J. Oper. Res. Soc. China (2014) 2:39–56
DOI 10.1007/s40305-014-0039-x

Analysis of Sparse Quasi-Newton Updates with Positive Definite Matrix Completion

Yu-Hong Dai · Nobuo Yamashita

Received: 24 January 2014 / Revised: 17 February 2014 / Accepted: 18 February 2014 / Published online: 20 March 2014

Abstract Based on the idea of maximum determinant positive definite matrix completion, Yamashita (Math Prog 115(1):1–30, 2008) proposed a new sparse quasi-Newton update, called MCQN, for unconstrained optimization problems with sparse Hessian structures. In exchange of the relaxation of the secant equation, the MCQN update avoids solving difficult subproblems and overcomes the ill-conditioning of approximate Hessian matrices. However, local and superlinear convergence results were only established for the MCQN update with the DFP method. In this paper, we extend the convergence result to the MCQN update with the whole Broyden's convex family. Numerical results are also reported, which suggest some efficient ways of choosing the parameter in the MCQN update the Broyden's family.

Keywords Quasi-Newton method · Large-scale problems · Sparsity · Positive definite matrix completion · Superlinear convergence

This work was supported by the Chinese NSF Grants (Nos. 11331012 and 81173633), the China National Funds for Distinguished Young Scientists (No. 11125107) and the CAS Program for Cross & Coorperative Team of the Science & Technology Innovation.

Y.-H. Dai (✉)
State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100080, People's Republic of China
e-mail: dyh@lsec.cc.ac.cn

N. Yamashita
Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan
e-mail: nobuo@i.kyoto-u.ac.jp

1 Introduction

Consider the unconstrained optimization problem

$$\min f(x), \quad x \in \mathbb{R}^n, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and its gradient ∇f is available. If the dimension n is not large, the quasi-Newton method is an ideal choice for solving (1.1) because of its superlinear convergence and no need to calculate the function Hessian. Assuming that x_k is the current iterate and H_k is the approximation to the inverse Hessian, the quasi-Newton method generates the next iteration by

$$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k), \tag{1.2}$$

where $\alpha_k > 0$ is a stepsize obtained via some line search and updates the approximation H_k to H_{k+1} so that

$$H_{k+1} y_k = s_k, \tag{1.3}$$

where $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$. The update formula in Broyden's family is

$$H_{k+1}^B = H_k - \frac{H_k y_k y_k^{\mathrm{T}} H_k}{y_k^{\mathrm{T}} H_k y_k} + \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \phi_k v_k v_k^{\mathrm{T}}, \tag{1.4}$$

where ϕ_k is a parameter and

$$v_k = \sqrt{y_k^{\mathrm{T}} H_k y_k} \left(\frac{s_k}{s_k^{\mathrm{T}} y_k} - \frac{H_k y_k}{y_k^{\mathrm{T}} H_k y_k} \right). \tag{1.5}$$

The choice $\phi_k \equiv 0$ is corresponding to the DFP update and $\phi_k \equiv 1$ gives rise to the BFGS update. In addition, the restricted Broyden's family with $\phi_k \in [0, 1]$ is called as Broyden's convex family.

If the dimension of the problem (1.1) is large, then the direct use of the quasi-Newton method is not possible due to the storage of an $n \times n$ matrix. In order to overcome this difficulty, several methods have been proposed. The limited-memory BFGS (L-BFGS) method [13, 15] is to only store a few pairs (s_i, y_i) in the construction of the Hessian approximation. Since there is no need to know any information about the Hessian, the L-BFGS method is now widely used in practice. For many large-scale problems, the function f can be written into the form

$$f(x) = \sum_{i=1}^{n_e} f_i(x),$$

where each of the n_e element functions, f_i, depends only on a few variables. In this case, the partitioned quasi-Newton method, developed by Griewank and Toint (see [9–11] and the references therein), performs very well in practice and is now regarded as one of the important practical optimization algorithms. Their basic idea is to update an approximation B_k^i to the Hessian of each element function f_i and then to assemble these matrices to define an approximation B_k to the Hessian of f. More exactly, they determine the search direction by solving the linear system

$(\sum_{i=1}^{n_e} B_k^i)d_k = -\nabla f(x_k)$. Their method was implemented with the trust region strategy since the matrix B_k is not positive definite in general.

There are also many large-scale problems where the function Hessian $\nabla^2 f(x)$ is sparse and the sparsity structure is available. Suppose that for all $x \in \mathbb{R}^n$,

$$[\nabla^2 f(x)]_{i,j} = 0, \quad (i,j) \in F, \tag{1.6}$$

where F is some subset of $I \times I$ and $I = \{1, 2, \cdots, n\}$. In this case, it is possible to establish faster optimization methods by exploiting the sparsity structure of the Hessian. Toint [18] and Fletcher [7] studied such updates and required H_{k+1} to meet the sparsity condition, namely $H_{k+1}(i,j) = 0$ when $(i,j) \in F$, and the secant equation (1.3) simultaneously. As a result, their methods involve the solution of a convex programming problem at each iteration. If some component of s_k is zero, the obtained approximate Hessians may be ill-posed (see Sorensen's [17] example).

Inspired by the successful use of positive definite matrix completion in [9] for semidefinite programming, Yamashita [19] proposed another type of quasi-Newton update for problem (1.1) with sparse Hessian structure. Let $\psi : \mathbb{R}^{n\times n} \to \mathbb{R}$ is a strictly convex function defined by

$$\psi(A) = \mathrm{tr}(A) - \ln\det(A) \tag{1.7}$$

(This function is introduced in [1] as a powerful tool for the convergence analysis of quasi-Newton methods). Yamashita determines the new approximation matrix H_{k+1} by two steps:

(i) update H_k to H^{QN} by certain quasi-Newton formula;
(ii) obtain H_{k+1} by solving the following subproblem

$$\begin{array}{ll} \min & \psi(H_k^{-1/2} H H_k^{-1/2}) \\ \text{s.t.} & H_{ij} = H_{i,j}^{\mathrm{QN}}, \quad (i,j) \in F, \\ & (H^{-1})_{ij} = 0, \quad (i,j) \notin F, \\ & H \in S_+. \end{array} \tag{1.8}$$

Here S_+ denotes the set of symmetric positive definite matrices. As in [20], we call the above update by MCQN. If the intermediate matrix H^{QN} is obtained by the DFP (BFGS) formula, we call the update by MCQN with DFP (BFGS). Similarly, we define MCQN with Broyden's family.

