

## A BRIEF SURVEY OF METHODS FOR SOLVING NONLINEAR LEAST-SQUARES PROBLEMS

HASSAN MOHAMMAD<sup>1</sup> \*, MOHAMMED YUSUF WAZIRI<sup>1</sup> AND  
SANDRA AUGUSTA SANTOS<sup>2</sup>

<sup>1</sup>Department of Mathematical Sciences  
Faculty of Physical Sciences  
Bayero University, Kano, 700241, Nigeria

<sup>2</sup>Institute of Mathematics, Statistics and Scientific Computing  
University of Campinas, Campinas, SP, 13083-970, Brazil

(Communicated by Hongchao Zhang)

ABSTRACT. In this paper, we present a brief survey of methods for solving nonlinear least-squares problems. We pay specific attention to methods that take into account the special structure of the problems. Most of the methods discussed belong to the quasi-Newton family (i.e. *the structured quasi-Newton methods (SQN)*). Our survey comprises some of the traditional and modern developed methods for nonlinear least-squares problems. At the end, we suggest a few topics for further research.

**1. Introduction.** This paper provides a survey of methods to find a minimizer for the nonlinear least-squares problem (NLS). Specifically, we consider the problem of determining a minimizer  $x^* \in \mathbb{R}^n$  to the problem of the form

$$\min_{x \in \mathbb{R}^n} f(x), \quad f(x) = \frac{1}{2} \|F(x)\|^2 = \frac{1}{2} \sum_{i=1}^m (F_i(x))^2. \quad (1)$$

Here  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  ( $m \geq n$ ),  $F(x) = (F_1(x), F_2(x), \dots, F_m(x))^T$ ,  $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ , and  $F$  is a twice continuously differentiable mapping. This type of problem arises in many applications such as data fitting, optimal control, parameter estimation, experimental design, data assimilation, and imaging problems (see, e.g. [12, 17, 22, 25, 28, 33, 34, 56] and references therein). These problems are solved either by general unconstrained minimization algorithms, or by specialized algorithms that take into account the particular structure of the objective function  $f$  [55, 68].

The gradient and Hessian of the objective function (1) have a special form and are given by

$$\nabla f(x) = \sum_{i=1}^m F_i(x) \nabla F_i(x) = J(x)^T F(x) \quad (2)$$

---

2010 *Mathematics Subject Classification.* Primary: 93E24, 49M37; Secondary: 65K05, 90C53, 49J52.

*Key words and phrases.* Nonlinear least squares, structured quasi-Newton methods, sizing techniques, hybridization methods, derivative-free methods.

\* Corresponding author: Hassan Mohammad.

$$\nabla^2 f(x) = \sum_{i=1}^m \nabla F_i(x) \nabla F_i(x)^T + \sum_{i=1}^m F_i(x) \nabla^2 F_i(x) \quad (3)$$

$$= J(x)^T J(x) + B(x), \quad (4)$$

where  $F_i$  is the  $i$ -th component of  $F$  and  $\nabla^2 F_i(x)$  is its Hessian,  $J(x) = \nabla F(x)^T$  is the Jacobian of the so-called *residual function*  $F$  at  $x$  and  $B(x)$  is a square matrix representing the second-order term of (3).

Basically, the iterative scheme for solving (1) has the general form: given an initial approximation  $x_0$ , a sequence of iterates  $\{x_k\}$  is obtained via

$$x_{k+1} = x_k + s_k, \quad k = 0, 1, \dots \quad (5)$$

where  $s_k = \alpha_k d_k$ ,  $\alpha_k$  is the step length obtained by a suitable line search procedure, and  $d_k$  is the search direction. Along the text, we adopt the notation  $F_k = F(x_k)$ ,  $J_k = J(x_k)$  and  $g_k = \nabla f(x_k) = J(x_k)^T F(x_k)$ .

It is vital to notice that the Hessian matrix (3) consists of two parts; the first-order term contains only the first derivatives of the residual, and the second-order term contains the second derivatives of the residuals.

There are many methods for solving (1), either first- or second-order based. Most classical and modified algorithms require the computation of the gradient, and also demand at least the first order-term of the Hessian matrix of the objective function (1), which may be costly as dimension increases.

However, if the residue, i.e. the objective function value, is large at the minimizer, then the computation of the exact second-order derivatives of  $F$ , although expensive, might become relevant. Nonzero residues are present in many practical applications (see [14, 55]). Unfortunately, Newton-type methods that require the exact Hessian matrix are not suitable for large-scale problems. As a consequence, alternative approaches, which make use of the first derivative information, or just rest upon function evaluations have been developed (see [42] for details).

The Gauss-Newton method (GN) [24] is a well-known method for solving (1), and it requires the computation of the Jacobian  $J_k$  of the function  $F_k$ , and the solution of a linear system at each iteration. Provided  $J_k^T J_k$  is nonsingular, the Gauss-Newton direction can be written as

$$d_k^{GN} = -(J_k^T J_k)^{-1} g_k, \quad (6)$$

where  $g_k = J_k^T F_k$ .

Another method for solving (1) is the Levenberg-Marquardt method (LM) [31, 36, 40]. This method adds some positive scalar correction to the term  $J_k^T J_k$  in the GN method. The correction is a regularization parameter added specifically to remedy the possibility of the Jacobian matrix to become rank-deficient. The LM direction is given by

$$d_k^{LM} = -(J_k^T J_k + \mu_k I)^{-1} g_k, \quad (7)$$

where  $\mu_k$  is a positive regularization or damping parameter.

