NEW QUASI-NEWTON METHOD FOR SOLVING SYSTEMS OF NONLINEAR EQUATIONS

Ladislav Lukšan, Jan Vlček, Praha

Received September 13, 2016. First published March 3, 2017.

Abstract. We propose a new Broyden method for solving systems of nonlinear equations, which uses the first derivatives, but is more efficient than the Newton method (measured by the computational time) for larger dense systems. The new method updates QR or LU decompositions of nonsymmetric approximations of the Jacobian matrix, so it requires $O(n^2)$ arithmetic operations per iteration in contrast with the Newton method, which requires $O(n^3)$ operations per iteration. Computational experiments confirm the high efficiency of the new method.

Keywords: nonlinear equation; system of equations; trust-region method; quasi-Newton method; adjoint Broyden method; numerical algorithm; numerical experiment

MSC 2010: 65K10

1. Introduction

Consider the system of nonlinear equations

$$f(x) = 0, \tag{1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear mapping, and denote by $J(x)$ the Jacobian matrix of f at the point x. We suppose that the Jacobian matrix is dense of a dimension which is not small, so methods saving matrix operations are preferred. We will use the following assumptions concerning the mapping f.

Assumption J1. The mapping $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable on the level set $\mathcal{D}(\overline{F}) = \{x \in \mathbb{R}^n\colon \|f(x)\| \leqslant \overline{F}\}$, where $\overline{F}$ is a suitable upper bound,

This work was supported by the Grant Agency of the Czech Republic, project No. 13-06684S, and the Institute of Computer Science of the AS CR (RVO: 67985807).

DOI: 10.21136/AM.2017.0253-16

and the Jacobian matrix J is Lipschitz continuous on $\mathcal{D}(\overline{F})$, i.e., there is a constant $\overline{L} > 0$ such that

$$\|J(y) - J(x)\| \leqslant \overline{L}\|y - x\| \quad \forall\, x, y \in \mathcal{D}(\overline{F}). \tag{2}$$

Assumption J2. There is a constant $\overline{J} > 0$ such that

$$\|J(x_i)s\| \leqslant \overline{J}\|s\| \quad \forall\, i \in \mathbb{N} \quad \forall\, s \in \mathbb{R}^n. \tag{3}$$

Notice that Assumption J2 follows from Assumption J1 if $\mathcal{D}(\overline{F})$ is compact.

Assumption J3. There is a constant $\underline{J} > 0$ such that

$$\|J(x_i)s\| \geqslant \underline{J}\|s\| \quad \forall\, i \in \mathbb{N}\; \forall\, s \in \mathbb{R}^n, \tag{4}$$

where $x_i \in \mathbb{R}^n$, $i \in \mathbb{N}$, are points generated by a chosen solution method.

We restrict our attention to iterative methods of the form $x_{i+1} = x_i + \alpha_i s_i$, $i \in \mathbb{N}$, with $A_i s_i + f_i \approx 0$, $f_i = f(x_i)$, $A_i \approx J_i = J(x_i)$ and $\alpha_i \geqslant 0$, which generate a monotone non-increasing sequence of norms $\|f(x_i)\|$, $i \in \mathbb{N}$. Since the norm $\|f(x)\|$ is a non-smooth function, we use the scaled squared norm $F(x) = \|f(x)\|^2/2$ as a merit function and assume that its gradient $\nabla F(x) = J(x)^{\mathrm{T}} f(x)$ is computed either analytically or by reverse automatic differentiation. The Newton method, which is the most widely known and rapidly convergent method of this type, uses matrices $A_i = J_i$, $i \in \mathbb{N}$. Since the Jacobian matrix J_i is completely recomputed in every iteration, the solution of the linear system $J_i s_i + f_i = 0$ requires $O(n^3)$ arithmetic operations per iteration to obtain a matrix factorization. This fact prolongs the computational time, so quasi-Newton methods, which update factorizations of matrices A_i, $i \in \mathbb{N}$, in $O(n^2)$ arithmetic operations, can be more efficient for larger n.

In this paper, we propose a new quasi-Newton method (31), which is a good approximation of the two-sided adjoint quasi-Newton method (26). Two-sided adjoint quasi-Newton methods have sophisticated theoretical (Theorem 8) and excellent numerical properties. Surprisingly, the new method is numerically perfect as well, but, unlike the method (26), it does not require additional computation of directional derivatives $J_{i+1} d_i$, $i \in \mathbb{N}$ (the computation of gradients $J_{i+1}^{\mathrm{T}} f_{i+1}$, $i \in \mathbb{N}$, suffices, see Section 3).

The paper is organized as follows. In Section 2, we briefly describe the trust region approach used in the implementation of quasi-Newton methods. Section 3, which is devoted to quasi-Newton methods and their properties, introduces a new quasi-Newton method. Section 4 contains results of computational experiments, which confirm the high efficiency of the new method. We follow results introduced in [3], [4] and [14]–[15]. Further information can be found in [8], [10]–[11] and [16]–[17].

2. Trust region methods

We restrict our attention to trust region methods, which have shown more successful than line-search methods in our numerical experiments. In the description of trust region methods, we utilize the knowledge of gradients $g_i = \nabla F(x_i)$, $i \in \mathbb{N}$, and denote

$$Q_i(s) = \frac{1}{2} s^{\mathrm{T}} A_i^{\mathrm{T}} A_i s + g_i^{\mathrm{T}} s$$

for the predicted decrease and

$$\varrho_i(s) = \frac{F(x_i + s) - F_i(x_i)}{Q_i(s)}$$

for the ratio of the actual and the predicted decreases of the merit function. Detailed desctiption of trust region methods is presented in [3], where also Definition 1 and Theorem 1 can be found.