Further, Yamashita showed that, if the sparsity of the Hessian is of such pattern that there is no fill-in in its Cholesky factorization, or equivalently, the graph induced by the Hessian is chordal (see [19] for details), the problem (1.8) is equivalent to finding a maximum-determinant positive definite matrix completion of $H_{ij}^{QN}, (i,j) \in F$:

$$\begin{array}{ll} \max & \det(H) \\ \text{s.t.} & H_{ij} = H_{ij}^{QN}, \ (i,j) \in F, \\ & H \in S_+. \end{array} \tag{1.9}$$

The above problem can then be easily solved by analyzing the clique tree of the graph induced by the Hessian (see [19] for details). In addition, it is shown in [19] that the update does not suffer from Sorensen's example. Therefore, by relaxing the secant equation, the MCQN update is easy to be implemented and is well posed.

Numerical experiments in [19] show that the MCQN update with BFGS obviously performs better than that with DFP. However, local and superlinear convergence results are only established for the latter. The purpose of this paper is to analyze and investigate the MCQN update with Broyden's family, in which the intermediate matrix H^{QN} is given by (1.4). After giving some preliminaries in the next section, we will show in Sect. 3 that the MCQN update with Broyden's convex family is locally and superlinearly convergent under appropriate conditions. Numerical results are reported in Sect. 4 for MCQN with Broyden's positive family and some discussions are made in the last section.

2 Properties of the MCQN Update

We consider the MCQN update with Broyden's convex family, namely, H^{QN} is given by (1.4) with $\phi_k \in [0,1]$. To facilitate our analysis, we introduce some notations at first. For any fixed invertible matrix P with its inverse having the same sparsity pattern with the function Hessian, namely,

$$(P^{-1})_{i,j} = 0, \quad \mathit{forall}\,(i,j) \in F, \tag{2.1}$$

we denote

$$\tilde{s}_k = P^{-1/2}s_k,\ \tilde{y}_k = P^{1/2}y_k,\ \tilde{H}_k = P^{-1/2}H_kP^{-1/2},\ \tilde{H}^B = P^{-1/2}H^BP^{-1/2},$$

where $H^B = H^B_{k+1}$ is given by (1.4). Then, it follows (1.4) that

$$\tilde{H}^B = \tilde{H}_k - \frac{\tilde{H}_k\tilde{y}_k\tilde{y}_k^{\mathrm{T}}\tilde{H}_k}{\tilde{y}_k^{\mathrm{T}}\tilde{H}_k\tilde{y}_k} + \frac{\tilde{s}_k\tilde{s}_k^{\mathrm{T}}}{\tilde{s}_k^{\mathrm{T}}\tilde{y}_k} + \phi_k\tilde{v}_k\tilde{v}_k^{\mathrm{T}}, \tag{2.2}$$

where

$$\tilde{v}_k = \sqrt{\tilde{y}_k^{\mathrm{T}}\tilde{H}_k\tilde{y}_k}\left(\frac{\tilde{s}_k}{\tilde{s}_k^{\mathrm{T}}\tilde{y}_k} - \frac{\tilde{H}_k\tilde{y}_k}{\tilde{y}_k^{\mathrm{T}}\tilde{H}_k\tilde{y}_k}\right). \tag{2.3}$$

Further, assume the Euclidean norm in default and write

$$\tau_k = \frac{\tilde{y}_k^{\mathrm{T}}\tilde{H}_k\tilde{y}_k}{\|\tilde{y}_k\|\|\tilde{H}_k\tilde{y}_k\|}, \quad q_k = \frac{\tilde{y}_k^{\mathrm{T}}\tilde{H}_k\tilde{y}_k}{\|\tilde{y}_k\|^2}, \quad \eta_k = \frac{\tilde{s}_k^{\mathrm{T}}\tilde{H}_k\tilde{y}_k}{\tilde{s}_k^{\mathrm{T}}\tilde{y}_k}, \quad m_k = \frac{\tilde{s}_k^{\mathrm{T}}\tilde{y}_k}{\tilde{y}_k^{\mathrm{T}}\tilde{y}_k},$$
$$M_k = \frac{\|\tilde{s}_k\|^2}{\tilde{s}_k^{\mathrm{T}}\tilde{y}_k}, \quad \beta_k = \frac{\tilde{s}_k^{\mathrm{T}}(\tilde{H}_k)^{-1}\tilde{s}_k}{\tilde{s}_k^{\mathrm{T}}\tilde{y}_k}, \quad \gamma_k = \frac{\tilde{y}_k^{\mathrm{T}}\tilde{H}_k\tilde{y}_k}{\tilde{s}_k^{\mathrm{T}}\tilde{y}_k}.$$

Similar to [2], we can get from (2.2), the following relations

$$\mathrm{tr}(\tilde{H}^B) = \mathrm{tr}(\tilde{H}_k) - (1-\phi_k)\frac{q_k}{\tau_k^2} - 2\phi_k\eta_k + \left(1 + \phi_k\frac{q_k}{m_k}\right)M_k \tag{2.4}$$

and

$$\det(\tilde{H}^B) = \det(\tilde{H}_k)[1 + \phi_k(\beta_k\gamma_k - 1)]/\gamma_k. \tag{2.5}$$

For the MCQN update with Broyden's family, we now establish the following relations between $\tilde{H}_{k+1}$ and $\tilde{H}^B$:

$$\mathrm{tr}(\tilde{H}_{k+1}) = \mathrm{tr}(\tilde{H}^B), \quad \det(\tilde{H}_{k+1}) \geqslant \det(\tilde{H}^B). \tag{2.6}$$

In fact, it follows from (1.8) that $(H_{k+1} - H^B)_{i,j} = 0$ for $(i,j) \in F$. This and (2.1) indicate that for any $(i,j) \in I \times I$, at least one of the elements $(H_{k+1} - H^B)_{i,j}$ and $(P^{-1})_{i,j}$ is equal to zero. Consequently, we have that

$$\begin{aligned}\mathrm{tr}(\tilde{H}_{k+1} - \tilde{H}^B) &= \mathrm{tr}(P^{-1/2}(H_{k+1} - H^B)P^{-1/2}) \\ &= \mathrm{tr}(P^{-1}(H_{k+1} - H^B)) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}(P^{-1})_{ij}((H_{k+1} - H^B)_{i,j} = 0.\end{aligned} \tag{2.7}$$

On the other hand, since the matrix H^B itself satisfies the constraints in (1.9), we must have that $\det(H_{k+1}) \geqslant \det(H^B)$. It follows:

$$\begin{aligned}\det(\tilde{H}_{k+1}) &= \det(P^{-1/2})\det(H_{k+1})\det(P^{-1/2}) \\ &\geqslant \det(P^{-1/2})\det(H^B)\det(P^{-1/2}) \\ &= \det(\tilde{H}^B).\end{aligned} \tag{2.8}$$

Thus, the relations in (2.6) hold. Furthermore, we know by (1.7) and (2.6) that

$$\psi(\tilde{H}_{k+1}) \leqslant \psi(\tilde{H}^B). \tag{2.9}$$

Substituting (2.4) and (2.5) into (2.9), we establish the relation between $\psi(\tilde{H}_{k+1})$ and $\psi(\tilde{H}_k)$:

$$\begin{aligned}\psi(\tilde{H}_{k+1}) \leqslant \psi(\tilde{H}_k) - (1 - \phi_k)\frac{q_k}{\tau_k^2} - 2\phi_k\eta_k + \left(1 + \phi_k\frac{q_k}{m_k}\right)M_k \\ - \ln[1 + \phi_k(\beta_k\gamma_k - 1)] + \ln\gamma_k.\end{aligned} \tag{2.10}$$

Although the above relation does not hold as an equality unlike the ordinary Broyden's family of methods, it suffices us to analyze the local and superlinear convergence of the MCQN update with Broyden's convex family.

