Some Iterative Regularized Methods for Highly Nonlinear Least Squares Problems*

I. Kangro and O. Vaarmann

Tallinn University of Technology, Institute of Cybernetics

Akadeemia tee 21, 12618 Tallinn, Estonia

E-mail(*coresp.*): `inga@cs.ioc.ee`

E-mail: `vaarmann@staff.ttu.ee`

Abstract. This report treats numerical methods for highly nonlinear least squares problems for which procedural and rounding errors are unavoidable, e.g. those arising in the development of various nonlinear system identification techniques based on input-output representation of the model such as training of artificial neural networks. Let F be a Frechet-differentiable operator acting between Hilbert spaces H_1 and H_2 and such that the range of its first derivative is not necessarily closed. For solving the equation $F(x) = 0$ or minimizing the functional $f(x) = \frac{1}{2} \|F(x)\|^2$, $x \in H_1$, two-parameter iterative regularization methods based on the Gauss-Newton method under certain condition on a test function and the required solution are developed, their computational aspects are discussed and a local convergence theorem is proved.

Key words: nonlinear least squares, parameter identification, artificial neural networks, iterative regularization, Gauss-Newton type methods.

1 Introduction

To translate a problem into a mathematical model is a typical task for operational and management science and scientific computing. Provided, that model formulation is successful there is a strong requirement for effective numerical techniques. Industrial and scientific problems lead frequently to complicated non-linear optimization problems that can be formulated as one of the mathematical programming:

$$\min \left\{ f(x) : \ x \in Q \right\}, \tag{1.1}$$

where Q is a closed subset of Hilbert space H_1. It also contains the problem of computing a solution of the equation

$$F(x) = 0 \tag{1.2}$$

---

with

$$f(x) = \frac{1}{2} \left\| F(x) \right\|^2, \quad Q = H_1, \tag{1.3}$$

where F is a nonlinear operator from a Hilbert space H_1 into another space H_2. For a differentiable operator F, depending on the properties of F', the problems under consideration may have a unique solution or infinitely many solutions or no solution in a classical sense. Real-life problems, which rest on mathematical modelling and simulation are frequently ill-posed. For instance, those arise in inverse problems because they typically involve the estimation of certain quantities based on indirect measurements. A well known example of great practical interest of an application of an inverse problem is that of computerized tomography (see e.g. [12]). The problem of the determination of the heat conductivity coefficient from available measurement data is also extensively studied (see e.g. [7, 13]). The important practical problem of determination of the heat conductivity coefficient of electrical cables is considered in [5, 4]. Since for ill-posed problems we cannot assume the existence of $(F')^{-1}$ or its uniform boundedness then some kind of regularization is needed. Ill-posedness is an important aspect of computation that is expressed in instability of the solution process (solution is unstable under data perturbations), thus the usage of non-regularized methods can produce uncontrollable and unacceptable error propagation.

Multilayer feed-forward network can easily cope with many complicated problems such as predicting fluctuations of stock market, energy planning, parameter identification etc., provided there are enough hidden layers in the network [10, 15, 18, 19]. For predicting fluctuations of stock market neural networks are useful and can serve as good financial advisors, in particular, as a tool to classify the alternatives into prescribed groups. The problem is then to find the "best weights" for the network, i.e., the weights that give the "best fit" for a given set of output vectors and given set of "target" output vectors. Similar problems arise in energy planning and in other fields of electrical engineering when we use neural networks. For short-time forecasting of the power load of a electric system a corresponding mathematical model can be given in the form of a three layered feed-forward neural network, where the activation (or transfer) function is, in general, a sigmoid function in respect of each predicted load [10]. Neural networks are intrinsically function approximators. In most applications activation function is a sigmoidal one, typically of the form $1/(1 - e^{-x})$, which leads to highly nonlinear least squares problems.

For solving those problems, regularizing algorithms containing several regularization parameters may be fruitful, because a proper choice of additional parameters enables to improve the convergence rate properties and to reduce the amount of computation.

Many problems in science and engineering can be formulated as an inverse problem. The field of inverse problems has experienced a rapid progress in the last few decades. Recently, many monographs, e.g., [1, 3, 9], and a big number of papers, e.g., [8, 16, 17, 26], are devoted to this topic. In the works mentioned above , as a rule, problems with the inaccurate data are considered and the convergence of the Gauss-Newton method is proved under the assumption on

source-like representation of the exact solution x^*:

$$x^* - u_0 = (F'^*(x^*)F'(x^*))^p \nu, \quad 0 < p \le 1.$$

In order to obtain simple estimates of the terms including approximations to the inverse operators, here we use the value $p = 1$. This paper is a continuation of papers [11, 14, 21, 22] and it treats approximate Gauss-Newton type methods for solving problems for which procedural and rounding errors are unavoidable, e.g, those arising in mathematical modelling and in the development of various system identification techniques based on input-output representation of the model. The purpose of this paper is to discuss basic concepts of approach and therefore the emphasis is on the theoretical background rather than the practical solution.