Definition 1. We say that an iterative method $x_{i+1} = x_i + \alpha_i s_i$, $i \in \mathbb{N}$, for solving a system of nonlinear equations $f(x) = 0$, is a trust region method, if the following conditions hold.

(T1) Direction vectors $s_i \in \mathbb{R}^n$, $i \in \mathbb{N}$, are determined in such a way that

$$\|s_i\| \leqslant \Delta_i, \tag{5}$$

$$\|s_i\| < \Delta_i \Rightarrow A_i s_i + f_i = 0, \tag{6}$$

$$Q_i(s_i) \leqslant \underline{\sigma} \min_{\alpha \|g_i\| \leqslant \Delta_i} Q_i(-\alpha g_i), \tag{7}$$

where $0 < \underline{\sigma} < 1$.

(T2) Step-sizes $\alpha_i \geqslant 0$, $i \in \mathbb{N}$, are selected so that

$$\varrho_i(s_i) \leqslant 0 \Rightarrow \alpha_i = 0, \tag{8}$$

$$\varrho_i(s_i) > 0 \Rightarrow \alpha_i = 1. \tag{9}$$

(T3) Trust region radii $0 < \Delta_i \leqslant \overline{\Delta}$, $i \in \mathbb{N}$, are chosen by the rule

$$\varrho_i(s_i) < \underline{\varrho} \Rightarrow \underline{\beta}\|s_i\| \leqslant \Delta_{i+1} \leqslant \overline{\beta}\|s_i\|, \tag{10}$$

$$\underline{\varrho} \leqslant \varrho_i(s_i) \leqslant \overline{\varrho} \Rightarrow \Delta_{i+1} = \Delta_i, \tag{11}$$

$$\varrho_i(s_i) > \overline{\varrho} \Rightarrow \Delta_i \leqslant \Delta_{i+1} \leqslant \min(\gamma \Delta_i, \overline{\Delta}), \tag{12}$$

where $0 < \underline{\beta} \leqslant \overline{\beta} < 1 < \gamma$ and $0 < \underline{\varrho} < \overline{\varrho} < 1$.

The direction vector $s_i \in \mathbb{R}^n$ satisfying conditions (5)–(7) can be computed in various ways. We have chosen the dog-leg strategy, introduced in [12], which uses the formulas

$$s_i = -\frac{\Delta_i}{\|g_i\|}, \quad \|s_i^C\| \geqslant \Delta_i, \tag{13}$$

$$s_i = s_i^C + \lambda_i(s_i^N - s_i^C), \quad \|s_i^C\| < \Delta_i < \|s_i^N\|, \tag{14}$$

$$s_i = s_i^N, \quad \|s_i^N\| \leqslant \Delta_i, \tag{15}$$

where

$$s_i^C = -\frac{\|g_i\|^2}{\|A_i g_i\|^2} g_i, \quad s_i^N = -A_i^{-1} f_i, \tag{16}$$

and λ_i is a number selected in such a way that $\|s_i\| = \Delta_i$. It is known (see [3]) that the direction vector s_i computed by (13)–(16) satisfies conditions (5)–(7) with $\underline{\sigma} = 1/2$.

The following assertion follows from the theorem introduced in [13].

Theorem 1. *Let the mapping* $f\colon \mathbb{R}^n \to \mathbb{R}^n$ *satisfy assumptions* J1–J3 *and matrices* A_i, $i \in \mathbb{N}$, *have bounded norms. Let* $x_i \in \mathbb{R}^n$, $i \in \mathbb{N}$, *be a sequence generated by the trust region method* (T1)–(T3). *Then* $f(x_i) \to 0$.

Notice that the sequence generated by the trust-region method (T1)–(T3) can converge to a stationary point of the function $F(x)$ which is not a solution of the system $f(x) = 0$, when Assumption J3 is not satisfied.

In the subsequent considerations, we assume that matrices $A_i \approx J_i$, $i \in \mathbb{N}$, used in Definition 1, are obtained by quasi-Newton updates described in the next section. In this case, a safeguard against the loss of convergence is necessary. In our implementation of the trust region method, we use restarts, which consist in setting $A_i = J_i$ and repeating the computation of s_i by (T1) when $A_i \neq J_i$ and $\varrho_i(s_i) \leqslant 0$.