The introduction of the matrix P plays an auxiliary role in establishing the relation (2.10). We will choose $P = (\nabla^2 f(x_*))^{-1}$ in our superlinear convergence analysis of the next section, where x_* is the solution point. In the superlinear convergence analysis of the ordinary quasi-Newton method, we can assume that $\nabla^2 f(x_*)$ is the identity matrix due to the invariance property under affine transformations. However, we can show by an example that the MCQN update does not possess the invariance property. Thus, we cannot assume that $\nabla^2 f(x_*)$ is the identity matrix and hence the introduction of P is necessary in the superlinear convergence analysis of the MCQN update. We can only prove that the MCQN update is invariant under those affine transformations which keep the sparsity structure.

Assume that H_k is the following 3×3 matrix

$$H_k = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

and $A = H_k$. Obviously, the matrices A and H_k have the same sparsity pattern. In this case, we can directly calculate the matrix AH_kA^{T} and its corresponding maximum determinant positive definite matrix completion, which is denoted by $MD(AH_kA^{\mathrm{T}})$ and only changes the (1, 3) and (3, 1) entries of AH_kA^{T}:

$$AH_kA^{\mathrm{T}} = \begin{pmatrix} 14 & -14 & 6 \\ -14 & 20 & -14 \\ 6 & -14 & 14 \end{pmatrix},$$
$$MD(AH_kA^{\mathrm{T}}) = \begin{pmatrix} 14 & -14 & 9.8 \\ -14 & 20 & -14 \\ 9.8 & -14 & 14 \end{pmatrix}.$$

On the other side, we have that

$$MD(H_k) = \begin{pmatrix} 2 & -1 & 0.5 \\ -1 & 2 & -1 \\ 0.5 & -1 & 2 \end{pmatrix},$$
$$A[MD(H_k)]A^{\mathrm{T}} = \begin{pmatrix} 14 & 15 & 8 \\ -15 & 21 & -15 \\ 8 & -15 & 14 \end{pmatrix}.$$

The inconsistency of the matrices $MD(AH_kA^{\mathrm{T}})$ and $A[MD(H_k)]A^{\mathrm{T}}$ indicates that the procedure of the maximum determinant positive definite matrix completion does not have the invariance property and neither does the MCQN update.

3 Local and Superlinear Convergence

In this section, we analyze the MCQN update with Broyden's convex family under the sparse structure (1.6) and Assumption 3.1.

Assumption 3.1 Let x_* be a solution of (1.1) and let $\mathcal{B} = \{x \in R^n \mid \|x - x_*\| \leqslant b\}$ with a positive constant b.

(i) The objective function f is twice continuously differentiable on $\mathcal{B}$.
(ii) There exist positive constants m and M such that

$$m\|z\|^2 \leqslant z^{\mathrm{T}}(\nabla^2 f(x))^{-1} z \leqslant M\|z\|^2, \quad \text{for all } z \in \mathbb{R}^n \text{ and } x \in \mathcal{B}. \tag{3.1}$$

If the second-order sufficient optimality condition holds at the solution x_*, then the (ii) in Assumption 3.1 holds. From the (i) in Assumption 3.1, ∇f and $\nabla^2 f(x)$ are Lipschitz continuous on $\mathcal{B}$, namely, there exist some constants $\hat{L}$, $L > 0$ such that

$$\|\nabla f(x) - \nabla f(z)\| \leqslant \hat{L}\|x - z\|, \quad \text{for all } x, z \in \mathcal{B}, \tag{3.2}$$

$$\|\nabla^2 f(x) - \nabla^2 f(z)\| \leqslant L\|x - z\|, \quad \text{for all } x, z \in \mathcal{B}. \tag{3.3}$$

Let us define

$$\varepsilon_k = \max\{\|x_k - x_*\|, \|x_{k+1} - x_*\|\}. \tag{3.4}$$

Noting that $y_k = \int_0^1 \nabla^2 f(x_k + ts_k)\mathrm{d}t$, we can obtain from this and (3.3) that

$$\begin{aligned} \|y_k - \nabla^2 f(x_*)s_k\| &= \left\| \int_0^1 [\nabla^2 f(x_k + ts_k) - \nabla^2 f(x_*)]s_k \mathrm{d}t \right\| \\ &\leqslant \int_0^1 \|\nabla^2 f(x_k + ts_k) - \nabla^2 f(x_*)\|\|s_k\|\mathrm{d}t \\ &\leqslant \int_0^1 L\|(1-t)(x_k - x_*) + t(x_{k+1} - x_*)\|\|s_k\|\mathrm{d}t \\ &\leqslant \int_0^1 L\varepsilon_k\|s_k\|\mathrm{d}t = L\varepsilon_k\|s_k\|. \end{aligned} \tag{3.5}$$

Denote $G_* = \nabla^2 f(x_*)$, $H_* = (G_*)^{-1}$ and take $P = H_*$ which satisfies (2.1). Define $\tilde{s}_k, \tilde{y}_k, \tau_k$, etc., as before. Then, we know by (3.5) that

$$\|\tilde{y}_k - \tilde{s}_k\| \leqslant \|H_*^{1/2}\|\|y_k - s_k\| \leqslant L\|H_*^{1/2}\|\varepsilon_k\|s_k\| \leqslant L\|H_*^{1/2}\|^2\varepsilon_k\|\tilde{s}_k\|. \tag{3.6}$$

Further, we can prove without difficulty that, there must exist positive constants $c_1 > 0$ and $c_2 \in (0, b)$ such that all the quantities

$$M_k - 1, \ \mu_k \triangleq \frac{2 - M_k - m_k}{m_k}, \ \bar{\mu}_k \triangleq \frac{(\tilde{y}_k - \tilde{s}_k)^{\mathrm{T}}\tilde{H}_k\tilde{y}_k}{\mathrm{tr}(\tilde{H}_k)\tilde{s}_k^{\mathrm{T}}\tilde{y}_k}, \ \ln m_k \leqslant \frac{1}{2}c_1\varepsilon_k \tag{3.7}$$

whenever $\varepsilon_k < c_2$. Now let us denote

$$\rho_k = q_k - 1 - \ln q_k, \quad \zeta_k = (1 - \phi_k)q_k(\tau_k^{-2} - 1), \quad \xi_k = \ln[1 + \phi_k(\beta_k\gamma_k - 1)]. \tag{3.8}$$