Both GN and LM methods neglect the second order part of the Hessian matrix of the objective function (1). As a result, they are expected to perform well when solving small residual problems. Nevertheless, when solving large residual problems, these methods may perform poorly [14, 55]. The need for storing the Jacobian matrix of the GN and LM methods also contributes to their inefficiency in the case of solving large-scale problems. Thus, derivative-free approaches are needed for solving such cases (see, for example, [7, 29, 47, 49–51, 59–61, 69, 70]).

Our survey focuses mainly on two categories of methods for solving nonlinear least-squares problems. The first category involves the use of structured quasi-Newton methods (SQN), whereas the second one concerns derivative-free methods for minimizing nonlinear least squares based on the Gauss-Newton method and hybrid methods. These categories are detailed in Sections 2 and 3, respectively. In Section 4 we describe additional works that do not fit into these two categories, mainly concerning miscellaneous methods as well as convergence analysis results. Final remarks and open research questions are raised in Section 5.

**2. Structured Quasi-Newton Methods for Nonlinear Least Squares.** Structured quasi-Newton methods (SQN) were devised to overcome the difficulties that arise when solving large residual problems using the Gauss-Newton method [44, 55]. These methods combine Gauss-Newton and quasi-Newton ideas in order to make good use of the structure of the Hessian matrix of the objective function. The search direction is obtained from

$$(A_k + B_k)d_k^{SQN} = -g_k, \quad (8)$$

where  $A_k = J_k^T J_k$ , and  $B_k$  is an approximation to the second-order part of the Hessian matrix. The matrix  $B_k$  is updated using a suitable quasi-Newton update such that

$$B_{k+1}s_k = y_k - A_{k+1}s_k \quad (9)$$

or

$$B_{k+1}s_k = (J_{k+1} - J_k)^T F_{k+1}, \quad (10)$$

where  $s_k = x_{k+1} - x_k$  and  $y_k = g_{k+1} - g_k$ .

In [8], Brown and Dennis developed an algorithm for nonlinear least-squares curve fitting. In [13] Dennis presented the structured quasi-Newton method as a technique for solving large residual NLS problems. Although SQN methods are shown to be very efficient (especially for large residual problems), there are some aspects associated with such approaches that require special attention. Firstly, a suitable quasi-Newton update has to be found, together with secant equations, such as (9) or (10), listed above. Secondly, there is a demand for algorithms that automatically switch to the Gauss-Newton method when solving small residual problems, or that adopt a quasi-Newton update when solving large residual problems.

**2.1. Sizing Techniques on SQN.** Sizing techniques were introduced as a remedy for the difficulties involved when solving small (or in particular zero) residual problems using SQN methods. This is because, in general, quasi-Newton updates do not approach zero matrices; therefore, SQN methods do not perform as well as the GN and LM methods do for zero residual problems at the solution (i.e. when  $F(x^*) = 0$ ).

In [3], Bartholomew-Biggs considered some SQN methods based on the observation of McKeown [37] about the incapability of GN-like algorithms on minimizing some class of functions that are sums of squares. Bartholomew-Biggs was aware of the shortcomings of the SQN methods when solving small residual problems, so he proposed a sizing parameter

$$\zeta_k = \frac{F_{k+1}^T F_k}{F_k^T F_k} \quad (11)$$

on the matrix  $B_k$ , and updated it using the symmetric rank one (SR1) formula

$$B_{k+1} = \zeta_k B_k + \frac{(y_k^* - \zeta_k B_k s_k)(y_k^* - \zeta_k B_k s_k)^T}{(y_k^* - \zeta_k B_k s_k)^T s_k}, \quad (12)$$

where

$$y_k^* = (J_{k+1} - J_k)^T F_{k+1}. \quad (13)$$

Dennis et al. [15] proposed another successful SQN algorithm called NL2SOL. The NL2SOL maintains the secant approximation of the second-order part of the Hessian and automatically switches to such an approximation when solving large residual problems. The sizing parameter used by the NL2SOL algorithm is given by

$$\hat{\zeta} = \min \left\{ \left| \frac{s_k^T y_k^*}{s_k^T B_k s_k} \right|, 1 \right\}, \quad (14)$$

where the scalar  $\frac{s_k^T y_k^*}{s_k^T B_k s_k}$  corresponds to the one used by Oren and Luenberger [46] and Oren [45] in their popular self-sizing quasi-Newton algorithms. The approximated second-order part of the Hessian matrix in the NL2SOL algorithm was updated using

$$B_{k+1} = \hat{\zeta}_k B_k + \frac{(y_k^* - \hat{\zeta}_k B_k s_k) y_k^T + y_k (y_k^* - \hat{\zeta}_k B_k s_k)^T}{s_k^T y_k} - \frac{s_k^T (y_k^* - \hat{\zeta}_k B_k s_k)}{(s_k^T y_k)^2} y_k y_k^T.$$

The SQN method with sizing parameter (14) yield an adjustable algorithm because, if the residual turns to  $F_{k+1} = 0$ , then  $y_k^* = 0$  and  $\hat{\zeta}_k = 0$ . As a result, the updated matrix  $B_{k+1}$  will be the zero matrix.