3. Quasi-Newton methods

Quasi-Newton methods, which are surveyed in [4] and [16], use nonsingular matrices A_i, $i \in \mathbb{N}$, which are computed recursively by the formula $A_{i+1} = A_i + u_i v_i^{\mathrm{T}}$ to satisfy the quasi-Newton condition $A_{i+1} d_i = y_i$, where $d_i = x_{i+1} - x_i$ and $y_i = f_{i+1} - f_i$. It can be easily shown that the quasi-Newton condition holds if $v_i^{\mathrm{T}} d_i \neq 0$ and $u_i = (y_i - A_i d_i)/v_i^{\mathrm{T}} d_i$. To simplify the notation, we frequently omit the index i and replace $i+1$ by the symbol $+$. Thus we can write

$$A_+ = A + \frac{(y - Ad)v^{\mathrm{T}}}{v^{\mathrm{T}} d}, \tag{17}$$

where the vector v is a free parameter. Setting $v = d$, we get an efficient and broadly used Broyden's good method [2]. Further efficient methods can be obtained by minimizing the condition number $\kappa(M) = \|M\|\|M^{-1}\|$ or the number $\|I - M\|\|I - M^{-1}\|$, where

$$M = A^{-1}A_+ = I - \frac{(d - A^{-1}y)v^{\mathrm{T}}}{v^{\mathrm{T}}d} = I - \frac{(d - w)v^{\mathrm{T}}}{v^{\mathrm{T}}d} \tag{18}$$

(with $w = A^{-1}y$). The following theorem is proved in [7].

Theorem 2. *Let A_+ be the matrix determined by formula* (17), *so that* (18) *holds. Assume that vectors d and w are linearly independent and denote $a = d^{\mathrm{T}}d$, $b = d^{\mathrm{T}}w$, $c = w^{\mathrm{T}}w$, so that $a > 0$, $b > 0$ and $ac > b^2$. Then $\|I - M\|\|I - M^{-1}\|$ is minimized if and only if $v = \theta d - w = \theta d - A^{-1}y$, where*

$$\theta = \sqrt{c/a} \quad \text{if } b \leqslant 0,$$
$$\theta = -\sqrt{c/a} \quad \text{if } b > 0.$$

Quasi-Newton methods find the solution of a linear system after a finite number of steps. The following theorem is proved in [5].

Theorem 3. *Let x_i, $i \in \mathbb{N}$, be a sequence generated by a quasi-Newton method of the form* (17) *with $A_i s_i + f_i = 0$ and $\alpha_i = 1$ (so $d_i = s_i$), $i \in \mathbb{N}$, applied to the system of linear equations $J(x - x^*) = 0$ with a nonsingular matrix J. Let $f_i = J(x_i - x^*) \neq 0$, $1 \leqslant i \leqslant 2n$. Then $f_{2n+1} = J(x_{2n+1} - x^*) = 0$ and $x_{2n+1} = x^*$.*

Quasi-Newton methods can be derived variationally by the following theorem [4].

Theorem 4. *Let W be a square nonsingular matrix of order n. Then the matrix A_+, which is a solution of the variational problem*

$$\|(A_+ - A)W^{-1}\|_F = \min_{\widetilde{A}} \|(\widetilde{A} - A)W^{-1}\|_F \quad \text{s.t. } \widetilde{A}d = y, \tag{19}$$

can be expressed in the form (17), *where $v = W^{\mathrm{T}}Wd$.*

Setting $W = I$ in (19), we obtain Broyden's good update, which corresponds to the orthogonal projection of A into the linear manifold defined by the quasi-Newton condition $A_+d = y$. Such update satisfies the bounded deterioration principle: there exists a constant $\overline{c}$ such that

$$\|A_{i+1} - J_{i+1}\| \leqslant \|A_i - J_i\| + \overline{c}\|d_i\|, \quad i \in \mathbb{N}. \tag{20}$$

The bounded deterioration principle can be used for proving the following local convergence theorem [4].

Theorem 5. *Let $x^* \in \mathbb{R}^n$ be a point such that $f(x^*) = 0$ and the Jacobian matrix $J(x^*)$ is nonsingular. Then there are numbers $\overline{\delta} > 0$ and $\overline{\theta} > 0$ such that if $\|x_1 - x^*\| \leqslant \overline{\delta}$ and $\|A_1 - J_1\| \leqslant \overline{\theta}$, the sequence x_i, $i \in \mathbb{N}$, generated by Broyden's good quasi-Newton method with unit step-sizes ($\alpha_i = 1$, $i \in \mathbb{N}$), converges Q-superlinearly to the point x^*.*

If the first derivatives are available, the standard quasi-Newton condition can be replaced by a stronger condition $A_{i+1} d_i = J_{i+1} d_i$. Alternatively, the adjoint quasi-Newton condition $A_{i+1}^{\mathrm{T}} w_i = J_{i+1}^{\mathrm{T}} w_i$ can be used (if $w_i = f_{i+1}$, then $g_{i+1} = J_{i+1}^{\mathrm{T}} f_{i+1} = A_{i+1}^{\mathrm{T}} f_{i+1}$). In this way, we obtain adjoint quasi-Newton methods, where matrices A_i, $i \in \mathbb{N}$, are chosen recursively by the formula $A_{i+1} = A_i + u_i v_i^{\mathrm{T}}$ and satisfy the adjoint quasi-Newton condition $A_{i+1}^{\mathrm{T}} w_i = J_{i+1}^{\mathrm{T}} w_i$. It can be easily shown that the adjoint quasi-Newton condition holds if $w_i^{\mathrm{T}} u_i \neq 0$ and $v_i = (J_{i+1} - A_i)^{\mathrm{T}} w_i / w_i^{\mathrm{T}} u_i$. Thus, we can write

$$A_+ = A + \frac{u w^{\mathrm{T}} (J_+ - A)}{w^{\mathrm{T}} u}. \tag{21}$$

Using the well known Sherman-Morrison formula, we can see that A_+ is nonsingular if and only if $w^{\mathrm{T}} J_+ A^{-1} u \neq 0$. In the subsequent considerations, we will assume that $w^{\mathrm{T}} u \neq 0$ and $w^{\mathrm{T}} J_+ A^{-1} u \neq 0$. These conditions are usually checked algorithmically and the updates are skipped if necessary.

