New Quasi-Newton Method for Solving Systems of Nonlinear Equations

Lukšan, Ladislav
2016

Institute of Computer Science
Academy of Sciences of the Czech Republic

New quasi-Newton method for solving systems of nonlinear equations

Ladislav Lukšan, Jan Vlček

Technical report No. 1233

September 2016

Pod Vodárenskou věží 2, 182 07 Prague 8 phone: +420 2 688 42 44, fax: +420 2 858 57 89,
e-mail:e-mail:ics@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

New quasi-Newton method for solving systems of nonlinear equations

Ladislav Lukšan, Jan Vlček [1]

Technical report No. 1233

September 2016

Abstract:

In this paper, we propose the new Broyden method for solving systems of nonlinear equations, which uses the first derivatives, but it is more efficient than the Newton method (measured by the computational time) for larger dense systems. The new method updates QR 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 equations, systems of equations, trust-region methods, quasi-Newton methods, adjoint Broyden methods, numerical algorithms, numerical experiments.

[1] 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).

1 Introduction

Consider the system of nonlinear equations

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

where $f : \mathcal{R}^n \to \mathcal{R}^n$ is a nonlinear mapping, and denote $J(x)$ 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 mapping f.

Assumption J1. The mapping $f : \mathcal{R}^n \to \mathcal{R}^n$ is continuously differentiable on the level set $\mathcal{D}(\overline{F}) = \{x \in \mathcal{R}^n : \|f(x)\| \le \overline{F}\}$, where $\overline{F}$ is a suitable upper bound, and the Jacobian matrix J is Lipschitz continuous on $\mathcal{D}(\overline{F})$, i.e., there is a constant $\overline{L} > 0$ such that

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

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

$$\|J(x_i)s\| \le \overline{J}\|s\| \qquad \forall i \in \mathcal{N} \qquad \forall s \in \mathcal{R}^n. \tag{3}$$

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

$$\|J(x_i)s\| \ge \underline{J}\|s\| \qquad \forall i \in \mathcal{N} \qquad \forall s \in \mathcal{R}^n. \tag{4}$$

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

We restrict our attention on iterative methods of the form $x_{i+1} = x_i + \alpha_i s_i$, $i \in \mathcal{N}$, with $A_i s_i + f_i \approx 0$, $f_i = f(x_i)$, $A_i \approx J_i = J(x_i)$ and $\alpha_i \ge 0$, which generate a monotone non-increasing sequence of norms $\|f(x_i)\|$, $i \in \mathcal{N}$. 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 $\nabla F(x) = J(x)^T f(x)$ is computed either analytically or by reverse automatic differentiation. The Newton method, which is the most known and rapidly convergent method of this type, uses matrices $A_i = J_i$, $i \in \mathcal{N}$. Since the Jacobian matrix J_i is completely recomputed in every iteration, the solution of 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, $i \in \mathcal{N}$, 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, against the method (26), it does not require additional computation of directional derivatives $J_{i+1}d_i$, $i \in \mathcal{N}$ (the computation of gradients $J_{i+1}^T f_{i+1}$, $i \in \mathcal{N}$, 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 [2], [3] and [13]–[14]. Further information can be found in [7] and [9]–[10].

2 Trust region methods

We restrict our attention on 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 $g_i = \nabla F(x_i)$, $i \in \mathcal{N}$, and denote

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

for the predicted decrease and

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

for the ratio of both the actual and the predicted decreases of the merit function. Detailed desctiption of trust region methods is introduced in [2], where also Definition 1 and Theorem 1 can be found.