In [26], Huschens argued that there are no theoretical results backing the commonly used sizing techniques discussed above. He proposed a self-sizing strategy on the SQN approach (and named it a *totally structured secant method*) that converges locally and superlinearly for non-zero residual problems and converges locally and quadratically for zero residual problems. Specifically, he dealt with the convex class of the Broyden's family.

Instead of obtaining the SQN direction using equation (8), Huschens proposed an alternative way to compute the direction using

$$(A_k + \|F_k\| B_k) d_k^{SQN} = -g_k, \quad (15)$$

where the matrix  $B_k = \sum_{i=1}^m \frac{F_i(x)}{\|F_i(x)\|} \nabla^2 F_i(x)$  is the  $k$ -th approximation of the second-order part of the Hessian. The matrix  $B_k$  is updated so that the secant equation

$$B_{k+1} s_k = \hat{y}_k \quad (16)$$

is satisfied, where  $\hat{y}_k = A_{k+1} s_k + \frac{\|F_{k+1}\|}{\|F_k\|} y_k^*$  and  $y_k^*$  is given in Equation (13).

Another version of the convergence proof by Huschens was given by Yabe and Ogasawara in [63], where they extend the result from a restricted convex class to a wider class of the structured Broyden's family.

In [73], Zhang et al. take advantage of the good convergence properties of the *totally structured secant method* of Huschens and proposed a modified quasi-Newton equation for nonlinear least-squares problems. Based on the numerical experiments presented in the paper, the performance of the Zhang et al. modified secant equation was shown to be better than the well-known structured secant equations (9) and (10).

In summary, the introduction of SQN methods reduced some of the drawbacks concerning the minimization of large residual problems. These approaches contributed to choosing the best secant equation from (9) and (10) for the quasi-Newton update (see [15]). Sizing techniques on SQN provide an environment for developing switching algorithms that make use of the good performance of the Gauss-Newton method when dealing with zero residual problems.

**2.2. Hybridization Methods for Solving Nonlinear Least Squares.** Hybrid methods for minimizing nonlinear least-squares problems take advantage of two approaches in just one algorithm. In [37], McKeown illustrates several examples of the quasi-Newton methods that do not take into consideration the special form of the Hessian matrix (3), but even though, perform much better than GN and LM methods.

Betts in [5] developed a hybrid algorithm for solving nonlinear least-squares problems. Betts algorithm considered the estimate for the Hessian matrix as the sum of the first and second order term, where the first order term is the usual Gauss-Newton method and the second-order term is obtained using a rank one quasi-Newton updating formula.

Based on the arguments of McKeown, Nazareth [42] reviewed some of the methods that implicitly or explicitly considered approximating the second order term of the Hessian matrix. In [41], Nazareth presented his contribution for minimizing a sum of squares of nonlinear functions. He considered a convex combination of the first order term and a matrix obtained by adopting a quasi-Newton approximation on the second-order term of the Hessian. The Hessian approximation proposed by Nazareth is given by

$$H_k = \alpha_k A_k + (1 - \alpha_k) B_k + \lambda_k I, \quad (17)$$

where  $\alpha_k \in [0, 1]$ ,  $\lambda_k > 0$  is a regularization parameter, and  $I$  is the identity matrix in  $\mathbb{R}^{n \times n}$ . The search direction is obtained by solving

$$H_k d_k = -g_k. \quad (18)$$

It can be observed that if  $\alpha_k = 1$ , the resulting direction corresponds to the Levenberg-Marquardt direction, while if  $\alpha_k = 0$ , we have the regularized quasi-Newton direction. Thus, the direction obtained by solving (18) contains the classical Levenberg-Marquardt, and the regularized quasi-Newton directions as special cases.

Al-Baali and Fletcher [2] considered various possibilities for hybridizing the Gauss-Newton method with quasi-Newton updates for nonlinear least squares. They used the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update in approximating the second-order term of the Hessian matrix. The authors have also used a switching criterion that measures the difference between the current Hessian and any positive definite matrix. Depending on the size of this measure, their algorithms switch to either the GN or the BFGS direction. The matrix updating is defined by

$$B_{k+1} = \begin{cases} B_k + \frac{y_k^* y_k^{*T}}{s_k^T y_k^*} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, & \text{if } \Delta(B_k) < \Delta(A_k) \\ A_{k+1}, & \text{otherwise,} \end{cases}$$

where  $\Delta(\cdot)$  is the Hessian error measure.

In [16], Fletcher and Xu pointed out that the rate of convergence of the algorithm proposed in [2] is not superlinear as claimed. They argued that the measure  $\Delta(\cdot)$  does not, in general, distinguish between zero and non-zero residual problems, they

also constructed a counter-example to support their observations (details can be found in [16]). In view of this, Fletcher and Xu proposed a general and most efficient algorithm with a switching strategy that uses the condition

$$(f_k - f_{k+1}) \leq \epsilon f_k, \quad (19)$$

where  $0 < \epsilon < 1$ . If (19) is satisfied then the algorithm switches to SQN direction, otherwise the GN direction is used by the algorithm. Therefore, the algorithm automatically adopts the GN direction for zero residual problems, and the SQN direction for non-zero residual problems.

Many studies have been made to the development of hybrid SQN methods (see, for example, [1, 35, 38, 57]). Spedicato and Vespucci [54] presented an extensive set of numerical experiments concerning 24 hybrid algorithms using the test problems introduced by McKeown [37]. The results show that most of the hybrid SQN methods performed well especially for large residual problems, with better results than those of the specialized GN method.

