimax case, we can replace (6.9) by the quadratic programming subproblem

$$(d^i, \xi^i) = \arg\min_{(d,\xi) \in \mathcal{B}_i^{n+1}} \left\{ \frac{1}{2} d^T B_i d + \xi \right\}, \tag{6.10}$$

where

$$\mathcal{B}_i^{n+1} = \{(d, \xi) \in \mathcal{R}^{n+1} : -\beta_j^i + (g_j^i)^T d \le \xi, \quad j \in \mathcal{J}_i\}. \tag{6.11}$$

Once the direction vector $d_i \in \mathcal{R}$ is determined, a special line search procedure which produces either serious or null steps has to be used. Then

$$x_{i+1} = x_i + \alpha_L^i d_i, \quad z_{i+1} = z_i + \alpha_R^i d_i, \tag{6.12}$$

where $0 \le \alpha_L^i \le \alpha_R^i \le 1$. Serious steps, characterized by the relation $\alpha_R^i = \alpha_L^i$, i.e. $z_{i+1} = x_{i+1}$, are typical for smooth optimization methods. In the nonsmooth case, null steps with $\alpha_L^i \ne \alpha_R^i$, i.e. $z_{i+1} \ne x_{i+1}$, are essential, since the bundle information is obtained from a larger domain which can include points lying on opposite sides of a possible discontinuity of subgradients. We refer to [74], [68], [90], or [89] for more details.

Another important part of the bundle method is the bundle update. We can always set $\mathcal{J}_{i+1} = \mathcal{J}_i \cup \{i+1\}$, but this simple strategy leads to unbounded cardinality of the bundle set. To keep this cardinality sufficiently small without a loss of global convergence, the aggregation strategy was introduced in [73]. This strategy consists in replacing the set $\mathcal{J}_i$ by $\mathcal{J}_i \cup \{0\}$ in the i-th QP subproblem, where the quantities with the index 0 are obtained by aggregation of the previous bundle elements. On the other words, if u_i is the Lagrange multiplier vector at the solution of the i-th QP subproblem, then we use the quantities $F_0^{i+1} = f_i^T u_i + (g_0^{i+1})^T (x_{i+1} - x_i)$, $g_0^{i+1} = J_i^T u_i$ and $\beta_0^{i+1} = F(x_{i+1}) - F_0^{i+1}$ in the next iteration, where f_i^T contains elements F_j^i, $j \in \mathcal{J}_i \cup \{0\}$ and J_i^T contains columns g_j^i, $j \in \mathcal{J}_i \cup \{0\}$. If the above aggregation is used, then we can delete, from $\mathcal{J}_{i+1}$, some indices different from $i+1$.

The last question, which has to be answered in connection with the bundle approach, is the choice of matrices B_i, $i \in \mathcal{N}$. The most popular proximal bundle methods [74], [68], [90] are based on the choice $B_i = \mu_i I$ where μ_i, $i \in \mathcal{N}$ are weighting coefficients. These methods require only $O(n)$ operations for solving a system with the matrix B_i so that they are most efficient measured by the computational time. Another possibility is to use aggregated Hessian matrices. The resulting so-called bundle-Newton method [89] is very efficient measured by the number of iterations and function evaluations, but it requires $O(n^3)$ operations for solving a system with the matrix B_i. A natural idea is to generate matrices B_i, $i \in \mathcal{N}$, by using variable metric updates of the form (2.13). However, the situation is more complicated, in comparison with the minimax optimization, since a suitable Lagrangian function, defining the vectors y_i, $i \in \mathcal{N}$, does not exist. The simple choices $y_i = g_{i+1} - g_i$ or $y_i = g_0^{i+1} - g_0^i$ lead to methods which do not overcome the efficiency of proximal bundle methods (see e.g. [76]). The most significant results are presented in [77], where reversal quasi-Newton updates together with a special curvilinear search procedure are used. The resulting method is superlinearly convergent in particular cases. Notice that the quasi-Newton condition can be also used for determining weight coefficients μ_i, $i \in \mathcal{N}$, in proximal bundle methods, see [87].