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

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

$$\begin{aligned} \|s_i\| &\le \Delta_i, & (5)\\ \|s_i\| &< \Delta_i \quad \Rightarrow \quad A_i s_i + f_i = 0, & (6)\\ Q_i(s_i) &\le \underline{\sigma} \min_{\alpha\|g_i\| \le \Delta_i} Q_i(-\alpha g_i), & (7) \end{aligned}$$

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

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

$$\begin{aligned} \rho_i(s_i) \le 0 \quad &\Rightarrow \quad \alpha_i = 0, & (8)\\ \rho_i(s_i) > 0 \quad &\Rightarrow \quad \alpha_i = 1. & (9) \end{aligned}$$

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

$$\begin{aligned} \rho_i(s_i) < \underline{\rho} \quad &\Rightarrow \quad \underline{\beta}\|s_i\| \le \Delta_{i+1} \le \overline{\beta}\|s_i\|, & (10)\\ \underline{\rho} \le \rho_i(s_i) \le \overline{\rho} \quad &\Rightarrow \quad \Delta_{i+1} = \Delta_i, & (11)\\ \rho_i(s_i) > \overline{\rho} \quad &\Rightarrow \quad \Delta_i \le \Delta_{i+1} \le \min(\gamma\Delta_i, \overline{\Delta}), & (12) \end{aligned}$$

where $0 < \underline{\beta} \le \overline{\beta} < 1 < \gamma$ and $0 < \underline{\rho} < \overline{\rho} < 1$.

Direction vector $s_i \in \mathcal{R}^n$ satisfying conditions (5)–(7) can be computed by various ways. We have chosen the dog-leg strategy, introduced in [11], which uses the following formulas

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

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

$$s_i = s_i^N, \qquad \|s_i^N\| \leq \Delta_i, \tag{15}$$

where

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

and λ_i is a number selected in such a way that $\|s_i\| = \Delta_i$. It is known (see [2]) that 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 [12].

Theorem 1 *Let the mapping $f : \mathcal{R}^n \to \mathcal{R}^n$ satisfy assumptions J1 – J3 and matrices A_i, $i \in \mathcal{N}$, have bounded norms. Let $x_i \in \mathcal{R}^n$, $i \in \mathcal{N}$, 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 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 $A_i \approx J_i$, $i \in \mathcal{N}$, 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 $\rho_i(s_i) \leq 0$.

3 Quasi-Newton methods

Quasi-Newton methods, which are surveyed in [3], use matrices A_i, $i \in \mathcal{N}$, which are computed recursively by the formula $A_{i+1} = A_i + u_i v_i^T$ 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 $v_i^T d_i \neq 0$ and $u_i = (y_i - A_i d_i)/v_i^T d_i$. To simplify the notation, we frequently omit index i and replace $i+1$ by symbol $+$. Thus we can write

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

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

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

(with $w = A^{-1}y$). The following theorem is proved in [6].

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

$$\begin{aligned} \vartheta &= \sqrt{c/a}, \quad \text{if} \quad b \le 0, \\ \vartheta &= -\sqrt{c/a}, \quad \text{if} \quad b > 0. \end{aligned}$$

Quasi-Newton methods find the solution of a linear system after a finite number of steps. The following theorem is proved in [4].

Theorem 3 *Let x_i, $i \in \mathcal{N}$, 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 $d_i = s_i$), $i \in \mathcal{N}$, applied to the system of linear equations $J(x - x^*) = 0$ with nonsingular matrix J. Let $f_i = J(x_i - x^*) \neq 0$, $1 \le i \le 2n$. 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 [3].

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

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

can be expressed in the form (17), where $v = W^T W d$.

Setting $W = I$ in (19), we obtain the 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

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

The bounded deterioration principle can be used for proving the following local convergence theorem [3].

Theorem 5 *Let $x^* \in \mathcal{R}^n$ be a point such that $f(x^*) = 0$ and the Jacobian matrix $J(x^*)$ is regular. Then there are numbers $\overline{\delta} > 0$ and $\overline{\vartheta} > 0$ such that if $\|x_1 - x^*\| \le \overline{\delta}$ and $\|A_1 - J_1\| \le \overline{\vartheta}$, the sequence x_i, $i \in \mathcal{N}$, generated by Broyden's good quasi-Newton method with the unit step-sizes ($\alpha_i = 1$, $i \in \mathcal{N}$), 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 $A_{i+1}^T w_i = J_{i+1}^T w_i$ can be used (if $w_i = f_{i+1}$, then $g_{i+1} = J_{i+1}^T f_{i+1} = A_{i+1}^T f_{i+1}$). In this way, we obtain adjoint quasi-Newton methods, where matrices A_i, $i \in \mathcal{N}$, are chosen recursively by the formula $A_{i+1} = A_i + u_i v_i^T$ and satisfy the adjoint quasi-Newton condition $A_{i+1}^T w_i = J_{i+1}^T w_i$. It can be easily shown that the adjoint quasi-Newton condition holds if $w_i^T u_i \neq 0$ and $v_i = (J_{i+1} - A_i)^T w_i / w_i^T u_i$. Thus we can write