**2.3. Factorized Versions of Structured Quasi-Newton Methods.** One of the issues associated with SQN methods is the need of the approximated Hessian matrix to maintain its positive definiteness so that a line search approach can be used to achieve global convergence. In an attempt to overcome this difficulty, Yabe and Takahashi [64–66], and Sheng and Zou [53], independently, proposed a factorized version of SQN methods which maintains the positive definiteness of the Hessian matrix in the system  $H_k d_k = -g_k$ . In the factorized version of SQN, the search direction is given by

$$d_k = -[(L_k + J_k)^T(L_k + J_k)]^{-1}g_k, \quad (20)$$

where the matrix  $L_k$  (called  $L$ -update) is the same size as the Jacobian matrix  $J_k$ , acting as its correction and expecting to generate a descent direction.

In [65], Yabe and Takahashi studied the modified version of Sheng and Zou factorized SQN method (called SZ-update), creating a generalized update which encompasses the SZ-update as a special case. By applying sizing techniques, Yabe [62], further proposed the factorized version of the structured Broyden-like family. In [67], Yabe and Yamaki established and proved the local and superlinear convergence of the proposed methods.

In [72], Zhang et al. considered sized factorized secant methods. Interestingly, their methods ensure quadratic convergence for zero residual problems, and super-linear convergence for nonzero residual problems. Instead of computing the search direction using (20), they consider

$$d_k = -[(\|F_k\|L_k + J_k)^T(\|F_k\|L_k + J_k)]^{-1}g_k, \quad (21)$$

where the factor  $\|F_k\|$  acts as the sizing factor of the matrix  $L_k$ , in which  $\|F_k\|L_k$  approaches zero for zero residual problems, even if the  $L_k$  update does not approach the zero matrix.

Zhou and Chen [75] proposed a global hybrid Gauss-Newton structured BFGS method that automatically reduces to Gauss-Newton step for zero residual problems, and to the structured BFGS step for nonzero residual problems. Unlike Fletcher and Xu, Zhou and Chen proposed a switching strategy for which

$$\frac{z_k^T s_k}{\|s_k\|^2} \geq \epsilon, \quad (22)$$

where  $0 < \epsilon \leq \frac{1}{2} \lambda_{\min}(H(x^*))$  is a constant and  $\lambda_{\min}(H(x^*))$  is the smallest eigenvalue of the Hessian  $H$  at the minimizer  $x^*$ .  $z_k$  is given by

$$z_k = y_k^* \frac{\|F_{k+1}\|}{\|F_k\|}.$$

Condition (22) plays a role similar to (19); in addition, it can be used to ensure that the update matrix in the BFGS step is safely positive definite.

Wang et al. [58] exploited the idea of the modified BFGS method for non-convex minimization proposed by Li and Fukushima [32], and proposed a hybrid Gauss-Newton structured modified BFGS method for solving nonlinear least-squares problems.

The hybrid Gauss-Newton structured quasi-Newton algorithms show compelling and improved numerical performance compared to the classical algorithm for solving nonlinear least squares problems. However, their performance on large-scale problems is very poor because they require a considerable large amount of storage. Based on this drawback, matrix-free approaches are preferable for solving large-scale nonlinear least-squares problems (cf. [39]).

### 3. Derivative-free Methods for Minimizing Sum of Squares of Functions.

Besides Gauss-Newton, Levenberg-Marquardt and SQN methods, other strategies based on the derivative-free approach for solving nonlinear least squares problems exist. These methods are intentionally developed to avoid computing the first derivative of the function when solving (1). To the best of our knowledge, the first derivative-free algorithm for nonlinear least-squares problems was devised in 1965 by Powell [49]. Powell's algorithm is similar to the Gauss-Newton step, except for the fact that the former approximates the first derivative along a set of  $n$  linearly independent directions rather than along the coordinate axis.

Peckham [47] also proposed a method for minimizing a function, given by the sum of squares, without computing the gradient. Peckham considered the different approach of approximating the first and second order terms of (3). In [7], Brown and Dennis proposed two derivative-free algorithms for minimizing sums of the squares of nonlinear functions. These methods are simply the analogs of the Gauss-Newton and Levenberg-Marquardt methods where the Jacobian matrix is approximated using finite difference techniques. The direction of these methods is given by

$$d_k = -(D_k^T D_k + \mu_k I)^{-1} D_k F_k, \quad (23)$$

where the matrix  $D_k$  is componentwise computed by finite differences as

$$D_k^{ij} = \frac{F_i(x_k + h_k^j u_j) - F_i(x_k)}{h_k^j}$$

with  $u_j$  the  $j$ -th unit column vector and  $\mu_k$  and  $h_k^j$  given by

$$\mu_k = c \|F_k\|,$$

$$\text{with } c = \begin{cases} 10, & \text{if } 10 \leq \|F_k\| \\ 1, & \text{if } 1 < \|F_k\| < 10 \\ 0.01, & \text{if } \|F_k\| \leq 1, \end{cases}$$

$$\text{and } h_k^j = \min\{\|F_k\|, \delta_k^j\}.$$

Notice that  $h_k$  and  $\delta_k$  are  $n$  dimensional vectors, the index  $j$  denotes the  $j$ -th component of these vectors, and

$$\delta_k^j = \begin{cases} 10^{-9}, & |x_k^j| < 10^{-6} \\ 0.001|x_k^j|, & \text{otherwise.} \end{cases} \quad (24)$$

Ralston and Jennrich [50] developed another derivative-free algorithm for minimizing sums of squares of a nonlinear function called DUD, for “don’t use derivatives”. The algorithm is designed specifically for data fitting applications in which the computations of the derivatives are too expensive. Ralston and Jennrich showed that their approach is competitive with the best existing derivative-free algorithms at that time, like Powell’s [49] and Peckham’s [47].