2 Methods and Convergence Theorem

A necessary condition for x to be a local minimum point of (1.3) is given by the equation

$$[F'(x)]^* F(x) = 0, \tag{2.1}$$

where $[\,\cdot\,]^*$ denotes the dual mapping.

If the operator F is highly nonlinear, then one possibility to handle nonlinear least squares (NLSQ) problems with small residual more safely is to use damped Gauss-Newton type methods. On the other hand, in the case of ill-conditioned F', acceptable values for the relaxation parameter ε can be extremely small what, in turn, means that the convergence speed may be drastically slowed down. The use of some iterative schemes which exploit the variable regularization parameter α and the variable relaxation parameter ε at each iteration step may be fruitful because that allows a wider choice of initial guesses x_0.

We shall consider the case when the range of F' is not necessarily closed and the product operator $[F']^*F'$ may not have a bounded inverse operator. The Gauss-Newton method

$$\begin{aligned} x_{k+1} &= x_k - [B(x_k)]^{-1} [F'(x_k)]^* F(x_k), \\ B(x_k) &= [F'(x_k)]^* F'(x_k), \quad k = 0, 1, \ldots, \end{aligned} \tag{2.2}$$

is equivalent to minimizing the linearized functional

$$g_k(h) = \frac{1}{2} \|F(x_k) + F'(x_k)h\|^2 \tag{2.3}$$

at every iteration point x_k for the correction term h. In order to cope with problems with the unbounded pseudoinverse of F' it is desirable to develop algorithms based on the Tikhonov functional

$$\Phi_k(h) = \frac{1}{2} \left(\|F(x_k) + F'(x_k)h\|^2 + \alpha \|\bar{x}_k - u_0\|^2 \right),$$

where $\alpha > 0$, $\bar{x}_k = x_k + h$, and u_0 is an element of H_1, the so-called test function.

Following the idea of iterative regularization [2, 23, 25], we suppose that α is a sequence $\{\alpha_k\}$ of properly chosen positive numbers. From the stationarity condition for the Tikhonov functional we obtain an one-parameter iterative Gauss-Newton method

$$x_{k+1} = x_k - M_k^{-1}\left[F'^*(x_k)F(x_k) + \alpha_k(x_k - u_0)\right], \tag{2.4}$$

where $M_k = B(x_k) + \alpha_k I$, I denotes the identity mapping, and $h_k = x_{k+1} - x_k$.

In this report we shall study a class of two-parameter regularized Gauss-Newton type methods

$$x_{k+1} = x_k - \varepsilon_k D_k\left[F'^*(x_k)F(x_k) + \alpha_k(x_k - u_0)\right], \tag{2.5}$$

where $0 < \varepsilon_k \le 1$ and D_k is an approximation to M_k^{-1} satisfying the condition

$$\|I - D_k\left(B\left(x_k\right) + \alpha_k I\right)\| \le \mu_k \le \mu_0 < 1, \quad k = 1, 2, \ldots.$$

The study of methods with approximate operators may give more realistic impression of the methods under discussion. Frequently, the use of finite-difference approximations to the derivatives gives rise to an inexact method. Approximation may be also regarded as a result of inevitable inaccuracy of computation or they are due to the strategy used for solving linear problems at each iteration, i.e. associated linear equations are solved intentionally approximately taking finitely many steps of an iterative procedure.

To get convergence results for the method (2.5), let us suppose that the operator F is twice Frechet-differentiable in the region under consideration with

$$\|F'(x)\| \le K_1, \quad \|F''(x)\| \le K_2, \quad K_1, K_2 > 0, \tag{2.6}$$

the equation (1.2) has a solution x^* with the representation

$$x^* - u_0 = F'^*(x^*)F'(x^*)\,v \tag{2.7}$$

holding for an element $v \in H_1$ and a test function u_0, and the inequality

$$\|F'^*(x^*)F'(x^*) - F'^*(x_k)F'(x_k)\| \le C_0\left\|x^* - x_k\right\|, \quad k = 1, 2, \ldots \tag{2.8}$$

is valid for some constant C_0.

