

# Successive Approximation of Nonlinear Confidence Regions (SANCR)

Thomas Carraro<sup>(✉)</sup> and Vladislav Olkhovskiy

Institute for Applied Mathematics, Heidelberg University,  
Im Neuenheimer Feld 205, 69120 Heidelberg, Germany  
thomas.carraro@iwr.uni-heidelberg.de

**Abstract.** In parameter estimation problems an important issue is the approximation of the confidence region of the estimated parameters. Especially for models based on differential equations, the needed computational costs require particular attention. For this reason, in many cases only linearized confidence regions are used. However, despite the low computational cost of the linearized confidence regions, their accuracy is often limited. To combine high accuracy and low computational costs, we have developed a method that uses only successive linearizations in the vicinity of an estimator. To accelerate the process, a principal axis decomposition of the covariance matrix of the parameters is employed. A numerical example illustrates the method.

**Keywords:** Parameter estimation · Nonlinear confidence region · Covariance matrix · Differential equations

## 1 Introduction

To simplify the notation, we consider a nonlinear model  $f(t, \theta)$ , with  $\theta \in \mathbb{R}^n$  and  $t \in \mathbb{R}$ , which does not depend on an additional (dynamical) system. We assume that  $f$  is differentiable with respect to  $\theta$  and continuous with respect to  $t$ .

We consider the approximation of a confidence region about parameter values estimated by nonlinear least squares. The parameters are estimated by using experimental data  $y_i$  in some given points  $t_i$  with  $i = 1, \dots, m$ . The observed values contain unknown errors  $e_i$  that we assume additive, so the response variable can be modeled by

$$y_i = f(t_i, \theta_{\text{true}}) + e_i, \quad (1)$$

where  $\theta_{\text{true}}$  is the unknown true value of the parameters. Therefore, the least squares estimator  $\hat{\theta}$  is the value that solves the following problem

$$\hat{\theta} = \operatorname{argmin} \frac{1}{2} S(\theta), \quad (2)$$

where  $S(\theta)$  is the residual sum of squares

$$S(\theta) = \sum_{i=1}^m (y_i - f(t_i, \theta))^2. \quad (3)$$

We assume that the model is correct and that the errors are normal, independent and identically distributed (iid) random variables with zero mean and variance  $\sigma^2$ , i.e.  $e_i \sim N(0, \sigma^2)$ .

The confidence regions are here interpreted (from the frequentistic perspective [14]) as the regions in the parameter space covering the true value of the parameters  $\theta_{\text{true}}$ , in large samples, with probability approximately  $1 - \alpha$ .

The use of linearized confidence regions with nonlinear algebraic models has been extensively treated in literature, see for example [1, 2, 6, 8, 11, 16]. In particular, it has been shown that confidence regions derived for the linear case can be used in linearized form also for nonlinear models, but in many cases with limited accuracy [18]. Furthermore, there are approximation techniques for nonlinear models that are not based on linearizations [3, 10, 17, 19].

To simplify the exposition, in this work we consider an algebraic model, but the method can be used for more complex models. In fact, the problem to approximate nonlinear confidence regions for implicit models, i.e. models based on a system of (differential) equations has been considered from different points of view and for different kind of applications by several authors. To cite only few of them, see the work [18] and the references therein for the design under uncertainty, [20] for an application to ground water flow, [13] for ecological systems, and [15] for additional examples. Newly, it has been presented a method based on second-order sensitivity for the approximation of nonlinear confidence regions applied to ODE based models [12]. It has been shown that higher order sensitivities give a higher accurate approximation of the confidence regions than methods using only the first order sensitivities.

With this work we show that the approximation using only linearized confidence regions can be substantially improved by a systematic successive application of linearizations, in the following called *Successive Approximation of Nonlinear Confidence Regions (SANCR)* method. We show results for the case with only two model parameters. An extension to more than two parameters is technically straightforward and could be partially parallelized, but the effect of successive linearizations in more than two (parameter space) dimensions has yet to be studied in this framework.

This paper is organized as follows (i) In Sect. 2 we report the two methods on which our approach is based; (ii) In Sect. 3 we describe the new method; (iii) In Sect. 4 we show a numerical realization of the SANCR method.

## 2 Linearized Confidence Region and Likelihood Ratio Test

As explained above, there are several methods to approximate (nonlinear) confidence regions. Our method is based on the following two approaches [19].