Adjoint quasi-Newton methods can be derived variationally by the following theorem.

Theorem 6. *Let W be a square nonsingular matrix of order n. Then the matrix A_+, which is a solution of the variational problem*

$$\|(A_+ - A)^{\mathrm{T}} W^{-1}\|_F = \min_{\widetilde{A}} \|(\widetilde{A} - A)^{\mathrm{T}} W^{-1}\|_F \quad \text{s.t. } \widetilde{A}^{\mathrm{T}} w = J_+^{\mathrm{T}} w, \tag{22}$$

can be expressed in the form (21), where $u = W^{\mathrm{T}} W w$.

P r o o f. The assertion follows from Theorem 4 after replacing A, A_+, d, and y by A^{T}, A_+^{T}, w, and $J_+^{\mathrm{T}} w$, respectively. □

Formula (21) contains two optional vectors u and w. Setting $u = (J_+ - A)d$, we get two-sided (or tangent) quasi-Newton methods

$$A_+ = A + \frac{(J_+ - A) d w^{\mathrm{T}} (J_+ - A)}{w^{\mathrm{T}} (J_+ - A) d}, \tag{23}$$

satisfying conditions $A_+^{\mathrm{T}} w = J_+^{\mathrm{T}} w$ and $A_+ d = J_+ d$. Setting $u = y - Ad$, we obtain secant quasi-Newton methods

$$A_+ = A + \frac{(y - Ad) w^{\mathrm{T}} (J_+ - A)}{w^{\mathrm{T}} (y - Ad)}. \tag{24}$$

Putting $w = f_+$, we obtain residual quasi-Newton methods

$$A_+ = A + \frac{u\, f_+^{\mathrm{T}} (J_+ - A)}{f_+^{\mathrm{T}} u}. \tag{25}$$

This class contains the very important two-sided residual quasi-Newton method, which uses the update

$$A_+ = A + \frac{(J_+ - A) d\, f_+^{\mathrm{T}} (J_+ - A)}{f_+^{\mathrm{T}} (J_+ - A) d} \tag{26}$$

satisfying conditions $A_+ d = J_+ d$ and $A_+^{\mathrm{T}} f_+ = J_+^{\mathrm{T}} f_+$. Setting $u = w$ (or $w = u$), we come to variationally derived adjoint quasi-Newton methods (Theorem 6) with $W = I$.

If $W = I$ in (22), we get the update which is an orthogonal projection of A into the linear manifold defined by the adjoint quasi-Newton condition $A_+^{\mathrm{T}} w = J_+^{\mathrm{T}} w$. Such update satisfies the bounded deterioration principle (20), so the following local convergence theorem holds [14].

Theorem 7. *Let $x^* \in \mathbb{R}^n$ be a point such that $f(x^*) = 0$ and the Jacobian matrix $J(x^*)$ is nonsingular. Then there are numbers $\overline{\delta} > 0$ and $\overline{\theta} > 0$ such that if $\|x_1 - x^*\| \leqslant \overline{\delta}$ and $\|A_1 - J_1\| \leqslant \overline{\theta}$, the sequence x_i, $i \in \mathbb{N}$, generated by the tangent (23) or the secant (24) or the residual (25) adjoint quasi-Newton method with $w_i = u_i$, $i \in \mathbb{N}$, and with the unit step-sizes ($\alpha_i = 1$, $i \in \mathbb{N}$), converges Q-superlinearly to the point $x^* \in \mathbb{R}^n$.*

Two-sided quasi-Newton methods have excellent properties expressed by the following theorem.

Theorem 8. *Let x_i, $i \in \mathbb{N}$, be a sequence generated by the two-sided quasi-Newton method with $A_i s_i + f_i = 0$ with A_i nonsingular and $\alpha_i = 1$ (so $d_i = s_i$), $i \in \mathbb{N}$, applied to the system of linear equations $J(x - x^*) = 0$ with a nonsingular matrix J. Let $f_i = J(x_i - x^*) \neq 0$, $1 \leqslant i \leqslant n + 1$. Then $f_{n+2} = J(x_{n+2} - x^*) = 0$ and $x_{n+2} = x^*$.*

P r o o f. Assume that $f_i \neq 0$, $1 \leqslant i \leqslant n+1$. We prove by induction that, for $1 \leqslant i \leqslant n$, the vector $d_i \neq 0$ is not a linear combination of vectors d_j, $1 \leqslant j < i$, and that, for $1 \leqslant j < i \leqslant n+1$, the equalities

$$(A_i - J)d_j = 0, \tag{27}$$

$$w_j^{\mathrm{T}}(A_i - J) = 0 \tag{28}$$

hold (these equalities are mentioned in [14] without proof). Let $i = 1$. Since $A_1 d_1 = A_1 s_1 = -f_1$, $f_1 \neq 0$, and the matrix A_1 is nonsingular, we can write $d_1 \neq 0$.

The induction step:

(a) Let $1 < i \leqslant n$. Since $A_i d_i = A_i s_i = -f_i$, $f_i \neq 0$, and the matrix A_i is nonsingular, we can write $d_i \neq 0$. Since

$$f_{i+1} = J(x_i + d_i - x^*) = f_i + J d_i \neq 0$$

by assumption, we obtain

$$(A_i - J)d_i = A_i s_i + f_i - J d_i - f_i = -(f_i + J d_i) \neq 0,$$

so the vector d_i is not a linear combination of vectors d_j, $1 \leqslant j < i$.