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

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 matrix A_+, which is a solution of the variational problem*

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

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

Proof The assertion follows from Theorem 4 after replacing A, A_+, d and y by A^T, A_+^T, w and $J_+^T w$. □

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

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

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

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

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

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

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

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

Setting $u = w$ (or $w = u$), we obtain 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 $A_+^T w = J_+^T w$. Such update satisfies the bounded deterioration principle (20), so the following local convergence theorem holds [13].

Theorem 7 *Let $x^* \in \mathcal{R}^n$ be a point such that $f(x^*) = 0$ and the Jacobian matrix $J(x^*)$ is regular. Then there are numbers $\overline{\delta} > 0$ and $\overline{\vartheta} > 0$ such that if $\|x_1 - x^*\| \le \overline{\delta}$ and $\|A_1 - J_1\| \le \overline{\vartheta}$, the sequence x_i, $i \in \mathcal{N}$, generated by the tangent (23) or the secant (24) or the residual (30) adjoint quasi-Newton method with $w_i = u_i$, $i \in \mathcal{N}$, and with the unit step-sizes ($\alpha_i = 1$, $i \in \mathcal{N}$), converges Q-superlinearly to the point $x^* \in \mathcal{R}^n$.*

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

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

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

$$\begin{aligned} (A_i - J)\, d_j &= 0, && (27) \\ w_j^T (A_i - J) &= 0 && (28) \end{aligned}$$

hold (these equalities are mentioned in [13] 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 $1 < i \le n$. 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 vector d_i is not a linear combination of vectors d_j, $1 \le j < i$.

(b) Using (26), we can write

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

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

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

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

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

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

$$f(x_{n+2}) = J(x_{n+2} - x^*) = J(x_{n+1} + d_{n+1} - x^*) = f_{n+1} + Jd_{n+1} = f_{n+1} + A_{n+1}s_{n+1} = 0.$$

□

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

Adjoint quasi-Newton methods use vector $J_+^T w$, which can be computed by backward automatic differentiation [5]. Two-sided quasi-Newton methods use vector $J_+ d$ as well, which can be computed by forward automatic differentiation [5] or by numerical differentiation. It can be also successfully approximated by vector $y = f_+ - f$.

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

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

where $h_+ = A^T f_+$.

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

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

(the directional derivative $J_+ d$ is replaced by vector y). This new method is not a two sided quasi-Newton method, since usually $y \ne J_+ d$, 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) has the form (17), where $v = g_+ - h_+$.

Changing denominator in (31) in such a way that

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

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

According to (6), quasi-Newton methods determine direction vector s by solving the system of linear equations $As + f = 0$, where A is a regular square matrix. Therefore it is advantageous to work with the orthogonal decomposition $A = QR$, where Q is an orthogonal matrix and R is an upper triangular matrix (this orthogonal decomposition can be obtained by the Householder method). Then the solution of system of nonlinear equations $QRs + f = 0$ requires $O(n^2)$ arithmetic operations. The following theorem, introduced, e.g., in [3], demonstrates the way for determining the orthogonal decomposition of matrix $\bar{A} = A + uv^T$ from the orthogonal decomposition of matrix A using $O(n^2)$ arithmetic operations.

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

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 [8]. 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 (23),
`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 $\underline{\rho} = 0.1$, $\overline{\rho} = 0.9$, $\underline{\beta} = 0.05$, $\overline{\beta} = 0.75$, $\gamma = 2$, termination criterion $\|f_i\| \le 10^{-8}$ and the restart strategy described in Section 2. All methods solve linear systems by using orthogonal decompositions of nonsymmetric matrices. Quasi-Newton methods use updates described in Theorem 9.