The inequality (2.10) can be written as

$$\psi(\tilde{H}_{k+1}) \leqslant \psi(\tilde{H}_k) - \rho_k - \zeta_k - \xi_k + (M_k - 1) + \phi_k q_k \mu_k + \phi_k \mathrm{tr}(\tilde{H}_k)\bar{\mu}_k + \ln m_k. \tag{3.9}$$

Since $\phi_k \in [0, 1], \gamma_k = q_k/m_k$ and $0 \leqslant q_k \leqslant \mathrm{tr}(\tilde{H}_k)$, we know from the above relation and (3.7) that

$$\psi(\tilde{H}_{k+1}) \leqslant \psi(\tilde{H}_k) - \rho_k - \zeta_k - \xi_k + c_1(1 + \mathrm{tr}(\tilde{H}_k))\varepsilon_k. \tag{3.10}$$

Using the fact that

$$\lambda - \ln\lambda \geqslant \max\left[\left(1 - \frac{1}{\mathrm{e}}\right)\lambda, 1\right], \quad \text{for any } \lambda > 0, \tag{3.11}$$

we can show that (see [20])

$$\psi(A) \geqslant \max\left[\left(1-\frac{1}{\mathrm{e}}\right)\mathrm{tr}(A), n\right], \tag{3.12}$$

for any symmetric positive definite matrix A. Denoting $c_3 = c_1[\frac{1}{n} + \frac{\mathrm{e}}{\mathrm{e}-1}]$, it follows from (3.10) that

$$\psi(\tilde{H}_{k+1}) \leqslant (1 + c_3\varepsilon_k)\psi(\tilde{H}_k) - \rho_k - \zeta_k - \xi_k. \tag{3.13}$$

Since $\tau_k^2 \leqslant 1$ and $\beta_k\gamma_k \geqslant 1$, we see that the quantities ρ_k, ζ_k, and ξ_k are all non-negative. Then, we obtain

$$\psi(\tilde{H}_{k+1}) \leqslant (1 + c_3\varepsilon_k)\psi(\tilde{H}_k). \tag{3.14}$$

In the following, we establish the local linear convergence of the MCQN update with Broyden's convex family with the help of (3.14).

Theorem 3.2 *Suppose that Assumption 3.1 holds. Consider the method* (1.2) *where* $\alpha_k \equiv 1$, *and* H_k *is obtained by the MCQN update with* $H^{QN} = H^B_{k+1}$ $(\phi_k \in [0,1])$. *Then, for any* $\alpha \in (0,1)$, *there exists* τ *such that* $\|x_0 - x_*\| \leqslant \tau$ *and* $\|H_0 - H_*\| \leqslant \tau$ *imply*

$$\|x_{k+1} - x_*\| \leqslant \alpha\|x_k - x_*\|, \quad \textit{for all } k. \tag{3.15}$$

Proof Suppose that $\alpha \in (0,1)$. From Lemma 4 in [19], there must exist constants $\bar{\tau} \in (0,b)$ and $\delta > 0$ such that, if $\|x_0 - x_*\| \leqslant \bar{\tau}$, then

$$\psi(\tilde{H}_0) - n \leqslant \frac{\delta}{2}, \tag{3.16}$$

$$\psi(\tilde{H}) - n \leqslant \delta \Longrightarrow \|H - H_*\| \leqslant \frac{\alpha}{2\hat{L}}, \tag{3.17}$$

where H is any symmetric positive definite matrix and $\tilde{H} = H_*^{-\frac{1}{2}} H H_*^{-\frac{1}{2}}$. Choose

$$\tau = \min\left\{\bar{\tau}, c_2, \frac{\alpha}{\hat{L}}, \frac{\alpha}{LM}, \frac{1-\alpha}{c_3}\ln\left(\frac{2(n+\delta)}{2n+\delta}\right)\right\}. \tag{3.18}$$

We will show by induction that the following inequalities hold for all k:

$$\|x_{k+1} - x_*\| \leqslant \alpha\|x_k - x_*\|, \tag{3.19}$$

$$\|H_k - H_*\| \leqslant \frac{\alpha}{2\hat{L}}. \tag{3.20}$$

To begin with, similar to (1.7), we can show by (3.3) and (3.1) that for any $x \in \mathcal{B}$,

$$
\begin{aligned}
\|x - x_* - H_* \nabla f(x)\| &\leqslant \|H_*\| \|(\nabla f(x) - \nabla f(x_*)) - G_*(x - x_*)\| \\
&\leqslant \|H_*\| \int_0^1 \|\nabla^2 f(x + t(x - x_*)) - G_*\| \|(x - x_*)\| \mathrm{d}t \\
&\leqslant \frac{1}{2} LM \|x - x_*\|^2.
\end{aligned}
\tag{3.21}
$$

Now, when $k = 0$, we see that (3.19) holds due to (3.16) and (3.17). Moreover, by (1.2), $\alpha_k \equiv 1$, (3.21) with $x = x_0$, and the choice of τ, we can get that

$$
\begin{aligned}
\|x_1 - x_*\| &= \|x_0 - H_0 \nabla f(x_0) - x_*\| \\
&\leqslant \|x_0 - x_* - H_* \nabla f(x_0)\| + \|(H_0 - H_*)(\nabla f(x_0) - \nabla f(x_*))\| \\
&\leqslant \frac{1}{2} LM \|x - x_*\|^2 + \|H_0 - H_*\| \|\nabla f(x_0) - \nabla f(x_*)\| \\
&\leqslant \left(\frac{1}{2} LM \|x_0 - x_*\| + \frac{\alpha}{2}\right) \|x_0 - x_*\| \\
&\leqslant \left(\frac{1}{2} LM \tau + \frac{\alpha}{2}\right) \|x_0 - x_*\| \ \leqslant \ \alpha \|x_0 - x_*\|.
\end{aligned}
\tag{3.22}
$$