A trial to develop a derivative-free hybrid SQN method for nonlinear least-squares problems was given by Xu in [59]. Xu proposed a modification to the hybrid Gauss-Newton BFGS method in such a way that the Broyden’s rank-one update is implemented for the algorithm to perform the Gauss-Newton step, whereas for the algorithm to take the BFGS step, the Jacobian matrix associated with the gradient is estimated using the finite difference operator. A good numerical behavior was observed from the algorithm, possibly because it combines the best features of the GN-Broyden and the finite-difference BFGS methods.

Working from a different perspective, but also as an attempt to develop a derivative-free algorithm for solving nonlinear least-squares problems, Shakhno and Gnatyshyn [52] approximated the Jacobian matrix in the Gauss-Newton step using a divided differences operator. Following the ideas of Shakhno and Gnatyshyn, Ren et al. [51] improved the local convergence result determining its radius of convergence.

Zhang et al. [70] and Zhang and Conn [69] developed a class of derivative-free algorithms for least-squares minimization. By taking advantage of the structure of the problem, they built polynomial interpolation models for each of the residual functions. The global convergence result of the proposed technique was obtained using the trust-region framework. It is worth mentioning that this new framework of derivative-free algorithms for solving nonlinear least-squares problems shows an excellent numerical performance compared with the algorithms that approximate the gradients using finite differences (cf. [69, 70]).

Recently, by taking advantage of the development of Automatic Differentiation (AD) technology [23], Xu et al. [60] proposed an efficient quasi-Newton approach for solving nonlinear least-squares problems. The method uses the application of AD to directly and efficiently compute the gradient of the objective function without computing the explicit Jacobian matrix, which is sometimes expensive. The Jacobian matrix appearing in the first order term of the Hessian matrix is approximated using the Broyden matrix  $B_k$  [9] updated in each iteration by the Broyden rank-one formula

$$B_{k+1} = B_k + \frac{(v_k - B_k s_k) s_k^T}{s_k^T s_k}, \quad (25)$$

such that  $B_{k+1} s_k = v_k$ , where  $v_k = F_{k+1} - F_k$ .

A fully Jacobian-free method for minimizing nonlinear least squares was recently proposed by Xu et al. [61]. This new approach was realized by the novel combination of preconditioning and AD techniques. In comparison to the previously discussed approaches, this new approach does not depend on the structure of the gradient

and the Hessian matrix of the objective function, making it suitable for general large-scale unconstrained optimization problems.

**4. Some Miscellaneous Methods for Nonlinear Least Squares.** Despite the fact that our survey is mainly on SQN methods, there exist some interesting methods in the literature that are worth mentioning. We list some of them here.

Gould et al. [20] introduced an algorithm for nonlinear least-squares problems using the combination of filter and trust region techniques. Bellavia et al. [4] presented a regularized Euclidean residual algorithm for nonlinear least squares and nonlinear equations. They proved the global and quadratic convergence of the algorithm without the non-singularity assumption of the Jacobian.

Although the structured nonlinear conjugate gradient methods are hardly studied, Kobayashi et al. [30] exploits the special structure of the Hessian of the nonlinear least-squares objective function and introduced some conjugate-gradient methods with structured secant condition for nonlinear least-squares problems. For box-constrained nonlinear least-squares problems, an inexact Gauss-Newton trust-region method was proposed by Porcelli in [48].

In another attempt, Gonçalves and Santos [19] proposed a simple spectral correction for the GN method, by approximating the residual Hessian in the second-order part of (3) using a sign-free spectral parameter. They showed that their approach is locally convergent when applied to the class of quadratic residual problems. In addition, convergence results were given for problems without ensured convergence of the GN method. Subsequently, Gonçalves and Santos [18] extended the idea, providing global convergence results to the case of non-quadratic residuals, by adopting the nonmonotone line search strategy of Zhang and Hager [71] as the globalization tool. Specifically, the direction of the approach of Gonçalves and Santos is obtained by solving

$$(J_k^T J_k + \mu_k I) d_k = -g_k, \quad (26)$$

where the parameter  $\mu_k$  is the spectral coefficient given by

$$\mu_k = \frac{s_k^T (J_{k+1} - J_k)^T F_{k+1}}{s_k^T s_k}. \quad (27)$$

In [27], Karas et al. have devised explicit algebraic rules for computing the regularization parameter of the LM method, with global and local convergence analysis. Resting upon the regularizing aspect of the LM method, further analysis of the performance of methods with second- and higher-order regularization terms have been developed (see, e.g. [6, 10, 11, 21, 43, 74]).

**5. Conclusions.** We have briefly reviewed methods for solving nonlinear least-squares problems, paying special attention to the structured quasi-Newton methods. The interested reader may refer to Ya-Xiang Yuan [68] review paper for some recent numerical methods based on Levenberg-Marquardt type, quasi-Newton's type, and trust-region algorithm.

Due to the nature of these problems, little attention was given to developing derivative-free methods for finding their solution. Based on this observation, we recommend further research on derivative-free methods for large-scale nonlinear least-squares problems.