The main deficiency of standard bundle methods is the necessity of solving rather extensive QP subproblem in every iteration, which is a time consuming procedure. On the other hand, standard variable metric methods described in Section 2 are relatively robust and efficient when they are applied to nonsmooth convex problems (see e.g. [76]). This fact indicates that special nonsmooth modifications of variable metric methods, not containing time consuming operations, could be developed. Roughly speaking, three basic ideas of bundle methods can be applied to variable metric methods for improving their efficiency and robustness. The essential feature is the utilization of null steps that serve for obtaining a sufficient information about a convex nondifferentiable function. Another approach consists in a simple aggregation of subgradients and application of linearization errors that guarantees convergence of subgradients to zero and allows us to evaluate a termination criterion. A globally convergent variable metric method based on these ideas is proposed in [88]. This method which utilizes a simple two-term aggregation in null steps can be described by the following procedure.

Starting with x_1, $F(x_1)$, $g_1 \in \partial F(x_1)$, H_1 positive definite (e.g. $H_1 = I$), $\tilde{g}_1 = g_1$, $\tilde{\beta}_1 = 0$, the i-th iteration begins with determination of the direction vector $d_i = -H_i \tilde{g}_i$ and computation of the stationarity measure $w_i = -\tilde{g}_i^T d_i + 2\nu \tilde{\beta}_i$, where $0 < \nu \leq 1$. If w_i is sufficiently small, then x_i is a good approximation of the minimizer. Otherwise, a steplength α_i is selected, e.g. using a piecewise linear approximation, together with $z_{i+1} = x_i + \alpha_i d_i$, $F(z_{i+1})$ and $g_{i+1} \in \partial F(z_{i+1})$. If

$$F(z_{i+1}) - F(x_i) \leq -\epsilon_1 \alpha_i w_i \tag{6.13}$$

(serious step), then we set $x_{i+1} = z_{i+1}$, $\tilde{g}_{i+1} = g_{i+1}$, compute $s_i = x_{i+1} - x_i$, $y_i = g_{i+1} - g_i$, determine H_{i+1} from H_i by the BFGS update and pass to the next iteration. If (6.13) is not satisfied (null step), then we compute $\beta_{i+1} = (F(x_i) - F(z_{i+1}))/\alpha_i + d_i^T g_{i+1}$, minimize the function

$$\phi_i(\lambda) = (\lambda g_{i+1} + (1-\lambda)\tilde{g}_i)^T H_i (\lambda g_{i+1} + (1-\lambda)\tilde{g}_i) + 2\omega(\lambda \beta_{i+1} + (1-\lambda)\tilde{\beta}_i)$$

with respect to λ and set

$$\tilde{g}_{i+1} = \lambda g_{i+1} + (1-\lambda)\tilde{g}_i, \quad \tilde{\beta}_{i+1} = \lambda \beta_{i+1} + (1-\lambda)\tilde{\beta}_i.$$

After this simple aggregation we compute $s_i = z_{i+1} - x_i$, $y_i = g_{i+1} - g_i$. If $\tilde{g}_i^T(s_i - H_i y_i) > 0$, then we construct H_{i+1} from H_i by the SR1 update. Finally, we go to the next iteration.

Although the method described in [88] is slightly more complicated, including more technical details, the above procedure is quite representative.

7. Conclusion. In this paper, we have given a review of variable metric or quasi-Newton methods for unconstrained optimization. Quasi-Newton methods can be also used for solving systems of nonlinear equations, see e.g. [16], [43], [91], 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 overcome the scope of this contribution.

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. JOTA 7 (1971) 391-409.

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

[4] M.Al-Baali: Variational quasi-Newton methods for unconstrained optimization. JOTA 77 (1993) 127-143.

[5] M.Al-Baali: An efficient class of switching type algorithms in the Broyden family. Optimization Methods and Software 4 (1994) 29-46.

[6] M.Al-Baali: On measure functions for the self-scaling updating formulae for quasi-Newton methods. Optimization 32 (1995) 59-69.

[7] M.Al-Baali, R.Fletcher: Variational methods for nonlinear least squares. JOTA 36 (1985) 405-421.

[8] K.A.Ariyawansa: Deriving collinear scaling algorithms as extension of quasi-Newton methods and the local convergence of DFP and BFGS related collinear scaling algorithms. Math Programming 49 (1990) 23-48.

[9] K.A.Ariyawansa, D.T.M.Lau: Local and Q-superlinear convergence of a class of collinear scaling algorithms that extends quasi-Newton methods with Broyden's bounded-Ψ class of updates. Optimization 23 (1992) 323-339.

[10] L.Armijo: Minimization of functions having continuous partial derivatives. Pacific J. Math. 16 (1966) 1-3.