So (3.19) and (3.20) hold for $k = 0$. Suppose that (3.19) and (3.20) are true for $k = 0, 1, \cdots, l$. Then, we have that $\varepsilon_k = \|x_k - x_*\|$ and $\varepsilon_k \leqslant \alpha^k \varepsilon_0 \leqslant \alpha^k \tau$ for $k = 0, 1, \cdots, l$. Similar to (3.22), we get

$$
\begin{aligned}
\|x_{l+1} - x_*\| &= \|x_l - H_l \nabla f(x_l) - x_*\| \\
&\leqslant \|x_l - x_* - H_* \nabla f(x_l)\| + \|(H_l - H_*)(\nabla f(x_l) - \nabla f(x_*))\| \\
&\leqslant \left(\frac{1}{2} LM \|x_l - x_*\| + \frac{\alpha}{2}\right) \|x_l - x_*\| \\
&\leqslant \left(\frac{1}{2} LM \alpha^l \tau + \frac{\alpha}{2}\right) \|x_l - x_*\| \ \leqslant \ \alpha \|x_l - x_*\|.
\end{aligned}
\tag{3.23}
$$

On the other hand, by the choice of τ, we have that

$$
c_3 \sum_{k=0}^{l} \varepsilon_k \leqslant c_3 \tau \sum_{k=0}^{l} \alpha^k = c_3 \tau \frac{1 - \alpha^{l+1}}{1 - \alpha} \leqslant \frac{c_3 \tau}{1 - \alpha} \leqslant \ln \frac{2(n + \delta)}{2n + \delta}. \tag{3.24}
$$

It follows from this, (3.16) and (3.14) that

$$
\begin{aligned}
\psi(\tilde{H}_{l+1}) - n &\leqslant (\psi(\tilde{H}_0) - n) + \left(\prod_{k=0}^{l} (1 + c_3 \varepsilon_k) - 1\right) \psi(\tilde{H}_0) \\
&\leqslant \frac{\delta}{2} + \left(n + \frac{\delta}{2}\right) \left(\prod_{k=0}^{l} \mathrm{e}^{c_3 \varepsilon_k} - 1\right) \\
&\leqslant \frac{\delta}{2} + \left(n + \frac{\delta}{2}\right) \left(\mathrm{e}^{c_3 \sum_{k=0}^{l} \varepsilon_k} - 1\right) \\
&\leqslant \frac{\delta}{2} + \left(n + \frac{\delta}{2}\right) \left(\frac{2(n + \delta)}{2n + \delta} - 1\right) = \delta.
\end{aligned}
\tag{3.25}
$$

Thus, by (3.17), relation (3.20) holds for $k = l + 1$. Therefore, by induction, (3.19) and (3.20) are true for all $k \geqslant 0$. This completes the proof. □

Now, we are ready to establish the superlinear convergence for the MCQN update with Broyden's convex family. The proof is based on the local linear convergence of the update and the relation (3.14).

Theorem 3.3 *Suppose that Assumption 3.1 holds. Consider the method* (1.2) *where* $\alpha_k \equiv 1$, *and* H_k *is obtained by the MCQN update with* $H^{QN} = H^B_{k+1}$ $(\phi_k \in [0,1])$. *Then, there exists a positive constant* τ *such that, if* $\|x_0 - x_*\| \leqslant \tau$ *and* $\|H_0 - H_*\| \leqslant \tau$, *then*

$$\lim_{k\to\infty} \frac{\|(H_k - H_*)y_k\|}{\|y_k\|} = 0. \tag{3.26}$$

Thus, the sequence $\{x_k\}$ *generated by the method converges to* x_* *superlinearly.*

Proof For any $\alpha \in (0,1)$, we choose τ as in the proof of Theorem 3.2. Then, (3.19) and (3.20) hold for all k. At the same time, we have that

$$\psi(\tilde{H}_k) - n \leqslant \delta \quad \text{for all}\, k. \tag{3.27}$$

Note that (3.13) implies that

$$\rho_k + \zeta_k + \xi_k \leqslant (\psi(\tilde{H}_{k+1}) - \psi(\tilde{H}_k)) + c_3\varepsilon_k\psi(\tilde{H}_k). \tag{3.28}$$

Summing the above relation and using (3.24) and (3.27), we obtain

$$\sum_{k\geqslant 1}(\rho_k + \zeta_k + \xi_k) \leqslant c_3 \sum_{k\geqslant 1} \varepsilon_k \psi(\tilde{H}_k) \leqslant c_3(n+\delta)\ln\frac{2(n+\delta)}{2n+\delta} < \infty. \tag{3.29}$$

Since the quantities ρ_k, ζ_k, and ξ_k are all nonnegative, the relation (3.29) indicates that they are all tend to zero as $k \to \infty$. Further, by their definitions in (3.8), we know that (*i*) $q_k \to 1$; (ii) if $\phi_k \leqslant \frac{1}{2}$, $\tau_k \to 1$; and (iii) if $\phi_k > \frac{1}{2}$, $\beta_k\gamma_k \to 1$.

Now, we consider the quantity $\|(H_k - H_*)y_k\|/\|y_k\|$. On one side, we have that

$$\begin{aligned} \frac{\|H_*^{-1/2}(H_k - H_*)y_k\|^2}{\|H_*^{\frac{1}{2}}y_k\|^2} &= \frac{\|\tilde{H}_k\tilde{y}_k - \tilde{y}_k\|^2}{\|\tilde{y}_k\|^2} = \frac{\|\tilde{H}_k\tilde{y}_k\|^2 - 2\tilde{y}_k^T\tilde{H}_k\tilde{y}_k + \|\tilde{y}_k\|^2}{\|\tilde{y}_k\|^2} \\ &= \frac{q_k}{\tau_k^2} - 2q_k + 1. \end{aligned} \tag{3.30}$$

On the other side, we have that

$$\begin{aligned} \frac{\|\tilde{H}_k\tilde{y}_k - \tilde{s}_k\|^2}{\|\tilde{y}_k\|^2} &\leqslant \frac{\|\tilde{H}_k{}^{1/2}\|^2\|\tilde{H}_k{}^{1/2}\tilde{y}_k - (\tilde{H}_k)^{-1/2}\tilde{s}_k\|^2}{\|\tilde{y}_k\|^2} \\ &= \frac{\|\tilde{H}_k{}^{1/2}\|^2(\tilde{y}_k^T\tilde{H}_k\tilde{y}_k - 2\tilde{s}_k^T\tilde{y}_k + \tilde{s}_k^T(\tilde{H}_k)^{-1}\tilde{s}_k)}{\|\tilde{y}_k\|^2} \\ &= \|\tilde{H}_k{}^{1/2}\|^2\left(q_k - 2m_k + \frac{\beta_k\gamma_k}{q_k}\right). \end{aligned} \tag{3.31}$$

In addition, note that

$$\left| \frac{\|\tilde{H}_k \tilde{y}_k - \tilde{y}_k\|}{\|\tilde{y}_k\|} - \frac{\|\tilde{H}_k \tilde{y}_k - \tilde{s}_k\|}{\|\tilde{y}_k\|} \right| \leqslant \frac{\|\tilde{y}_k - \tilde{s}_k\|}{\|\tilde{y}_k\|} \to 0. \tag{3.32}$$