We shall show that the sequence $\{x_k\}$ defined by (2.5) converges to a solution x^* of (1.2) provided $\|v\|$ and μ_k are sufficiently small, i.e.

$$\mu_k \le \mu_0 \le \|v\|, \quad k = 1, 2, \ldots, \tag{2.9}$$

$$2\varepsilon_0\left(K_1 K_2\|v\|\right)^{1/2}\left(1 + \mu_0\right) + \left[1 + \varepsilon_0(1 + \mu_0)C_0\right]\|v\| = q < 1, \tag{2.10}$$

and the parameter α_k is chosen as

$$\alpha_k = \frac{1}{2}\left[\frac{K_1 K_2}{\|v\|}\right]^{1/2} q^k \tau_0, \tag{2.11}$$

where $\tau_0 \ge \|x_0 - x^*\|$.

Theorem 1. *Let F be twice Frechet-differentiable operator and the relations* (2.6) – (2.11) *hold. If $1 - \varepsilon_k(1 - \mu_k) \leq \|v\|$ then the sequence $\{x_k\}$ defined by* (2.5) *converges to the solution x^* with*

$$\|x_k - x^*\| \leq q^k \tau_0, \quad k = 1, 2, \ldots$$

where $\|x_0 - x^*\| \leq \tau_0$.

Proof. Denote $\tau_k = q^k \tau_0$. Assume that $\|x_k - x^*\| \leq \tau_k$ for some $k \geq 0$ and show that then

$$\|x_{k+1} - x^*\| \leq \tau_{k+1}.$$

On the basis of the Taylor formula

$$F'^*(x_k) \left[F(x_k) - F(x^*) \right] = F'^*(x_k) F'(x_k)(x_k - x^*) + G,$$

where

$$\|G\| \leq \frac{K_1 K_2}{2} \|x_k - x^*\|^2. \tag{2.12}$$

Taking into account that $F(x^*) = 0$ and the relations (2.5)–(2.12) we have

$$
\begin{aligned}
x_{k+1} - x^* &= x_k - x^* - \varepsilon_k D_k \left[F'^*(x_k) F(x_k) + \alpha_k(x_k - u_0) \right] \\
&= x_k - x^* - \varepsilon_k D_k \Big\{ F'^*(x_k) \left[F(x_k) - F(x^*) \right] \\
&\qquad\qquad + \alpha_k(x_k - x^*) + \alpha_k(x^* - u_0) \Big\} \\
&= x_k - x^* - \left[M_k^{-1} + (\varepsilon_k - 1) M_k^{-1} + \varepsilon_k(D_k - M_k^{-1}) \right] \\
&\qquad\qquad \times \left[M_k(x_k - x^*) + G + \alpha_k(x^* - u_0) \right] \\
&= - \Big[(\varepsilon_k - 1)(x_k - x^*) + \varepsilon_k(D_k M_k - I)(x_k - x^*) \\
&\qquad\qquad + \varepsilon_k D_k G + \varepsilon_k \alpha_k D_k(x^* - u_0) \Big].
\end{aligned}
$$

On account of the relations (2.7) and (2.8) we obtain

$$
\begin{aligned}
x_{k+1} - x^* = \; &- \left[\varepsilon_k - 1 + \varepsilon_k(D_k M_k - I) \right] (x_k - x^*) \\
&- \varepsilon_k D_k M_k M_k^{-1} G - \varepsilon_k \alpha_k D_k M_k M_k^{-1} F'^*(x_k) F'(x_k) v \\
&- \varepsilon_k \alpha_k D_k M_k M_k^{-1} \left[F'^*(x^*) F'(x^*) - F'^*(x_k) F'(x_k) \right] v.
\end{aligned}
$$

Clearly,

$$M_k^{-1} F'^*(x_k) F'(x_k) v = M_k^{-1} \left[F'^*(x_k) F'(x_k) + \alpha_k I - \alpha_k I \right] v$$

$$= \left(I - \alpha_k M_k^{-1} \right) v$$

and due to the relation $\left\| M_k^{-1} \right\| \leq \frac{1}{\alpha_k}$ we get

$$\left\| M_k^{-1} F'^*(x_k) F'(x_k) v \right\| \leq 2 \|v\|.$$

Making use of the assumptions (2.8) and (2.11) and the inequalities

$$1 - \varepsilon_k(1 - \mu_k) \leq \|v\|, \quad \|D_k M_k\| = \|I + D_k M_k - I\| \leq 1 + \mu_k$$

we obtain

$$
\begin{aligned}
\|x_{k+1} - x^*\| \quad \leq \quad & \frac{\varepsilon_k K_1 K_2 (1 + \mu_k)}{2\alpha_k} \|x_k - x^*\|^2 + \|v\| \|x_k - x^*\| \\
& + 2\varepsilon_k \alpha_k (1 + \mu_k) \|v\| + \varepsilon_k (1 + \mu_k) C_0 \|x^* - x_k\| \|v\| \\
\leq \quad & \varepsilon_k \left[\frac{K_1 K_2 (1 + \mu_k)}{2\alpha_k} \tau_k^2 + 2\alpha_k (1 + \mu_k) \|v\| \right] \\
& + [1 + \varepsilon_k (1 + \mu_k) C_0] \|v\| \tau_k.
\end{aligned}
$$