(b) Using (26), we can write

$$A_{i+1} - J = A_i - J + \frac{(J - A_i)d_i w_i^{\mathrm{T}}(J - A_i)}{w_i^{\mathrm{T}}(J - A_i)d_i}. \tag{29}$$

Equalities (27), which hold by the inductive assumption, and the relation (29) imply that $(A_{i+1} - J)d_j = 0$ for $1 \leqslant j < i$. Furthermore,

$$(A_{i+1} - J)d_i = (A_i - J)d_i + (J - A_i)d_i = 0,$$

so $(A_{i+1} - J)d_j = 0$ for $1 \leqslant j \leqslant i$.

(c) Equalities (28), which hold by the inductive assumption, and the relation (29) imply that $w_j^{\mathrm{T}}(A_{i+1} - J) = 0$ for $1 \leqslant j < i$. Moreover,

$$w_i^{\mathrm{T}}(A_{i+1} - J) = w_i^{\mathrm{T}}(A_i - J) + w_i^{\mathrm{T}}(J - A_i) = 0,$$

so $w_j^{\mathrm{T}}(A_{i+1} - J) = 0$ for $1 \leqslant j \leqslant i$.

The induction step is finished. Since vectors d_i, $1 \leqslant i \leqslant n$, are linearly independent and (27) implies $(A_{n+1} - J)d_i = 0$, $1 \leqslant i \leqslant n$, we can write $A_{n+1} = J$ and therefore,

$$\begin{aligned} f(x_{n+2}) = J(x_{n+2} - x^*) &= J(x_{n+1} + d_{n+1} - x^*) = f_{n+1} + J d_{n+1} \\ &= f_{n+1} + A_{n+1} s_{n+1} = 0. \end{aligned}$$

□

Theorem 8 is very strong, since it guarantees that the two-sided quasi-Newton method terminates after at most $n+1$ steps, if the system is linear and certain assumptions are satisfied. Note that quasi-Newton methods of the form (17) terminate after at most $2n$ steps under the same assumptions (Theorem 3).

Adjoint quasi-Newton methods use vector $J_+^{\mathrm{T}} w$, which can be computed by backward automatic differentiation [6]. Two-sided quasi-Newton methods use the vector $J_+ d$ as well, which can be computed by forward automatic differentiation [6] or by numerical differentiation. It can be also successfully approximated by the vector $y = f_+ - f$.

If the residual quasi-Newton method is used, then $J_+^{\mathrm{T}} w = J_+^{\mathrm{T}} f_+ = g_+$, where g_+ is the gradient of the function $F(x) = \|f(x)\|^2/2$ at the point x_+. Thus (25) with $u = w = f_+$ can be rewritten in the form

$$A_+ = A + \frac{f_+(g_+ - h_+)^{\mathrm{T}}}{f_+^{\mathrm{T}} f_+}, \tag{30}$$

where $h_+ = A^{\mathrm{T}} f_+$.

The update of two-sided residual quasi-Newton method (26) can be approximated by the expression

$$A_+ = A + \frac{(y - Ad)(g_+ - h_+)^{\mathrm{T}}}{(g_+ - h_+)^{\mathrm{T}} d} \tag{31}$$

(the directional derivative $J_+ d$ is replaced by the vector y). This new method is not a two sided quasi-Newton method, since usually $y \neq J_+ d$, but its properties are similar to the properties of the residual two-sided quasi-Newton method (26), since $y \approx J_+ d$. Notice that the method (31) satisfies the quasi-Newton condition $A_+ y = d$ and has the form (17), where $v = g_+ - h_+$.

Changing the denominator in (31) in such a way that

$$A_+ = A + \frac{(y - Ad)(g_+ - h_+)^{\mathrm{T}}}{f_+^{\mathrm{T}}(y - Ad)}, \tag{32}$$

we obtain the residual secant quasi-Newton method (24) with $w = f_+$, which satisfies the condition $A_+^{\mathrm{T}} f_+ = J_+^{\mathrm{T}} f_+$ and is also a good approximation of the two-sided residual quasi-Newton method (26). Methods (30)–(32) require the computation of the gradient $g_+ = J_+^{\mathrm{T}} f_+$, but the computation of the full Jacobian matrix J_+ or the vector $J_+ d$ is not necessary.

According to (6), quasi-Newton methods determine the direction vector s by solving the system of linear equations $As + f = 0$, where A is a nonsingular square matrix.

For this purpose, one can use either the orthogonal decomposition or the triangular decomposition of the matrix A. Both the orthogonal decomposition $A = QR$, where Q is an orthogonal matrix and R is an upper triangular matrix, and the triangular decomposition $A = LU$, where L is a lower triangular matrix with units on the diagonal and U is an upper triangular matrix, require $O(n^3)$ arithmetic operations, while solutions of linear systems $QRs + f = 0$ and $LUs + f = 0$, require only $O(n^2)$ arithmetic operations. Therefore, it is advantageous to update decompositions $A = QR$ and $A = LU$ by methods requiring $O(n^2)$ arithmetic operations.