For a given estimator  $\hat{\theta}$  of the parameter  $\theta$ , we consider:

- (i) The method derived from the likelihood ratio test (LR)

$$-2 \log(L(\theta)/L(\hat{\theta})) \leq \gamma^2, \quad (4)$$

from which it follows

$$S(\theta) - S(\hat{\theta}) \leq \gamma^2. \quad (5)$$

where  $L$  is the likelihood function and  $\gamma^2$  is the confidence level.

- (ii) The method based on the Wald test that leads to the linearized confidence regions (CL):

$$(\theta - \hat{\theta})^T Cov^{-1}(\theta - \hat{\theta}) \leq \gamma^2, \quad (6)$$

where  $Cov$  is the estimated covariance matrix of the parameters. There are several approximations of  $Cov$  [18], we use the one based on the Jacobian  $J$  of  $f$ :

$$Cov = s^2(J^T J)^{-1}, \quad (7)$$

where

$$J_{i,j} = \frac{\partial f(t_i, \theta)}{\partial \theta_j}. \quad (8)$$

The level  $\gamma^2 = \chi_{1-\alpha, n}^2$  is given by the  $1 - \alpha$  percentile of the chi-square distribution with  $n$  degrees of freedom in case  $\sigma^2$  is known, and it is  $\gamma^2 = s^2 n F_{(1-\alpha, n, m-n)}$  in case  $\sigma^2$  is unknown, but approximated by  $s^2 = S(\hat{\theta})/(m-n)$ . It has been proved [7] that these two confidence regions are asymptotically equivalent, but far from the asymptotic behavior, i.e. in case of a small number of data, they perform differently as presented in [18]. Additionally, our method show the limitation of linearized confidence regions based only on (6).

One of the major goals in defining the confidence regions is the reduction of the costs associated to their computation. From the perspective of the computational costs, the method CL is cheap since it needs only one evaluation of the covariance matrix at the parameter value  $\hat{\theta}$ , while the method LR is much more expensive because it is based on the evaluation of the functional  $S$  in an adequately high number of points  $\theta$  in the vicinity of  $\hat{\theta}$  to produce a contour. In addition, the extension of the confidence region is not known a priori. In practice, the number of function evaluations needed for the method LR is in the order of several thousands, for example in our case with two parameters we use a grid of  $10^4$  points for the method LR.

On the contrary, as indicated in the expression (7), the covariance matrix can be evaluated at the cost of building the Jacobian  $J$ . Therefore, the major computational costs for the method CL are given by the computation of the derivatives of the model  $f$  with respect to the parameters. Thus, we have few computations of a linearized model for the method CL while many thousand

computations of a nonlinear model are needed for the method LR. Unfortunately, the accuracy of these two methods is inversely related to their computational costs, with the CL method being much more inaccurate if the model is highly nonlinear. We remind that both methods are only asymptotically exact for linear models and their quality decreases far from the asymptotic behavior.

Therefore, a compromise between computational costs and precision is highly required for many practical applications especially in case the model is based on differential equations. To this aim we established a new method combining low computational costs and high accuracy.

### 3 Successive Linearizations of Nonlinear Confidence Regions

The SANCR method is based on the use of successive linearizations of the confidence region, starting from the estimated parameter value  $\hat{\theta}$  (see expression (2)) combined with the likelihood ratio test (5) as explained below exemplarily for a model with two parameters.

The likelihood ratio test is used to check whether a point belongs or not to the approximate nonlinear confidence region. Instead of testing all points in the vicinity of  $\hat{\theta}$  we use an educated guess, i.e. the likelihood ratio test is performed only on few points lying on the contour of the *linearized* confidence regions. In fact, linearized confidence regions are ellipsoids in the parameter space and the directions of the semi-axis are defined by the eigenvectors of the covariance matrix as can be deduced by the quadratic form (6). Note that the covariance matrix has dimension  $n \times n$ , where  $n$  is the number of parameters to estimate. Therefore, starting from  $\hat{\theta}$  we determine the directions of the principal axes and their length which is given by

$$\ell_i = \gamma \sqrt{\lambda_i},$$

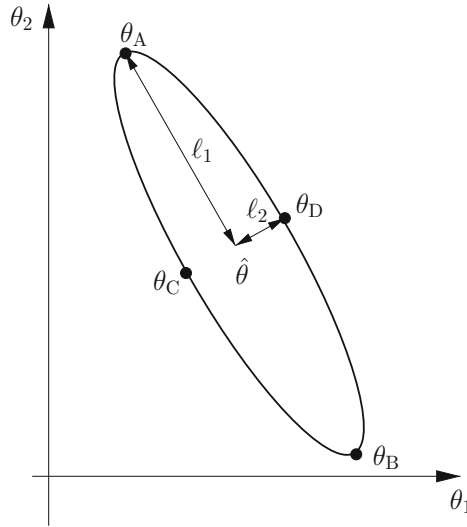
where  $\lambda_i$  is the eigenvalue corresponding to the  $i^{\text{th}}$  eigenvector. We perform the likelihood ratio test for the extreme points of the semi-axes, see points  $\theta_A, \theta_B, \theta_C, \theta_D$  in Fig. 1.