For the subsequence $\{k_i : \phi_{k_i} \leqslant \frac{1}{2}\}$, we know from $q_k \to 1, \tau_{k_i} \to 1$ and (3.30) that the limit relation (3.26) holds for the subsequence $\{k_i\}$. For the subsequence $\{k_i : \phi_{k_i} > \frac{1}{2}\}$, we know from $q_k \to 1, \beta_{k_i}\gamma_{k_i} \to 1, m_k \to 1$, the first equality in (3.30), (3.31), and (3.32) that the limit relation (3.26) holds for the subsequence $\{k_i\}$. Thus, combining the two cases, (3.26) is true. As addressed in [19], (3.26) implies the following relation

$$\lim_{k \to \infty} \frac{\|((H_k)^{-1} - G_*)s_k\|}{\|s_k\|} = 0. \tag{3.33}$$

Therefore, by [6], we know that $\{x_k\}$ is superlinearly convergent. □

4 Numerical Studies

In the previous section, we analyzed the convergence properties of the MCQN update with Broyden's convex family. In this section, we will investigate the numerical performance of the MCQN update with Broyden's positive family, in which case the ϕ_k in (1.4) is restricted to be nonnegative.

Five test problems in CUTEr [13] and [14] were used in our numerical studies, where x_{ini} is the standard initial point for each problem. For each problem, we tried four initial points x_{ini}, 4 x_{ini}, 7 x_{ini}, and 10 x_{ini}.

Problem 1 (TRIDIA)

$$f(x) = (x_1 - 1)^2 + \sum_{i=2}^{n} i(x_{i-1} - 2x_i)^2,$$
$$x_{\text{ini}} = (1, \cdots, 1)^{\mathrm{T}}.$$

Problem 2 (Extended Rosenbrock function)

$$f(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2,$$
$$x_{\text{ini}} = (-1.2, 1, -1.2, 1, \cdots, -1.2, 1)^{\mathrm{T}}.$$

Problem 3 (Extended Powell singular function)

$$f(x) = \sum_{i=1}^{n/4} \Big[10(x_{4i-3} - x_{4i})^4 + (x_{4i-2} - 2x_{4i-1})^4 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-3} + 10x_{4i-2})^2 \Big],$$
$$x_{\text{ini}} = (3, -1, 0.1, \cdots, 3, -1, 0, 1)^{\mathrm{T}}.$$

Problem 4 (Broyden tridiagonal function)

$$f(x) = (3x_1 - 2x_1^2 - 2x_2 + 1)^2 + (3x_n - 2x_n^2 - x_{n-1} + 1)^2 + \sum_{i=2}^{n-1} (3x_i - 2x_i^2 - x_{i-1} - 2x_{i+1} + 1)^2,$$
$$x_{\text{ini}} = (-1, \cdots, -1)^{\text{T}}.$$

Problem 5 (Broyden banded function)

$$f(x) = \sum_{i=1}^{n} \left(5x_i^3 + 2x_i + 1 - \sum_{j \in J_i} x_j(1 + x_j) \right)^2,$$

where

$$J_i = \{j \mid j \neq i, \max\{1, i-5\} \leqslant j \leqslant \min\{n, i+1\}\}.$$
$$x_{\text{ini}} = (-1, \cdots, -1)^{\text{T}}.$$

All the Hessians of the above problems are band matrices. Therefore, we can obtain the chordal extensions of their sparsity pattern without difficulty. The dimensions n of all problems are set to 1 000. In our tests, we are satisfied with an approximate stationary point with

$$\|\nabla f(x)\|_\infty \leqslant 10^{-5}.$$

Table 1 lists the iteration numbers required for each fixed value of ϕ_k. In each row of the table, we write the least iteration number(s) in the bold style. From Table 1, we can see that if ϕ_k is restricted to the interval [0, 1], the choice of $\phi_k = 1$ provides the best results. This indicates that, in the MCQN update, the BFGS formula is the best choice in Broyden's convex family. This conclusion is the same as in the ordinary quasi-Newton method (see [2, 16]). On the other hand, we can also see from the table that the numerical performance of the MCQN update can be improved by increasing the value of ϕ_k. It can be seen in (1.4) that the positive definiteness of the quasi-Newton matrix H_k is kept, if $\phi_k \geqslant 0$. Consequently, the MCQN update is well defined for all $\phi_k \geqslant 0$. Surely, if $\phi_k > 1$, the superlinear convergence analysis in Sect. 3 does not apply because the quantity ζ_k is not nonnegative any more, as is necessary for the deduction of the relation (3.14). Despite the lack of a strict superlinear convergence proof, we see from Table 1 that $\phi_k = 4$, 5 provide quite good numerical results, which are significantly superior to those of $\phi_k = 1$.

Although the MCQN update with BFGS can still be improved by increasing the value of ϕ_k, the best choice of ϕ_k varies from test problems. For example, for Problem 2 with the standard initial point x_{ini}, $\phi_k = 4$ is an ideal choice. If $10x_{\text{ini}}$ is used as the initial point, $\phi_k = 5$ is a better choice. Based on these observations, we feel that it might be worthwhile to dynamical choices of ϕ_k. We considered the following three ways:

Table 1 Testing MCQN with Broyden's family with fixed ϕ_k's

P	Intial point	ϕ_k 0	0.2	0.4	0.6	0.8	1	2	3	4	5	7	10
1	x_{ini}	1 727	359	272	243	217	219	188	170	156	142	**131**	146
	$4x_{\text{ini}}$	1 750	373	281	260	241	225	196	177	161	141	**136**	156
	$7x_{\text{ini}}$	1 774	376	289	252	232	225	191	181	166	156	**141**	158
	$10x_{\text{ini}}$	1 783	377	294	264	236	227	189	180	167	152	**139**	167
2	x_{ini}	6 589	4 909	4 005	3 608	3 342	3 279	2 909	2 744	**2 626**	2 652	2 709	3 916
	$4x_{\text{ini}}$	6 611	4 915	4 070	3 576	3 360	3 200	2 857	2 747	2 688	**2 653**	2 683	4 015
	$7x_{\text{ini}}$	1 096	165	142	122	116	109	93	88	83	83	86	**79**
	$10x_{\text{ini}}$	6 654	4 969	4 056	3 614	3 400	2 635	2 886	2 864	2 658	**2 584**	2 724	3 623
3	x_{ini}	1 919	1 178	702	1 297	1 025	971	653	608	560	**483**	502	522
	$4x_{\text{ini}}$	1 921	1 178	701	1 343	1 168	985	731	638	583	492	490	**468**
	$7x_{\text{ini}}$	1 923	1 206	927	1 290	1 071	987	699	617	564	**498**	495	562
	$10x_{\text{ini}}$	1 906	1 208	936	1 360	1 041	1 009	692	657	614	610	531	**471**
4	x_{ini}	283	128	78	65	62	59	**47**	49	**47**	58	53	85
	$4x_{\text{ini}}$	134	74	72	72	69	53	**44**	48	52	83	49	102
	$7x_{\text{ini}}$	563	213	404	146	132	182	95	**62**	86	131	77	68
	$10x_{\text{ini}}$	637	418	320	207	145	134	79	**78**	94	86	92	103
5	x_{ini}	3 169	656	364	578	147	123	143	138	**77**	87	122	169
	$4x_{\text{ini}}$	107	60	52	49	46	46	46	46	47	**38**	**38**	39
	$7x_{\text{ini}}$	493	136	99	85	109	90	76	76	62	**49**	55	104
	$10x_{\text{ini}}$	713	191	158	121	103	111	110	103	111	146	160	**64**