The next theorem, introduced, e.g., in [4], demonstrates the way for determining the orthogonal decomposition of the matrix $\bar{A} = A + uv^{\mathrm{T}}$ from the orthogonal decomposition of the matrix A by using $O(n^2)$ arithmetic operations.

Theorem 9. *Let $\bar{A} = A + uv^{\mathrm{T}}$, where $A = QR$, Q is an orthogonal matrix and R is an upper triangular matrix. Let $\widetilde{u} = Q^{\mathrm{T}}u$ and let $\widetilde{Q}^{\mathrm{T}}$ be an orthogonal matrix (the product of Givens elementary rotation matrices) such that $\widetilde{Q}^{\mathrm{T}}\widetilde{u} = \|\widetilde{u}\|e_1$, and the matrix $\widetilde{R} = \widetilde{Q}R$ is upper Hessenberg. Let $\widehat{Q}^{\mathrm{T}}$ be an orthogonal matrix (the product of Givens elementary rotation matrices) such that the matrix $\overline{R} = \widehat{Q}^{\mathrm{T}}(\widetilde{R} + \|\widetilde{u}\|e_1v^{\mathrm{T}})$ is upper triangular. Then $\bar{A} = \overline{Q}\,\overline{R}$, where $\overline{Q} = Q\widetilde{Q}\widehat{Q}$.*

There are two efficient methods for updating the triangular decomposition using $O(n^2)$ arithmetic operations. The simplest of them, proposed in [1], is based on the following theorem (an alternative proof is introduced).

Theorem 10. *Let L, $\overline{L}$ be lower triangular matrices with units on the diagonal and let U, $\overline{U}$ be upper triangular matrices such that*

$$\overline{L}\,\overline{U} = LU + pq^{\mathrm{T}}. \tag{33}$$

Let l_i, $\bar{l}_i$, $1 \leqslant i \leqslant n$, be the columns of matrices L, $\overline{L}$ and u_i, $\overline{u}_i$, $1 \leqslant i \leqslant n$, the transposed rows of matrices U, $\overline{U}$. Then

$$\overline{u}_i = u_i + p_{ii}q_i, \tag{34}$$

$$\bar{l}_i = \frac{u_{ii}}{\overline{u}_{ii}}l_i + \frac{q_{ii}}{\overline{u}_{ii}}p_i = l_i + \frac{q_{ii}}{\overline{u}_{ii}}p_{i+1} \tag{35}$$

for $1 \leqslant i \leqslant n$, where p_{ii}, q_{ii} are the ith entries of the vectors p_i, q_i and u_{ii}, $\overline{u}_{ii}$ are the ith diagonal entries of the matrices U, $\overline{U}$ (so $\overline{u}_{ii} = u_{ii} + p_{ii}q_{ii}$ by (34)). The vectors p_i, q_i, $1 \leqslant i \leqslant n$, are computed recursively by the relations

$$p_{i+1} = p_i - p_{ii}l_i, \tag{36}$$

$$q_{i+1} = \frac{u_{ii}}{\overline{u}_{ii}}q_i - \frac{q_{ii}}{\overline{u}_{ii}}u_i = q_i - \frac{q_{ii}}{\overline{u}_{ii}}u_{i+1} \tag{37}$$

where $p_1 = p$, $q_1 = q$.

P r o o f. The theorem is proved by induction. Assume that

$$\sum_{j=i}^{n} \bar{l}_j \overline{u}_j^{\mathrm{T}} = \sum_{j=i}^{n} l_j u_j^{\mathrm{T}} + p_i q_i^{\mathrm{T}} \tag{38}$$

holds for some index $1 \leqslant i < n$. This equality is satisfied for $i = 1$ with $p_1 = p$ and $q_1 = q$, since (33) can be written in the form

$$\sum_{j=1}^{n} \bar{l}_j \overline{u}_j^{\mathrm{T}} = \sum_{j=1}^{n} l_j u_j^{\mathrm{T}} + p q^{\mathrm{T}}.$$

Since the vectors l_j, $\bar{l}_j$, u_j, $\overline{u}_j$, $i \leqslant j \leqslant n$, have the first $j-1$ entries equal to zero, the matrix (38) has the first $i-1$ rows and $i-1$ columns equal to zero and the entries in its ith row and ith column are fully determined by the vectors u_i and l_i by formulas $\bar{l}_{ii}\overline{u}_i = l_{ii}u_i + p_{ii}q_i$ and $\bar{l}_i\overline{u}_{ii} = l_i u_{ii} + p_i q_{ii}$. Since $\bar{l}_{ii} = l_{ii} = 1$, one can write

$$\overline{u}_i = u_i + p_{ii} q_i, \tag{39}$$

$$\bar{l}_i = \frac{u_{ii}}{\overline{u}_{ii}} l_i + \frac{q_{ii}}{\overline{u}_{ii}} p_i. \tag{40}$$