Finally, we hope this small review will help students and researchers familiarize themselves with the required literature needed to understand classical and modern methods for solving nonlinear least-squares problems.

We conclude this brief survey stating some open (*based on this survey and to the best of our knowledge*) research topics:

1. It seems that hybrid GN-SQN algorithms perform better than the individual GN or SQN algorithm in practice. Is it possible to develop a single algorithm that solves nonlinear least-squares problems with better efficiency than the hybrid algorithm?
2. The hybrid algorithm proposed by Nazareth [41], adaptively solves nonlinear least-squares problems. We recommend further research on choosing a best convex combination parameter and, at the same time developing the convergence analysis of the algorithm.
3. Considering the good convergence property of the class of SQN algorithms, more research is needed to develop a simplified version of factorized SQN algorithms.
4. Due to the advancement of automatic differentiation (AD), the efficiency of the derivative-free hybrid method presented by Xu in [59] may likely improve by incorporating AD techniques.
5. Huschens-like methods for nonlinear least squares exhibit good local convergence properties. Does this property hold globally?
6. For large-scale problems, assembling a complete Hessian matrix may cause insufficient memory for storage. Keeping this in mind, is it possible to develop an algorithm that does not require the space for storing a Hessian matrix for minimizing the sums of squares of functions, considering its structure?

**Acknowledgements.** We would like to thank three anonymous referees for their valuable comments and for recommending additional references that enhance the quality of this review paper. We also wish to thank Hasan Shahid, of Brown University, USA, for his careful reading and helpful comments.

#### REFERENCES

- [1] M. Al-Baali, [Quasi-newton algorithms for large-scale nonlinear least-squares](#), in *High Performance Algorithms and Software for Nonlinear Optimization*, Springer, (2003), 1–21.
- [2] M. Al-Baali and R. Fletcher, Variational methods for non-linear least-squares, *J. Oper. Res. Soc.*, **36** (1985), 405–421.
- [3] M. C. Bartholomew-Biggs, [The estimation of the Hessian matrix in nonlinear least squares problems with non-zero residuals](#), *Math. Program.*, **12** (1977), 67–80.
- [4] S. Bellavia, C. Cartis, N. I. M. Gould, B. Morini and P. L. Toint, [Convergence of a regularized Euclidean residual algorithm for nonlinear least-squares](#), *SIAM J. Numer. Anal.*, **48** (2010), 1–29.
- [5] J. T. Betts, [Solving the nonlinear least square problem: Application of a general method](#), *J. Optim. Theory Appl.*, **18** (1976), 469–483.
- [6] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos and P. L. Toint, [Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models](#), *Math. Program.*, **163** (2017), 359–368.
- [7] K. M. Brown and J. E. Dennis, [Derivative-free analogues of the Levenberg-Marquardt and Gauss algorithms for nonlinear least squares approximation](#), *Numer. Math.*, **18** (1971), 289–297.
- [8] K. M. Brown and J. E. Dennis, A new algorithm for nonlinear least-squares curve fitting, in *Mathematical Software* (ed. J. R. Rice), Academic Press, New York, (1971), 391–396.
- [9] C. G. Broyden, [A class of methods for solving nonlinear simultaneous equations](#), *Math. Comput.*, **19** (1965), 577–593.
- [10] C. Cartis, N. I. M. Gould and P. L. Toint, [On the evaluation complexity of cubic regularization methods for potentially rank-deficient nonlinear least-squares problems and its relevance to constrained nonlinear optimization](#), *SIAM J. Optim.*, **23** (2013), 1553–1574.