[11] O.Axelsson: Iterative solution methods. Cambridge University Press, Cambridge 1996.

[12] J.T.Betts: Solving the nonlinear least squares problem: Application of a general method. JOTA 18 (1976) 469-483.

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

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

[15] K.W.Brodlie, A.R.Gourlay, J.Greenstadt: Rank-one and rank-two corrections to positive definite matrices expressed in product form. J. Inst. Maths. Applics. 11 (1973), 73-82.

[16] C.G.Broyden: A class of methods for solving nonlinear simultaneous equations. Math. of Comput. 19 (1965) 577-593.

[17] 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.

[18] K.M.Brown, J.E.Dennis: A new algorithm for nonlinear least squares curve fitting. In: "Mathematical Software" (J.Rice ed.) Academic Press, London 1971.

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

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

[21] 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.

[22] 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.

[23] 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.

[24] R.H.Byrd, J.Nocedal, C.Zhu: Towards a discrete Newton method with memory for large-scale optimization. In "Nonlinear Optimization and Applications" (G.Di Pillo, G.Giannessi, eds.) Plenum Press, 1996.

[25] R.H.Byrd, R.B.Schnabel, G.A.Shultz: Approximate solution of the trust region problem by minimization over two-dimensional subspaces. Math. Programming 40 (1988) 247-263.

[26] A.R. Conn, N.I.M. Gould, P.L. Toint: Testing a class of methods for solving minimization problems with simple bounds on the variables. Math. Comput. 50 (1988) 399-430.

[27] M. Contreras, R.A. Tapia: Sizing the BFGS and DFP updates: a numerical study. JOTA 78 (1993) 93-108.

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

[29] W.C.Davidon: Optimally conditioned optimization algorithms without line searches. Math. Programming 9 (1975) 1-30.

[30] W.C.Davidon: Conic approximations and collinear scalings for optimizers. SIAM J. Numer. Anal. 17 (1980) 268-281.

[31] W.C.Davidon: Conjugate directions for conic functions. In: "Nonlinear Optimization 1981" (M.J.D.Powell ed.), Academic Press, New York (1992) 23-28.

[32] M.J.Daydé, J.Y.L'Excellent, N.I.M.Gould: On the use of element-by-element preconditioners to solve large scale partially separable optimization problems. Report No. RAL 95-010, Central Computing Department, Atlas Centre, Rutherford Appleton Laboratory, 1995.

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

[34] N.Y. Deng, Z.F. Li: Nonquadratic model method in unconstrained optimization. In: "Algorithms for Continuous Optimization. The State of the Art" (E. Spedicato ed.), Kluwer, Amsterdam, 1994.

[35] N.Y.Deng, Z.F.Li: Some global convergence properties of a conic-variable metric algorithm for minimization with inexact line searches. Optimization Methods and Software 5 (1995) 105-122.

[36] N.Y.Deng, Y.Xiao, F.J.Zhou: Nonmonotonic trust region algorithm. JOTA 76 (1993) 259-285.

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

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

[39] J.E.Dennis, H.J.Martinez, R.A.Tapia: Convergence theory for the structured BFGS secant method with application to nonlinear least squares. JOTA 61 (1989) 161-177.

[40] 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.

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

[42] 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.

[43] J.E.Dennis, R.B.Schnabel: Numerical methods for unconstrained optimization and nonlinear equations. Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

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

[45] J.E.Dennis, R.E.Welsch: Techniques for nonlinear least squares and robust regression. Communications in Statistics B 7 (1978) 345-359.

[46] J.E.Dennis, H.Wolkowicz: Sizing and least-change secant methods. SIAM J. Numer. Anal. 30 (1993) 1291-1314.

[47] R.Fletcher: A new approach to variable metric algorithms. Computer J. 13 (1970) 317-322.

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

[49] R.Fletcher: A new variational result for quasi-Newton formulae. SIAM J. Optimization 1 (1991) 18-21.

[50] R.Fletcher: An optimal positive definite update for sparse Hessian matrices. SIAM J. Optimization 5 (1995) 192-218.

[51] R.Fletcher, M.J.D.Powell: A rapidly convergent descent method for minimization. Computer J. 6 (1963) 163-168.

[52] R.Fletcher, C.M.Reeves: Function minimization by conjugate gradients. Computer J. 7 (1964) 149-154.