(I) The first way is to pick an interval [L, U] with $L \geqslant 0$ and solve the one-dimensional problem

$$\begin{array}{ll} \min & \|H^{\mathrm{QN}} y_{k-1} - s_{k-1}\| \\ \text{s.t.} & L \leqslant \phi \leqslant U. \end{array} \tag{4.1}$$

In other words, we want H^{QN} to satisfy the secant condition $H^{\mathrm{QN}} y_{k-1} = s_{k-1}$ as possible as it can while it satisfies the one that $H^{\mathrm{QN}} y_k = s_k$ exactly. Since such ϕ_k sometimes becomes negative, we force ϕ_k to remain on [L, U]. Numerical results with three choices, [1, 5], [1, 10], and [1,50], of [L, U] are taken down in columns 3–5 of Table 2. Again, in each row of Table 2, we write the least iteration number(s) in the bold style.

(II) The second way comes from another observation of superlinear convergence proof and aims to restrict ϕ_k so that a relation similar to (3.14) can be achieved. Pick some constants $c_1, c_2 \in (0, 1)$. If the choice of ϕ_k is such that

$$-\zeta_k = (\phi_k - 1) q_k (\tau_k^{-2} - 1) \leqslant c_1 \rho_k + c_2 \xi_k, \tag{4.2}$$

we can still obtain

$$\psi(\tilde{H}_{k+1}) \leqslant (1 + c\varepsilon_k)\psi(\tilde{H}_k) - (1 - c_1)\rho_k - (1 - c_2)\xi_k. \tag{4.3}$$

If the above choice of ϕ_k is possible, we can prove that both ρ_k and ξ_k tend to zero and achieve the superlinear convergence. However, the relation (4.2) is related to $\tilde{y}_k$, $\tilde{s}_k$, and $\tilde{H}_k$, which depend on $\nabla^2 f(x_*)$. Since we do not know $\nabla^2 f(x_*)$ in general, we cannot directly exploit the inequality (4.2). Nevertheless, we consider to use y_k, s_k, and H_k to replace $\tilde{y}_k, \tilde{s}_k$, and $\tilde{H}_k$, respectively, in (4.2). In this case, since $H_k^{-1} s_k = -\alpha_k g_k$, the value of β_k and hence the value of ξ_k can be easily obtained. For different values of c_1 and c_2, we then choose ϕ_k such that

$$(\phi_k - 1) q_k (\tau_k^{-2} - 1) = c_1 \rho_k + c_2 \xi_k. \tag{4.4}$$

The quantity ϕ_k is involved in the calculations of ξ_k. We solve the equality (4.4) by Newton's method starting from

$$\phi_0 = c_1 \rho_k / [q_k (\tau_k^{-2} - 1)].$$

For numerical stability, we used a projection of a solution ϕ_k onto some interval [L, U]. We tested different choices of (c_1, c_2) and [L, U] and found that the numerical results are not sensitive to the choices of the parameters. The recommended intervals for c_1 and c_2 are both [0.7, 0.95]. Numerical results with fixed choice $(c_1, c_2) = (0.9, 0.9)$ and three choices, [1, 5], [1, 10], and [1,50], of [L, U] are taken down in columns 6–8 of Table 2.

(III) The third way comes from the proposal in [21] and aims to choose ϕ_k such that the search direction is close to the steepest descent direction. Denote

$$p_k = -\frac{g_{k+1}^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} s_k} s_k, \quad q_k = -\left(H_k g_k - \frac{y_k^{\mathrm{T}} H_k g_k}{y_k^{\mathrm{T}} H_k y_k} H_k y_k \right)$$

and

Table 2 Testing MCQN with Broyden's family with dynamical ϕ_k's

P	Intial point	Way (I)			Way (II)			Way (III)			
		[1, 5]	[1, 10]	[1, 50]	[1, 5]	[1, 10]	[1, 50]	[μ_k, 3μ_k]	[μ_k, 5μ_k]	[μ_k, 7μ_k]	[μ_k, 10μ_k]
1	x_{ini}	195	194	182	179	178	**162**	204	197	205	191
	$4x_{ini}$	205	200	182	193	**174**	176	206	203	198	206
	$7x_{ini}$	204	206	194	182	187	**178**	217	201	206	215
	$10x_{ini}$	211	208	188	192	**175**	184	203	203	209	215
2	x_{ini}	2 974	2 872	2 874	2 800	**2 745**	2 804	3 128	3 154	3 019	3 109
	$4x_{ini}$	2 943	2 921	2 879	2 814	2 762	**2 696**	3 107	3 033	3 025	3 103
	$7x_{ini}$	104	103	102	95	95	**81**	103	113	116	107
	$10x_{ini}$	**2 471**	2 509	2 501	2 782	2 789	2 771	3 104	3 116	3 028	3 387
3	x_{ini}	920	897	856	695	676	742	866	833	675	**639**
	$4x_{ini}$	829	852	685	699	685	792	899	837	729	**668**
	$7x_{ini}$	930	915	710	**709**	743	791	906	856	863	714
	$10x_{ini}$	907	853	711	735	730	845	920	845	792	**703**
4	x_{ini}	55	57	59	55	66	75	60	60	**54**	55
	$4x_{ini}$	50	67	**48**	82	57	59	54	53	51	53
	$7x_{ini}$	151	157	154	137	**96**	117	107	107	117	127
	$10x_{ini}$	111	101	111	92	91	79	77	**75**	99	84
5	x_{ini}	112	111	86	103	97	**82**	249	114	150	97
	$4x_{ini}$	46	46	46	43	46	42	45	43	54	**31**
	$7x_{ini}$	84	83	83	80	**57**	89	94	93	93	85
	$10x_{ini}$	101	98	97	108	**62**	85	108	99	154	113

$$d(\rho) = p_k + \rho q_k.$$

Let ρ_k^* be a solution of

$$\begin{array}{ll} \min & \frac{d(\rho)^{\mathrm{T}} g_{k+1}}{\|d(\rho)\| \|g_{k+1}\|} \\ \text{s.t.} & \rho_k^L \leqslant \rho \leqslant \rho_k^U, \end{array}$$

where ρ_k^L and ρ_k^U are the upper and lower bounds of ρ. Note that ρ_k^* is easily obtained by [(4.5)–(4.6), 21]. Then ϕ_k is given by

$$\phi_k = \frac{\rho_k^* - 1}{\mu_k - 1}$$

with

$$\mu_k = \frac{s_k^{\mathrm{T}} H_k^{-1} s_k y_k^{\mathrm{T}} H_k y_k}{(y_k^{\mathrm{T}} s_k)^2}.$$