Table 1 proposes results obtained by solving 62 problems with 100 equations, 62 problems with 200 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.

$n = 100$	NIT	NFV	NFJ	NDC	fails	time
TRNM	910	1019	910	837	2	1.46
TRBG	2305	2623	241	237	-	1.03
TRIT	2112	2487	261	257	-	1.02
TRRB	3505	3967	4396	485	-	1.98
TRRT	1553	1769	1918	207	-	0.87
TRRS	2230	2419	2547	186	1	1.01
TRNB	1500	1696	1823	186	-	0.79
$n = 200$	NIT	NFV	NFJ	NDC	fails	time
TRNM	1383	1486	1383	1279	-	17.39
TRBG	2651	3038	292	288	-	6.56
TRIT	2591	3015	294	294	-	6.73
TRRB	4454	4773	5049	336	-	9.96
TRRT	1736	1923	2049	186	-	4.69
TRRS	2253	2413	2518	167	-	5.17
TRNB	1643	1826	1934	168	-	4.39
$n = 400$	NIT	NFV	NFJ	NDC	fails	time
TRNM	1039	1136	1039	830	1	194.03
TRBG	2207	2525	210	206	-	59.07
TRIT	2076	2349	199	194	-	56.11
TRRB	3264	3646	3945	357	-	98.81
TRRT	2198	2373	2483	168	1	52.10
TRRS	1401	1557	1652	155	-	43.64
TRNB	1496	1680	1768	146	-	42.69

Table 1: TEST37

Results introduced in Table 1 imply 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 evaluations. Its

properties are similar to properties of residual adjoint quasi-Newton methods (23) and (32), but directional derivatives $J_{i+1}d_i$, $i \in \mathcal{N}$, need not be computed. The new method gives the best results in comparisons introduced in Table 1.

References

[1] C.G.Broyden: A class of methods for solving nonlinear simultaneous equations. Mathematics of Computation 19 (1965) 577-593.

[2] A.R. Conn, N.I.M. Gould, P.L. Toint: *Trust-Region Methods* SIAM, Philadelphia, 2000.

[3] J.E. Dennis, R.B. Schnabel: Numerical methods for unconstrained optimization and nolinear equations. SIAM, Philadelphia, 1996. (2002) 201-213.

[4] D.M. Gay: Some convergence properties of Broyden's method. SIAM J. Numerical Analysis 16 (1979), 623-630.

[5] A. Griewank, A. Walther: *Evaluating Derivatives* SIAM, Philadelphia, 2008.

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

[7] L. Lukšan: Inexact Trust Region Method for Large Sparse Systems of Nonlinear Equations. Journal of Optimization Theory and Applications, Vol.81, 1994, No. 3, pp.569-590.

[8] L. Lukšan, M. Tůma, J. Vlček, N. Ramešová, M. Šiška, J, Hartman, C. Matonoha: UFO 2013. Interactive System for Universal Functional Optimization. Technical Report V-1191. Prague, Institute of Computer Science AS CR, Prague 2013.

[9] L. Lukšan, J. Vlček: Truncated Trust Region Methods Based on Preconditioned Iterative Subalgorithms for Large Sparse Systems of Nonlinear Equations. Journal of Optimization Theory and Applications, Vol. 95, 1997, No. 3. pp.637-658.

[10] L. Lukšan, J. Vlček: Computational Experience with Globally Convergent Descent Methods for Large Sparse Systems of Nonlinear Equations. Optimization Methods and Software, Vol. 8, 1998, pp.201-223.

[11] 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.

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

[13] S.Schlenkrich, A.Griewank, A.Walther: On the local convergence of adjoint Broyden methods. Mathematical Programming 121 (2010),221-247.

[14] S.Schlenkrich, A.Walther: Global convergence of quasi-Newton methods based on adjoint Broyden updates. Applied Numeerical Mathematics 59 (2009),1120-1136.