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

[54] J.Greenstadt: Variations on variable metric methods. Math. Comput. 24 (1970) 1-18.

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

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

[57] P.E.Gill, W.Murray, M.H.Wright: Practical Optimization. Academic Press, London 1981.
[58] D.Goldfarb: A family of variable metric algorithms derived by variational means. Math Comput. 24 (1970) 23-26.
[59] G.H.Golub, C.F.Van Loan: Matrix Computations. Johns Hopkins University Press, Baltimore, 1989.
[60] H.Gourgeon, J.Nocedal: A conic algorithms for optimization. SIAM J. Sci. Stat. Comput. 6 (1985) 253-267.
[61] A.Griewank: The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients. Math. Programming 50 (1991) 141-175.
[62] A.Griewank, P.L.Toint: Partitioned variable metric updates for large scale structured optimization problems. Numer. Math. 39 (1982) 119-137.
[63] A.Griewank, P.L.Toint: Local convergence analysis for partitioned quasi-Newton updates. Numer. Math. 39 (1982) 429-448.
[64] L.Grippo, F.Lampariello, S.Lucidi: A nonmonotone line search technique for Newton's method. SIAM J. Numer. Anal. 23 (1986) 707-716.
[65] L.Grippo, F.Lampariello, S.Lucidi: A class of nonmonotone stabilization methods in unconstrained optimization. Numer. Math. 59 (1991) 779-805.
[66] S.P.Han: Variable metric methods for minimizing a class of nondifferentiable functions. Math. Programming 20 (1981) 1-13.
[67] M.R.Hestenes, C.M.Stiefel: Methods of conjugate gradient for solving linear systems. J. Res. NBS 49 (1964) 409-436.
[68] J.B. Hiriart-Urruty, C.Lemarechal: Convex Analysis and Minimization Algorithms I and II. Springer, Berlin, 1993.
[69] R.Hooke, T.A.Jeeves: Direct search solution of numerical and statistical problems. J. Assoc. Comp. Mach. 8 (1961) 212-221.
[70] S.Hoshino: A formulation of variable metric methods. J. Inst. Math. Appl. 10 (1972) 394-403.
[71] Y.F.Hu, C.Storey: Motivating quasi-Newton updates by preconditioned conjugate gradient methods. Report No. A150, Dept. of Math. Sci., Loughborough University of Technology, Loughborough, 1991.
[72] J.Huschens: On the use of product structure in secant methods for nonlinear least squares. SIAM J. Optimization 4 (1994) 108-129.
[73] K.C. Kiwiel: An aggregate subgradient method for nonsmooth convex minimization. Math. Programming, 27 (1983) 320-341.
[74] K.C. Kiwiel: Methods of Descent for Nondifferentiable Optimization. Lecture Notes in Mathematics 1133, Springer-Verlag, Berlin, 1985.
[75] H.Kleinmichel: Quasi-Newton Verfahren vom Rang-Eins-Typ zur Loesung unrestringierter Minimierungsprobleme. Teil 1: Verfahren und grundlegende Eigenshafte. Numer. Math. 38 (1981) 219-228
[76] C. Lemarechal: Numerical experiments in nonsmooth optimization. Proc. IIASA workshop "Progress in Nondifferentiable Optimization", December 1978.
[77] C. Lemarechal, C. Sagastizábal: Variable metric bundle methods: From conceptual to implementable forms. Math. Programming, 76 (1997) 393-410.
[78] D.C.Liu, J.Nocedal: On the limited memory BFGS method for large-scale optimization. Math. Programming 45 (1989) 503-528.
[79] L.Lukšan: Dual method for solving a special problem of quadratic programming as a subproblem at linearly constrained nonlinear minimax approximation. Kybernetika 20 (1984) 445-457.
[80] L.Lukšan: Conjugate direction algorithms for extended conic functions. Kybernetika 22 (1986) 31-46.
[81] L.Lukšan: Variable Metric Methods. Academia, Prague, 1990.
[82] L.Lukšan: Computational experience with improved variable metric methods for unconstrained minimization. Kybernetika 26 (1990) 415-431.
[83] L.Lukšan: Variationally derived scaling and variable metric updates from the preconvex part of the Broyden family. JOTA 73 (1992) 299-307.
[84] L.Lukšan: Computational experience with known variable metric updates. JOTA 83 (1994) 27-47.
[85] L.Lukšan: Hybrid methods for large sparse nonlinear least squares. JOTA 89 (1996) 575-595.
[86] 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.
[87] L.Lukšan, J.Vlček: PBUN, PNEW - a bundle type algorithms for nonsmooth optimization.