- [11] C. Cartis, N. I. M. Gould and P. L. Toint, [On the evaluation complexity of constrained nonlinear least-squares and general constrained nonlinear optimization using second-order methods](#), *SIAM J. Numer. Anal.*, **53** (2015), 836–851.
- [12] A. Cornelio, [Regularized nonlinear least squares methods for hit position reconstruction in small gamma cameras](#), *Appl. Math. Comput.*, **217** (2011), 5589–5595.
- [13] J. E. Dennis, Some computational techniques for the nonlinear least squares problem, in *Numerical Solution of Systems of Nonlinear Algebraic Equations* (eds. G. D. Byrne and C. A. Hall), Academic Press, New York, (1973), 157–183.
- [14] J. E. Dennis and R. B. Schnabel Jr, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, 1983.
- [15] J. E. Dennis Jr, D. M. Gay and R. E. Walsh, [An adaptive nonlinear least-squares algorithm](#), *ACM Trans. Math. Software (TOMS)*, **7** (1981), 348–368.
- [16] R. Fletcher and C. Xu, [Hybrid methods for nonlinear least squares](#), *IMA J. Numer. Anal.*, **7** (1987), 371–389.
- [17] G. Golub and V. Pereyra, [Separable nonlinear least squares: the variable projection method and its applications](#), *Inverse Problems*, **19** (2003), R1–R26.
- [18] D. S. Gonçalves and S. A. Santos, [A globally convergent method for nonlinear least-squares problems based on the Gauss-Newton model with spectral correction](#), *Bull. Comput. Appl. Math.*, **4** (2016), 7–26.
- [19] D. S. Gonçalves and S. A. Santos, [Local analysis of a spectral correction for the Gauss-Newton model applied to quadratic residual problems](#), *Numer. Algorithms*, **73** (2016), 407–431.
- [20] N. Gould, S. Leyffer and P. L. Toint, [A multidimensional filter algorithm for nonlinear equations and nonlinear least-squares](#), *SIAM J. Optim.*, **15** (2004), 17–38.
- [21] G. N. Grapiglia, J. Yuan and Y. X. Yuan, [On the convergence and worst-case complexity of trust-region and regularization methods for unconstrained optimization](#), *Math. Program.*, **152** (2015), 491–520.
- [22] S. Gratton, A. S. Lawless and N. K. Nichols, [Approximate Gauss-Newton methods for nonlinear least squares problems](#), *SIAM J. Optim.*, **18** (2007), 106–132.
- [23] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, SIAM, Vol. 105, 2008.
- [24] H. O. Hartley, [The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares](#), *Technometrics*, **3** (1961), 269–280.
- [25] S. Henn, [A Levenberg-Marquardt scheme for nonlinear image registration](#), *BIT*, **43** (2003), 743–759.
- [26] J. Huschens, [On the use of product structure in secant methods for nonlinear least squares problems](#), *SIAM J. Optim.*, **4** (1994), 108–129.
- [27] E. W. Karas, S. A. Santos and B. F. Svaiter, [Algebraic rules for computing the regularization parameter of the Levenberg-Marquardt method](#), *Comput. Optim. Appl.*, **65** (2016), 723–751.
- [28] S. Kim, K. Koh, M. Lustig, S. Boyd and D. Gorinevsky, [An interior-point method for large-scale  \$\ell\_1\$ -regularized least squares](#), *IEEE. J. Sel. Top. Signal.*, **1** (2007), 606–617.
- [29] D. A. Knoll and D. E. Keyes, [Jacobian-free Newton–Krylov methods: a survey of approaches and applications](#), *J. Comput. Phys.*, **193** (2004), 357–397.
- [30] M. Kobayashi, Y. Narushima and H. Yabe, [Nonlinear conjugate gradient methods with structured secant condition for nonlinear least squares problems](#), *J. Comput. Appl. Math.*, **234** (2010), 375–397.
- [31] K. Levenberg, [A method for the solution of certain non-linear problems in least squares](#), *Quarter. Appl. Math.*, **2** (1944), 164–168.
- [32] D.-H. Li and M. Fukushima, [A modified BFGS method and its global convergence in non-convex minimization](#), *J. Comput. Appl. Math.*, **129** (2001), 15–35.
- [33] J. Li, F. Ding and G. Yang, [Maximum likelihood least squares identification method for input nonlinear finite impulse response moving average systems](#), *Math. Comput. Model.*, **55** (2012), 442–450.
- [34] D. C. López, T. Barz, S. Korkel and G. Wozny, [Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design](#), *Comput. Chem. Eng.*, **77** (2015), 24–42.
- [35] L. Lukšan, [Hybrid methods for large sparse nonlinear least squares](#), *J. Optim. Theory Appl.*, **89** (1996), 575–595.
- [36] D. W. Marquardt, [An algorithm for least-squares estimation of nonlinear parameters](#), *J. Soc. Ind. Appl. Math.*, **11** (1963), 431–441.