Using relations (39)–(40) and formula (37), we obtain

$$\overline{u}_{ii}\bar{l}_i = u_{ii}l_i + q_{ii}p_i = (\overline{u}_{ii} - p_{ii}q_{ii})l_i + q_{ii}p_i = \overline{u}_{ii}l_i + q_{ii}(p_i - p_{ii}l_i) = \overline{u}_{ii}l_i + q_{ii}p_{i+1},$$

which after dividing it by $\overline{u}_{ii}$ gives the second equality in (35). To finish the induction step, we have to prove the relation

$$l_i u_i^{\mathrm{T}} - \bar{l}_i \overline{u}_i^{\mathrm{T}} = p_{i+1} q_{i+1}^{\mathrm{T}} - p_i q_i^{\mathrm{T}}, \tag{41}$$

which follows by subtracting (38) with $j = i + 1$ from the same equation with $j = i$. But

$$\begin{aligned} l_i u_i^{\mathrm{T}} - \bar{l}_i \overline{u}_i^{\mathrm{T}} &= l_i u_i^{\mathrm{T}} - \Big(l_i + \frac{q_{ii}}{\overline{u}_{ii}} p_{i+1}\Big)(u_i + p_{ii}q_i)^{\mathrm{T}} \\ &= l_i u_i^{\mathrm{T}} - l_i u_i^{\mathrm{T}} - p_{ii} l_i q_i^{\mathrm{T}} - \frac{q_{ii}}{\overline{u}_{ii}} p_{i+1}(u_i + p_{ii}q_i) \\ &= -p_{ii} l_i q_i^{\mathrm{T}} - \frac{q_{ii}}{\overline{u}_{ii}} p_{i+1} u_{i+1} \end{aligned}$$

and

$$p_{i+1} q_{i+1}^{\mathrm{T}} - p_i q_i^{\mathrm{T}} = (p_i - p_{ii} l_i)\Big(q_i - \frac{q_{ii}}{\overline{u}_{ii}} u_{i+1}\Big) - p_i q_i^{\mathrm{T}} = -p_{ii} l_i q_i^{\mathrm{T}} - \frac{q_{ii}}{\overline{u}_{ii}} p_{i+1} u_{i+1},$$

which proves relation (41), so the induction step is completed. □

4. Computational experiments

Methods for solving systems of nonlinear equations were tested on 62 problems with selected dimensions taken from the collection TEST37 contained in the software system for universal functional optimization UFO [9]. Table 1 contains results obtained by the following methods:

`TRNM`—Newton's method,

`TRBG`—Broyden's good method,

`TRIT`—the method of Ip and Todd (Theorem 2),

`TRRB`—residual basic adjoint quasi-Newton method (30),

`TRRT`—residual tangent adjoint quasi-Newton method (26),

`TRRS`—residual secant adjoint quasi-Newton method (32),

`TRNB`—new quasi-Newton method (31).

The above methods were implemented as dog-leg trust-region methods (5)–(16) with parameters $\underline{\varrho} = 0.1$, $\overline{\varrho} = 0.9$, $\underline{\beta} = 0.05$, $\overline{\beta} = 0.75$, $\gamma = 2$, termination criterion $\|f_i\| \leqslant 10^{-8}$ and the restart strategy described in Section 2. All methods solve linear systems by using both the orthogonal and triangular decompositions of nonsymmetric matrices. Quasi-Newton methods use updates described in Theorem 9 and Theorem 10. If the denominator of the updating formula is zero, the corresponding update is skipped.

Table 1 proposes results obtained by solving 62 problems with 200 equations, 62 problems with 300 equations, 60 problems with 400 equations, and contains the total numbers of iterations `NIT`, function evaluations `NFV`, Jacobian (or gradient) evaluations `NFJ`, matrix decompositions `NDC`, the total number of failures (number od unsolved problems) and the total computational time.

The results contained in Table 1 lead to several conclusions:

- ▷ If elements of the Jacobian matrix are given analytically, the Newton method converges rapidly and requires lowest number of iterations and function evaluations. However, this method consumes $O(n^3)$ arithmetic operations per iteration, which prolongs the computational time for larger n.
- ▷ Quasi-Newton methods of the form (17) require more iterations and function evaluations in comparison with the Newton method, but they use $O(n^2)$ arithmetic operations in a greater part ($\approx 90\%$) of iterations.
- ▷ Adjoint quasi-Newton methods (23) and (32) converge faster than standard quasi-Newton methods and use $O(n^2)$ arithmetic operations in a greater part of iterations as well.
- ▷ The new method (31), which is of the form (17), is surprisingly competitive with the Newton method, measured by the number of iterations and function evalua-

tions. Its properties are similar to those of residual adjoint quasi-Newton methods (23) and (32), but directional derivatives $J_{i+1}d_i$, $i \in \mathbb{N}$, need not be computed. The new method shows to be very efficient in comparison with the methods introduced in Table 1.

▷ Triangular decompositions save arithmetic operations, so the methods with triangular decompositions are faster. At the same time, the methods with orthogonal decompositions are more stable.