Let be  $\theta_A$  the first point to be processed. If this point passes the test, i.e. if the following condition is fulfilled

$$S(\theta_A) - S(\hat{\theta}) \leq \gamma^2,$$

it is considered for the construction of the confidence region and the procedure continues along the second axis. On the contrary, if the point  $\theta_A$  does not pass the test, it is discarded and a new candidate in the same direction  $\hat{\theta}\theta_A$  is chosen.

A new point  $\theta'_A$  along the selected semi-axis is taken by scaling  $\ell_1$  by a factor  $\alpha < 1$  as shown in Fig. 2(a). This procedure is repeated with a new likelihood ratio test and possibly a rescaling (reducing  $\alpha$ ) until a point that satisfies the test



**Fig. 1.** Definition of the points to perform the likelihood ratio test for two parameters.

$$S(\theta'_A) - S(\hat{\theta}) \leq \gamma^2$$

is found. Once this point, say  $\theta_{\text{new}}$ , has been found, we linearize the confidence region around this new point. To this aim we calculate the Jacobian  $J(\theta_{\text{new}})$  (see (8)) and the covariance  $Cov(\theta_{\text{new}})$  (see (7)).

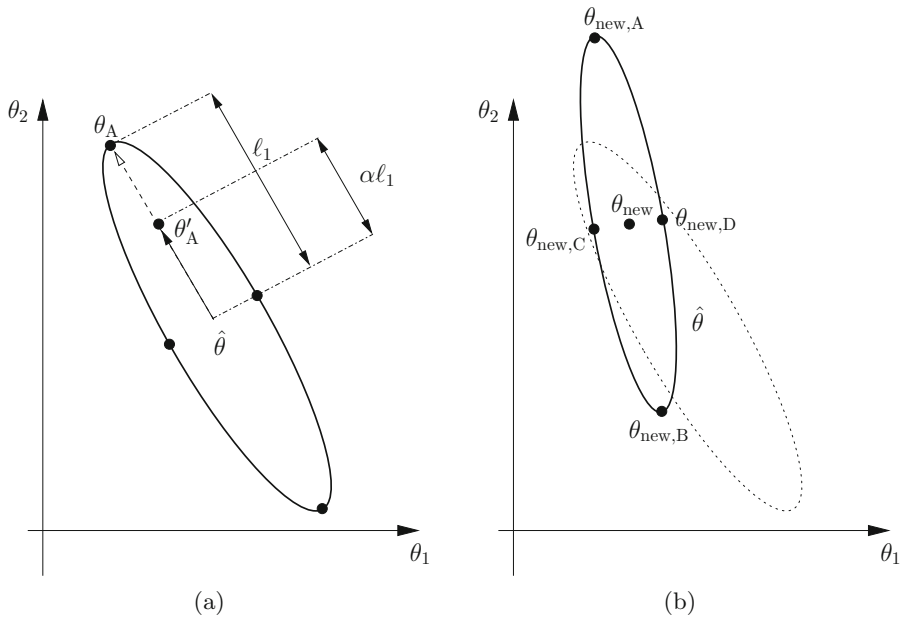
After performing the eigendecomposition of the new covariance matrix, the principal axes might have changed direction due to the nonlinearity of the model, see Fig. 2(b). Following the new principal directions, we can analogously find the next candidate points belonging to the confidence region, i.e. the points  $\theta_{\text{new,A}}, \theta_{\text{new,C}}$  and  $\theta_{\text{new,D}}$ , see Fig. 2(b). The point  $\theta_{\text{new,B}}$  is not considered because it is the opposite extremal point of the same principal axis. In fact, instead of taking  $\theta_{\text{new,B}}$ , we perform the same procedure starting from  $\theta_B$  to approximate the confidence region in the direction  $\hat{\theta}\theta_B$ . Therefore, this procedure is repeated along all principal axes considering both directions.