Denoting $a_k = K_1 K_2 (1 + \mu_k)$ and $b_k = 1 + \mu_k$ we can rewrite α_k as

$$
\alpha_k = \frac{1}{2} \left[\frac{a_k}{b_k \|v\|} \right]^{1/2} \tau_k.
$$

Since

$$
\frac{a_k}{[a_k / b_k \|v\|]^{1/2}} + \frac{a_k^{1/2} b_k \|v\|}{(b_k \|v\|)^{1/2}} = 2 (a_k b_k \|v\|)^{1/2}
$$

then $\|x_{k+1} - x^*\| \leq q_k \tau_k$, where

$$
q_k = 2\varepsilon_k (K_1 K_2 \|v\|)^{1/2} (1 + \mu_k) + [1 + \varepsilon_k (1 + \mu_k) C_0] \|v\| .
$$

Obviously, if $\mu_k \leq \mu_0 < 1$ and $\varepsilon_k \leq \varepsilon_0 \leq 1$, then $q_k \leq q$, and hence

$$
\|x_{k+1} - x^*\| \leq q\tau_k = \tau_{k+1}.
$$

Thus, according to the principle of mathematical induction, the statement of the theorem is proved. □

3   Concluding Remarks

The most widely used neural network activation functions are logistic function and hyperbolic tangent. The main advantage of using these functions is that they are always differentiable and it is very easy and fast to calculate the derivatives of these functions. To get more realistic impression of convergence properties of the methods under discussion their approximate analogs are studied here. Frequently, the use of finite difference approximations to the derivatives gives rise to an inexact method. An approximate variant of the method can also be obtained as a result of a strategy used for solving linear problems at each iteration step.

Performance of methods of the type (2.5) is equivalent to either solving the associated linear equations or computing inverses with an error at every iteration step. The strategy to solve corresponding linear problems inexactly can be used for purpose of economy. A strategy of problem solving that instead of finding the exact solution of a linear equation at every step solves it intentionally inexactly permits to save the computational work and is adaptive in the sense that one uses low accuracy numerical solutions of linear equations when the solution of primary problem is not reached yet and improves the accuracy as the

solution is approached. In many cases iterative methods are more appropriate and economical for linear problems than direct ones. Besides, iterative methods are self-correcting, and hence they are not sensitive to computational errors. But their convergence can be quite slow in the presence of ill-conditioning. In order to improve the stability and to save thereby laborious solving of linear auxiliary problems it might be fruitful to do some extra computational effort and implement preconditioning techniques.

Sometimes discretization serves as a regularizer [24]. For the nonlinear equation

$$F(x) = y \qquad (3.1)$$

with the compact operator F, a suitable precedent discretization can be obtained by means of projection methods replacing the equation (3.1) by the equation $Q_h F P_h x = Q_h y$, where P_h and Q_h are ortoprojectors [20]. For solving (3.1) frequently Galerkin method is used which has a remarkable property: if the formally written Galerkin approximation converges, then the limit is necessarily an exact solution [6].

References

[1] A. B. Bakushinskij and M. Y. Kokurin. *Iterative Methods for Approximate Solution of Inverse Problems.* Springer, 2004.

[2] A. B. Bakushinskij and B. T. Polyak. On solving variational inequalities. *DAN SSSR*, **219**:1038–1041, 1974. (in Russian)

[3] J. Chen. The convergence analysis of inexact Gauss-Newton methods for nonlinear problems. *Computational Optimization and Applications*, **40**(1):97–118, 2008.

[4] R. Čiegis, A. Ilgevičius, G. Jankevičiūtė and M. Meilūnas. Determination of heat conductivity coefficient of a cable bundle by inverse problem solution method. *Electronics and Electronical Engineering*, **2**(90):77–80, 2009.