	Orthogonal decomposition						Triangular decomposition					
$n = 200$	NIT	NFV	NFJ	NDC	Fail	Time	NIT	NFV	NFJ	NDC	Fail	Time
TRNM	1418	1525	1418	1319	–	13.40	1404	1559	1404	1236	–	1.20
TRBG	2605	3000	281	277	–	4.74	2103	2568	308	305	1	0.90
TRIT	1948	2212	215	210	1	3.88	2081	2525	282	279	1	0.92
TRRB	3576	3939	4239	360	–	6.98	2580	3047	3405	418	–	1.23
TRRT	1732	1915	2040	185	–	3.63	2222	2581	2810	291	–	1.24
TRRS	2277	2439	2544	165	–	3.99	1826	2164	2375	273	–	0.95
TRNB	1650	1838	1947	169	–	3.37	1923	2169	2343	234	–	0.95
$n = 300$	NIT	NFV	NFJ	NDC	Fail	Time	NIT	NFV	NFJ	NDC	Fail	Time
TRNM	1450	1591	1450	1379	–	44.30	1892	2087	1892	1759	–	6.13
TRBG	2865	3344	363	360	–	15.61	2299	2736	280	278	1	2.16
TRIT	2464	2893	270	268	1	13.86	2085	2498	280	278	1	2.34
TRRB	3804	4163	4464	361	1	19.46	3502	4163	4658	556	1	3.85
TRRT	2051	2237	2355	178	–	10.71	2302	2732	3018	346	–	3.02
TRRS	1807	1981	2094	173	–	9.71	2788	3125	3386	322	–	3.08
TRNB	1727	1984	2112	198	–	10.23	2227	2571	2789	278	–	2.59
$n = 400$	NIT	NFV	NFJ	NDC	Fail	Time	NIT	NFV	NFJ	NDC	Fail	Time
TRNM	1036	1134	1036	739	1	58.29	1128	1259	1128	1019	1	4.30
TRBG	2178	2488	201	197	–	23.30	2091	2394	195	192	–	3.66
TRIT	2140	2412	203	200	–	24 00	2199	2494	191	188	–	4.13
TRRB	2804	3103	3338	293	–	33.63	1854	2179	2388	266	–	3.93
TRRT	2200	2377	2489	170	1	23.07	1836	2137	2307	227	1	4.52
TRRS	1405	1561	1658	155	–	18.15	1863	2122	2300	235	–	4.18
TRNB	1474	1662	1751	147	–	17.94	1546	1787	1907	177	–	3.54

Table 1. TEST37

References

[1] *J. M. Bennett*: Triangular factors of modified matrices. Numer. Math. *7* (1965), 217–221. zbl MR doi

[2] *C. G. Broyden*: A class of methods for solving nonlinear simultaneous equations. Math. Comput. *19* (1965), 577–593. zbl MR doi

[3] *A. R. Conn, N. I. M. Gould, P. L. Toint*: Trust-Region Methods. MPS/SIAM Series on Optimization 1, Society for Industrial and Applied Mathematics, Philadelphia; Mathematical Programming Society, Philadelphia, 2000. zbl MR doi

[4] *J. E. Dennis,* Jr., *R. B. Schnabel*: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Classics in Applied Mathematics 16, Society for Industrial and Applied Mathematics, Philadelphia, 1996. zbl MR doi

[5] *D. M. Gay*: Some convergence properties of Broyden's method. SIAM J. Numer. Anal. *16* (1979), 623–630. zbl MR doi

[6] *A. Griewank, A. Walther*: Evaluating Derivatives. Principles and Techniques of Algorithmic Differentiation. Society for Industrial and Applied Mathematics, Philadelphia, 2008. zbl MR doi

[7] *C. M. Ip, M. J. Todd*: Optimal conditioning and convergence in rank one quasi-Newton updates. SIAM J. Numer. Anal. *25* (1988), 206–221. zbl MR doi

[8] *L. Lukšan*: Inexact trust region method for large sparse systems of nonlinear equations. J. Optimization Theory Appl. *81* (1994), 569–590. zbl MR doi

[9] *L. Lukšan, M. Tůma, J. Vlček, N. Ramešová, M. Šiška, J, Hartman, C. Matonoha*: UFO 2014. Interactive System for Universal Functional Optimization. Technical Report V-1218. Prague, Institute of Computer Science AS CR, Praha, 2014.

[10] *L. Lukšan, J. Vlček*: Truncated trust region methods based on preconditioned iterative subalgorithms for large sparse systems of nonlinear equations. J. Optimization Theory Appl. *95* (1997), 637–658. zbl MR doi

[11] *L. Lukšan, J. Vlček*: Computational experience with globally convergent descent methods for large sparse systems of nonlinear equations. Optim. Methods Softw. *8* (1998), 201–223. zbl MR doi

[12] *M. J. D. Powell*: A new algorithm for unconstrained optimization. Nonlinear Programming, Proc. Sympos., Univ. of Wisconsin, Madison (J. B. Rosen, O. L. Mangasarian, K. Ritter, eds.). Academic Press, London, 1970, pp. 31–65. zbl MR doi

[13] *M. J. D. Powell*: On the global convergence of trust region algorithms for unconstrained minimization. Math. Program. *29* (1984), 297–303. zbl MR doi

[14] *S. Schlenkrich, A. Griewank, A. Walther*: On the local convergence of adjoint Broyden methods. Math. Program. *121* (2010), 221–247. zbl MR doi

[15] *S. Schlenkrich, A. Walther*: Global convergence of quasi-Newton methods based on adjoint Broyden updates. Appl. Numer. Math. *59* (2009), 1120–1136. zbl MR doi

[16] *Y.-X. Yuan*: Recent advances in numerical methods for nonlinear equations and nonlinear least squares. Numer. Algebra Control Optim. *1* (2011), 15–34. zbl MR doi

[17] *Y.-X. Yuan*: Recent advances in trust region algorithms. Math. Program. *151* (2015), 249–281. zbl MR doi

Authors' address: *Ladislav Lukšan*, *Jan Vlček*, Institute of Computer Science, Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 07 Praha 8, Czech Republic, e-mail: `luksan@cs.cas.cz`, `vlcek@cs.cas.cz`.