Report No. 718, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague 1997.

[88] L.Lukšan, J.Vlček: Globally convergent variable metric method for convex nonsmooth unconstrained minimization. Report No. 741, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague 1998.

[89] L.Lukšan, J.Vlček: A bundle-Newton method for nonsmooth unconstrained minimization. To appear in Mathematical Programming.

[90] M.M. Mäkelä, P.Neittaanmäki: Nonsmooth Optimization. World Scientific Publishing Co., London, 1992.

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

[92] E.S.Marwil: Exploiting sparsity in Newton-like methods. Ph.D. Thesis, Cornell University, Ithaca 1978.

[93] R.B.Mifflin, J.L.Nazareth: The least-prior deviation quasi-Newton update. Technical Report, Dept. of Pure and Applied Math., Washington State University, Pullman 1991.

[94] J.J.Moré, B.S.Garbow, K.E.Hillström: Testing unconstrained optimization software. ACM Trans. Math. Software 7 (1981) 17-41.

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

[96] J.A.Nelder, R.Mead: A simplex method for function minimization. Computer J. 7 (1965) 308-313.

[97] J.Nocedal: Updating quasi-Newton matrices with limited storage. Math. Comput. 35 (1980) 773-782.

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

[99] J.Nocedal, Y.Yuan: Combining trust region and line search techniques. In: "Advances in Nonlinear Programming" (Y.Yuan ed.), Kluwer, Amsterdam, 1998.

[100] 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.

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

[102] 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.

[103] A.Perry: A modified conjugate gradient algorithm. Oper. Research 26 (1978) 1073-1078.

[104] E.Polak, G.Ribiére: Note sur la convergence des méthodes de directions conjugées. Revue Francaise Inform. Mech. Oper. 16-R1(1969) 35-43.

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

[106] M.J.D.Powell: Convergence properties of a class of minimization algoritms. In: "Nonlinear Programming 2" (O.L.Mangasarian, R.R.Meyer, S.M.Robinson eds.). Academic Press, London, 1975.

[107] M.J.D.Powell: A fast algorithm for nonlinearly constrained optimization calculations. In: "Numerical analysis" (G.A.Watson ed.). Springer Verlag, Berlin 1977.

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

[109] M.J.D.Powell: Updating conjugate directions by the BFGS formula. Math. Programming 38 (1987) 29-46.

[110] D.G.Pu, W.W.Tian: A class of modified Broyden algorithms. J. of Computational Mathematics 12 (1994) 366-379.

[111] H.Ramsin, P.A.Wedin: A comparison of some algorithms for the nonlinear least squares problem. BIT 17 (1977) 72-90.

[112] L.K.Schubert: Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian. Math. of Comput. 24 (1970) 27-30.

[113] D.F.Shanno: Conditioning of quasi-Newton methods for function minimization. Math. Comput. 24 (1970) 647-656.

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

[115] S.Sheng, Z.Zou: A new secant method for nonlinear least squares problems. Report NANOG-