[5] Raim. Čiegis, Rem. Čiegis, M. Meilūnas, G. Jankevičiūtė and V. Starikovičius. Parallel numerical algorithms for optimization of electrical cables. *Math. Model. Anal.*, **13**(4):471–482, 2008. (doi:10.3846/1392-6292.2008.13.471-482)

[6] E. Domanskij. Sequential regularizability in the sense of Maslov for improperly posed problems. *Mathematical Notes*, **4**:1036–1042, 1994.

[7] S. Guseinov and A. Buikis. Inverse heat transport problems for coeffcients in two-layer domains and methods for their solution. *Math. Model. Anal.*, **7**(2):217–228, 2002.

[8] B. Kaltenbacher and A. Neubauer. Convergence of projected iterative regularization methods for nonlinear problems with smooth solutions. *Inverse problems*, **22**:1105–1119, 2006.

[9] B. Kaltenbacher, A. Neubauer and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems.* Walter de Gruyter, 2008.

[10] K. Kamkamidze and A. Nemsadze. Application of neural networks for short-term forecasting of electric system's load. In *Proc. of the Intern. Scientific Conference Problems of Control and Power Engineering, Tbilisi, Georgia, 2004*, pp. 252–256, Georgian Academy of Sciences, 2004. A. Eliashvili Institute of Control Systems. (in Russian)

[11] I. Kangro and O. Vaarmann. Two-parameter regularized Gauss-Newton type methods for highly nonlinear least squares problems. In L. Sakalauskas et al(Ed.), *The 20th Inter. Conference EurOPT'2008, Selected papers, Neringa, Lithuania, 2008*, Continuous Optimization and Knowledge-Based Technologies, pp. 160–164, Institute of Mathematics and Informatics, 2008. VGTU Publishing House "Technica".

[12] F. Natterer. *The Mathematics of Computerized Tomography.* SIAM, 2001.

[13] E. Pais and J. Janno. Inverse problem to determine degenerate memory kernels in heat flux with third kind boundary conditions. *Math. Model. Anal.*, **11**(4):427–450, 2006.

[14] I. Parts and O. Vaarmann. Damped Gauss-Newton type methods for ill-posed optimization problems. In H. Pranevicius et al(Ed.), *Intern. Conference on Operational Research SOBI2006, Tallinn, Estonia, 2006*, Simulation and Optimisation in Business and Industry, pp. 119–123, Kaunas University of Technology, 2006. Technologija, Kaunas.

[15] E. Petlenkov. *Neural Networks Based Identification and Control of Nonlinear Systems: ANARX Model Based Approach.* Tallinn University of Technology, Tallinn, 2007. (Ph.D thesis)

[16] J. Qi-Nian. On the iteratively regularized Gauss-Newton method for solving nonlinear ill-posed problems. *Math. Computation*, **69**(232):1603–1623, 2000.

[17] J. Qinian. A convergence analysis of the iteratively regularized Gauss-Newton method under the Lipschitz condition. *Inverse problems*, **24**:045002(16pp), 2008.

[18] M. Rudnicki, P. S. Czczepaniak and R. Cholajda. Genetic algorithm as a tool for solving electrical engineering problems. *K. Mietinen et al (Ed.), Evolutionary Algorithms in Engineering and Computer Science, John Wiley & Sons, Inc.*, pp. 261–281, 1999.

[19] B. Sipra. Application of optimization. The best of all possible worlds. *SIAM NEWS*, **30**(2), 1997.

[20] A. V. Smirnova. Discretization and iterative approximation of solutions for nonlinear ill-posed problems. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, **8**:73–79, 1992. (in Russian)

[21] O. Vaarmann. Solution of nonlinear least squares problems by Levenberg-Marquardt type methods. *Proceedings of the Estonian Academy of Sciences: Physics, Mathematics*, **38**(2):146–153, 1989.

[22] O. Vaarmann. On solving ill-conditioned systems of nonlinear equations. *Proceedings of the Estonian Academy of Sciences: Physics, Mathematics*, **43**(2):49–63, 1994.

[23] O. Vaarmann. Some approximate methods for ill-posed problems. *Proceedings of the Estonian Academy of Sciences: Physics, Mathematics*, **48**(2):146–151, 1999.

[24] G. Vainikko and U. Hämarik. Projection methods and self-regularization in ill-posed problems. *Soviet Mathematics*, **29**:1–20, 1985.

[25] V. Vasin. Iterative regularization techniques for ill-posed problems. *Russian Mathematics (Iz. VUZ)*, **11**:69–84, 1995.

[26] V. V. Vasin and A. A. Mokrushin. Iterative process of Gauss-Newton type for ill-posed operator equations. *Dokl. Akad. Nauk.*, **371**(1):35–37, 2000.