- [37] J. J. McKeown, [Specialised versus general-purpose algorithms for minimising functions that are sums of squared terms](#), *Math. Program.*, **9** (1975), 57–68.
- [38] F. Modarres, M. A. Hassan and W. J. Leong, [Structured symmetric rank-one method for unconstrained optimization](#), *Int. J. Comput. Math.*, **88** (2011), 2608–2617.
- [39] H. Mohammad and S. A. Santos, [A structured diagonal Hessian approximation method with evaluation complexity analysis for nonlinear least squares](#), *Computational and Applied Mathematics*, online.
- [40] D. D. Morrison, Methods for nonlinear least squares problems and convergence proofs, in *Proceedings of the Seminar on Tracking Programs and Orbit Determination* (eds. J. Lorell and F. Yagi), Jet Propulsion Laboratory, Pasadena, (1960), 1–9.
- [41] L. Nazareth, *An Adaptive Method for Minimizing a Sum of Squares of Nonlinear Functions*, NASA Report WP-83-99, 1983.
- [42] L. Nazareth, [Some recent approaches to solving large residual nonlinear least squares problems](#), *SIAM Review*, **22** (1980), 1–11.
- [43] Y. Nesterov, [Modified Gauss-Newton scheme with worst case guarantees for global performance](#), *Optim. Methods Softw.*, **22** (2007), 469–483.
- [44] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Science, 2006.
- [45] S. S. Oren, [On the selection of parameters in self scaling variable metric algorithms](#), *Math. Program.*, **7** (1974), 351–367.
- [46] S. S. Oren and D. G. Luenberger, [Self-scaling variable metric \(SSVM\) algorithms: Part I: Criteria and sufficient conditions for scaling a class of algorithms](#), *Management Science*, **20** (1974), 845–862.
- [47] G. Peckham, [A new method for minimising a sum of squares without calculating gradients](#), *The Comput. J.*, **13** (1970), 418–420.
- [48] M. Porcelli, [On the convergence of an inexact Gauss–Newton trust-region method for nonlinear least-squares problems with simple bounds](#), *Optim. Lett.*, **7** (2013), 447–465.
- [49] M. J. D. Powell, [A method for minimizing a sum of squares of non-linear functions without calculating derivatives](#), *The Comput. J.*, **7** (1965), 303–307.
- [50] M. L. Ralston and R. I. Jennrich, [DUD, A derivative-free algorithm for nonlinear least squares](#), *Technometrics*, **20** (1978), 7–14.
- [51] H. Ren, I. K. Argyros and S. Hilout, [A derivative-free iterative method for solving least squares problems](#), *Numer. Algorithms*, **58** (2011), 555–571.
- [52] S. M. Shakhno and O. P. Gnatyshyn, [On an iterative algorithm of order 1.839 ... for solving the nonlinear least squares problems](#), *Appl. Math. Comput.*, **161** (2005), 253–264.
- [53] S. Sheng and Z. Zou, *A New Secant Method for Nonlinear Least Squares Problems*, Report NANOG 1988-03, Nanjing University, China, 1988.
- [54] E. Spedicato and M. T. Vespucci, [Numerical experiments with variations of the Gauss-Newton algorithm for nonlinear least squares](#), *J. Optim. Theory Appl.*, **57** (1988), 323–339.
- [55] W. Sun and Y. X. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, Springer Science, 2006.
- [56] L. Tang, [A regularization homotopy iterative method for ill-posed nonlinear least squares problem and its application](#), in *Advances in Civil Engineering, ICCET 2011*, vol. 90 of Applied Mechanics and Materials, Trans Tech Publications, 2011, 3268–3273.
- [57] P. L. Toint, [On large scale nonlinear least squares calculations](#), *SIAM J. Sci. Stat. Comput.*, **8** (1987), 416–435.
- [58] F. Wang, D. H. Li and L. Qi, [Global convergence of Gauss-Newton-MBFGS method for solving the nonlinear least squares problem](#), *Advan. Model. Optim.*, **12** (2010), 1–19.
- [59] C. X. Xu, [Hybrid method for nonlinear least-square problems without calculating derivatives](#), *J. Optim. Theory Appl.*, **65** (1990), 555–574.
- [60] W. Xu, T. F. Coleman and G. Liu, [A secant method for nonlinear least-squares minimization](#), *Comput. Optim. Appl.*, **51** (2012), 159–173.
- [61] W. Xu, N. Zheng and K. Hayami, [Jacobian-free implicit inner-iteration preconditioner for nonlinear least squares problems](#), *J. Sci. Comput.*, **68** (2016), 1055–1081.
- [62] H. Yabe, [Variations of structured Broyden families for nonlinear least squares problems](#), *Optim. Methods Softw.*, **2** (1993), 107–144.
- [63] H. Yabe and H. Ogasawara, [Quadratic and superlinear convergence of the Huschen’s method for nonlinear least squares problems](#), *Comput. Optim. Appl.*, **10** (1998), 79–103.
- [64] H. Yabe and T. Takahashi, [Structured quasi-newton methods for nonlinear least squares problems](#), *TRU Math.*, **24** (1988), 195–209.

- [65] H. Yabe and T. Takahashi, [Factorized quasi-newton methods for nonlinear least squares problems](#), *Math. Program.*, **51** (1991), 75–100.
- [66] H. Yabe and T. Takahashi, [Numerical Comparison among structured quasi-Newton methods for nonlinear least squares problems](#), *J. Oper. Res. Soc. Japan*, **34** (1991), 287–305.
- [67] H. Yabe and N. Yamaki, [Convergence of a factorized Broyden-like family for nonlinear least squares problems](#), *SIAM J. Optim.*, **5** (1995), 770–791.
- [68] Y. X. Yuan, [Recent advances in numerical methods for nonlinear equations and nonlinear least squares](#), *Numer. Algebr., Contr. Optim.*, **1** (2011), 15–34.
- [69] H. Zhang and A. R. Conn, [On the local convergence of a derivative-free algorithm for least-squares minimization](#), *Comput. Optim. Appl.*, **51** (2012), 481–507.
- [70] H. Zhang, A. R. Conn and K. Scheinberg, [A derivative-free algorithm for least-squares minimization](#), *SIAM J. Optim.*, **20** (2010), 3555–3576.
- [71] H. Zhang and W. W. Hager, [A nonmonotone line search technique and its application to unconstrained optimization](#), *SIAM J. Optim.*, **14** (2004), 1043–1056.
- [72] J. Z. Zhang, L. H. Chen and N. Y. Deng, [A family of scaled factorized Broyden-like methods for nonlinear least squares problems](#), *SIAM J. Optim.*, **10** (2000), 1163–1179.
- [73] J. Zhang, Y. Xue and K. Zhang, [A structured secant method based on a new quasi-newton equation for nonlinear least squares problems](#), *BIT*, **43** (2003), 217–229.
- [74] R. Zhao and J. Fan, [Global complexity bound of the Levenberg-Marquardt method](#), *Optim. Methods Softw.*, **31** (2016), 805–814.
- [75] W. Zhou and X. Chen, [Global convergence of new hybrid Gauss-Newton structured BFGS method for nonlinear least squares problems](#), *SIAM J. Optim.*, **20** (2010), 2422–2441.

Received September 2017; 1<sup>st</sup> revision April 2018; final revision May 2018.

*E-mail address:* [hmuhd.mth@buk.edu.ng](mailto:hmuhd.mth@buk.edu.ng)

*E-mail address:* [mywaziri.mth@buk.edu.ng](mailto:mywaziri.mth@buk.edu.ng)

*E-mail address:* [sandra@ime.unicamp.br](mailto:sandra@ime.unicamp.br)