Here, the definition of μ_k is different from the one in (3.7). The suggested interval $[\rho_k^L, \rho_k^U]$ in [21] is $[\mu_k, 5\ \mu_k]$. We tested four different intervals and took down the corresponding numerical results into columns 9–12 of Table 2.

Comparing with the MCQN update with BFGS (see the column $\phi_k = 1$ in Table 1), we see that all the three dynamical ways of choosing ϕ_k can lead to almost uniformly better numerical results. Among the three dynamical ways, way (II) seems to be the most efficient one.

5 Conclusions and discussions

In this paper, we have established the local and superlinear convergence of the MCQN update with Broyden's convex family under suitable assumptions. A global convergence analysis has been given for the MCQN update with Broyden's convex family assuming that the objective function is uniformly convex and its dimension is only two. Numerical results have been reported for the MCQN update with Broyden's positive family. They show that, to obtain the intermediate matrix H_k^{QN}, the BFGS formula is the best choice among Broyden's convex family. On the other hand, there are several ways to beat the BFGS formula if one considers Broyden's positive family with ϕ_k larger than or equal to 1.

It still remains to study whether the MCQN update with Broyden's convex family is globally convergent or not for uniformly convex functions of any dimension. As discussed with Professor Ya-xiang Yuan, this problem is much related to the global convergence problem of the ordinary DFP method with inexact line search for uniformly convex functions and hence may be very difficult. For general objective functions, it is easy to know that the MCQN update with BFGS or Broyden's convex family need not converge because that the MCQN update with BFGS corresponds with the ordinary BFGS method in case of full Hessian and that by [4] even the ordinary BFGS method is not necessarily globally convergent. Nevertheless, Ref. [5] established the global convergence of MCQN update with Broyden's convex family in the case when the objective function $f(x)$ is uniformly convex and there are only two variables.

By dynamically choosing ϕ_k in Broyden's positive family, we can obtain MCQN updates better than the MCQN update with BFGS. From Tables 1 and 2, however, we see that some fixed choices of ϕ_k give quite good numerical results and they are even better than those of dynamical ways. Therefore, we wonder whether there exist more efficient dynamical ways of choosing ϕ_k or not. In addition, it is also worthwhile how to order the two steps in the MCQN upate. See [3] for a useful try along this way.

Acknowledgments The authors are grateful to Professors Masao Fukushima and Ya-xiang Yuan for their warm encouragement and valuable suggestions. They also thank the two anonymous referees very much for their useful comments on an early version of this paper.

References

[1] Byrd, R., Nocedal, J.: A tool for the analysis of quasi-Newton methods with application to unconstrained optimization. SIAM J. Numer. Anal. **26**, 727–739 (1989)
[2] Byrd, R., Nocedal, J., Yuan, Y.: Global convergence of a class of quasi-Newton methods on convex problems. SIAM J. Numer. Anal. **24**, 1171–1189 (1987)
[3] Cheng, M., Dai, Y.H., Diao, R.: A new sparse quasi-Newton update method. Sultan Qaboos Univ. J. Sci. **17**(1), 30–43 (2012)
[4] Dai, Y.H.: Convergence properties of the BFGS algorithm. SIAM J. Optim. **13**(3), 693–701 (2003)
[5] Dai, Y.H., Yamashita, N.: Convergence analysis of sparse quasi-Newton updates with positive definite matrix completion for two-dimensional functions. Numer. Algebra Control Optim. **1**(1), 61–69 (2011)
[6] Dennis, J.E. Jr., Moré, J.J.: Quasi-Newton method, motivation and theory. SIAM Rev. **19**, 46–89 (1977)
[7] Fletcher, R.: An optimal positive definite update for sparse Hessian matrices. SIAM J. Optim. **5**, 192–218 (1995)
[8] Fukuda, M., Kojima, M., Murota, K., Nakata, K.: Exploiting sparsity in semidefinite programming via matrix completion I: general frameworks. SIAM J. Optim. **11**, 647–674 (2000)
[9] Griewank, A., Toint, Ph.L.: Local convergence analysis of partitioned quasi-Newton updates. Numer. Math. **39**, 429–448 (1982)
[10] Griewank, A., Toint, Ph.L.: Partitioned variable metric updates for large structure optimization problems. Numer. Math. **39**(1), 119–137 (1982)
[11] Griewank, A., Toint, Ph. L.: *Numerical experiments with partially separable optimization problems*. In: Griffiths D.F. (ed.) Numerical Analysis: Proceedings Dundee 1983. Lecture Notes in Mathematics, vol. 1066, pp. 203–220. Springer Verlag, Berlin (1984)
[12] Gould, N.I.M., Orban, D., Toint, Ph.L.: CUTEr, a constrained and unconstrained testing environment, revisited. ACM Trans. Math. Softw. **29**, 373–94 (2003)
[13] Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Math. Program. **45**, 503–528 (1989)
[14] Moré, J.J., Garbow, B.S., Hillstrom, K.E.: Testing unconstrained optimization software. ACM Trans. Math. Softw. **7**, 17–41 (1981)
[15] Nocedal, J.: Updating quasi-Newton matrices with limited storage. Math. Comput. **35**, 773–782 (1980)
[16] Powell, M.J.D.: How bad are the BFGS and DFP methods when the objective function is quadratic?. Math. Program. **34**, 34–47 (1986)
[17] Sorensen, D.C.: Collinear scaling and sequential estimation in sparse optimization algorithm. Math. Program. Study **18**, 135–159 (1982)

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

[19] Yamashita, N.: Sparse quasi-Newton updates with positive definite matrix completion. Math. Program. **115**(1), 1–30 (2008)

[20] Yuan, Y., Peng J., M.: Properties of Ψ-functoins and its applications. Math. Numer. Sin. **1**, 102–107 (1994)

[21] Zhang, Y., Tewarson, R.P.: Quasi-Newton algorithm with updates from the preconvex part of Broyden's family. IMA J. Numer. Anal. **8**, 487–500 (1988)