1988-03, Nanjing University, China 1988.
[116] D.Siegel: Updating of conjugate direction matrices using members of Broyden family. Math. Programming 60 (1993) 167-185.
[117] D.Siegel: Modifying the BFGS update by a new column scaling technique. Math. Programming 66 (1993) 45-78.
[118] F.Sloboda: A generalized conjugate gradient algorithm for minimization. Numer. Math. 35 (1980) 223-230.
[119] D.C.Sorensen: The Q-superlinear convergence of a collinear algorithm for unconstrained optimization. SIAM J. Numer. Anal. 17 (1980) 84-114.
[120] E.Spedicato: Problems related to the solution of the quasi-Newton equation. Report No. CISE-N-147, CISE - Documentation Service, Milano, 1971.
[121] E.Spedicato: Stability of Huang's update for the conjugate gradient method. JOTA 11 (1973) 469-479
[122] E.Spedicato: On condition numbers of matrices in rank-two minimization algorithms. In: "Towards Global Optimization" (L.C.W. Dixon, G.P.Szego eds.), North Holland, Amsterdam, 1975
[123] E.Spedicato: A bound on the condition number of rank-two corrections and applications to the variable metric method. Calcolo 12 (1975) 185-199.
[124] E.Spedicato: A variable-metric method for function minimization derived from invariancy to nonlinear scaling. JOTA 20 (1976) 315-329.
[125] E.Spedicato: On a conjecture of Dixon and other topics in variable metric methods. Math. Programming 15 (1978) 123-129.
[126] E.Spedicato: A note on the determination of the scaling parameters in a class of quasi-Newton methods for unconstrained minimization. J. Inst. Maths. Applics. (1978) 285-291.
[127] E.Spedicato: Parameter estimation and least squares. In: "Numerical Techniques for Stochastic Systems" (F.Archetti and M. Cugiani eds.), North Holland, Amsterdam, 1980
[128] E.Spedicato: A class of rank-one positive definite quasi-Newton updates for unconstrained minimization. Math. Operationsforsch. Statist. Ser. Optimization 14 (1983) 61-70.
[129] E. Spedicato: A class of sparse symmetric Quasi-Newton updates. Ricerca Operativa 22 (1992) 63-70.
[130] E. Spedicato, N.Y. Deng, L. Zhengfeng: On sparse Quasi-Newton quasi-diagonally dominant updates, Report DMSIA 1/96, University of Bergamo, 1996.
[131] E. Spedicato, Z. Huang: Numerical experience with Newton-like methods for nonlinear algebraic systems. Computing 58 (1997) 69-90.
[132] E. Spedicato, M.T. Vespucci, Numerical experiments with variations of the Gauss-Newton algorithm for nonlinear least squares. JOTA 57 (1988) 323-339.
[133] E. Spedicato, Z. Xia: Finding general solutions of the Quasi-Newton equation via the ABS approach. Optimization Methods and Software 1 (1992) 243-252.
[134] E. Spedicato, J. Zhao: Explicit general solution of the Quasi-Newton equation with sparsity and symmetry, Optimization Methods and Software 2 (1993) 311-319.
[135] T.Steihaug: Local and superlinear convergence for truncated iterated projections methods. Math. Programming 27 (1983) 176-190.
[136] T.Steihaug: The conjugate gradient method and trust regions in large-scale optimization. SIAM J. Numer. Anal. 20 (1983) 626-637.
[137] S.W. Thomas: Sequential estimation techniques for quasi-Newton algorithm, Report TR 75-227, Cornell University, 1975.
[138] P.L.Toint: On sparse and symmetric matrix updating subject to a linear equation. Math. of Comp. 31 (1977) 954-961.
[139] P.L.Toint: Global convergence of the partitioned BFGS algorithm for convex partially separable optimization. Math. Programming 36 (1986) 290-306.
[140] P.L.Toint: On large scale nonlinear least squares calculations. SIAM J. Sci. Stat. Comput. 8 (1987) 416-435.
[141] P.L.Toint: An assessment of non-monotone linesearch techniques for unconstrained optimization. Report No. 94/14, Facultées Universitaires de Namur, Namur 1994.
[142] M.Tuma: Sparse fractioned variable metric updates. Report No. 497, Institute of Computer and Information Sciences, Czechoslovak Academy of Sciences, Prague 1991.
[143] H.Wolkowicz: Measures for rank-one updates. Math. of Operations Research 19 (1994) 815-830.
[144] H.Wolkowicz: An all-inclusive efficient region of updates for least change secant methods. SIAM J. Optimization 5 (1996) 172-191.

[145] Y. Xie: A new variable metric formulation. Manuscript, Institute of Applied Mathematics, Academia Sinica, Beijing, 1996.

[146] H.Yabe: Variations of structured Broyden families for nonlinear least squares problems. Optimization Methods and Software 2 (1993) 107-144.

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

[148] Y.Yuan: On self-dual update formulae in the Broyden family. Optimization Methods and Software 1 (1992) 117-127.

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

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

[151] Y.Zhang, R.P.Tewarson: Least-change updates to Choleski factors subject to nonlinear quasi-Newton condition. IMA J. Numer. Anal. 7 (1987) 509-521.

[152] Y.Zhang, R.P.Tewarson: Quasi-Newton algorithms with updates from the preconvex part of Broyden's family. IMA J. Numer. Anal. 8 (1988) 487-509.

[153] X.Zhou, I.M.Navon, M.Berger, K.H.Phua, T.Schlick, F.X LeDimet: Numerical experience with limited-memory quasi-Newton and Truncated Newton methods. SIAM J. Optimization 3 (1993) 582-606.