*Stopping Criterion.* The search along one principal axis is stopped if the distance of the next accepted point, let's say  $\theta'_{\text{new,A}}$ , to the previous one is less than a given tolerance

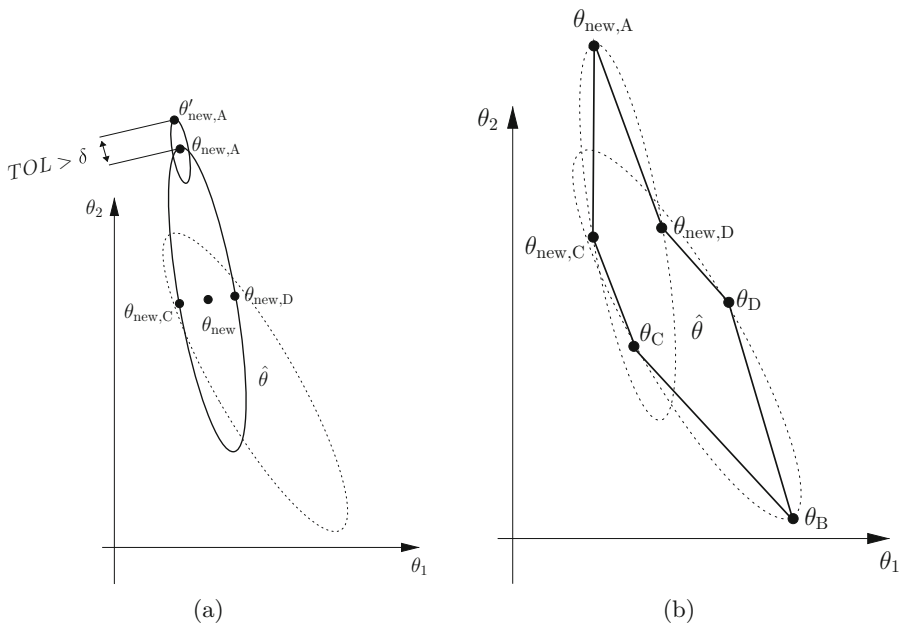
$$|\theta'_{\text{new,A}} - \theta_{\text{new,A}}| < TOL, \tag{9}$$

then the point  $\theta_{\text{new,A}}$  is retained to define the nonlinear confidence region, see Fig. 3.

*Contour Approximation.* The contour of the nonlinear confidence region is approximated by connecting all retained points, in our case  $\theta_{\text{new,A}}, \theta_{\text{new,C}}, \theta_{\text{new,D}}, \theta_C, \theta_D$  and  $\theta_B$ . These points are linearly connected as shown in Fig. 3(b).



**Fig. 2.** Scaling the semi-axes (a) and linearize at the new point (b).



**Fig. 3.** Stopping criterion (a) and interpolation (b).

## 4 Numerical Results

As an example the following model is considered

$$y = \theta_1 t^{\theta_2},$$

where the parameter  $\theta_1$  and  $\theta_2$  are estimated by the nonlinear least squares method. To simulate the parameter estimation process we have applied perturbed data generated using the “true” values of the parameters according to the following model response:

$$y_i = f(t_i, \theta_{\text{true}}) + e_i, \quad (11)$$

where  $e_i$  is a random variable distributed as  $N(0, \sigma^2)$ . The Table 1 indicates the values  $\theta_{\text{true}}$  and  $\sigma^2$  used in the calculations, and the least squares estimated values  $\hat{\theta}$  found by minimizing  $S(\theta)$  for a realization of the observations  $y_i$ . Additionally, the Table 2 includes the measurement positions  $t_i$ . One stopping criterion of the SANCR method is that the distance of two successive candidates is smaller than a given tolerance  $TOL$ , see (9). We have used  $TOL = 0.15$ .

To evaluate the results of our approach we compare it with a Markov Chain Monte Carlo (MCMC) method described in [9] using the associated MCMC toolbox for Matlab. In fact, an alternative way to perform a statistical analysis of nonlinear models is the use of the Bayes’s theorem [4]. Bayesian inference is not the focus of our work, therefore we refer for example to [5] for a presentation of the Bayesian approach. Since the MCMC method does not allow to easily define a stopping criterion to assure convergence, we have set to  $5 \cdot 10^6$  the number of model evaluations in the MCMC code.

In Fig. 4 the approximations of the confidence region using the four methods can be qualitatively compared. The blue dots (for the colors see the electronic version) are the points of the MCMC method. The cyan ellipse is the linearized confidence region of the method CL. The green curve is the confidence region approximated by the method LR and the red curve is the confidence region approximated by the SANCR method.

One can observe that the linearized confidence region CL is much smaller than the MCMC approximation and that it is not centered in it. The SANCR method is an approximation of the confidence region defined by the method LR obtained at a much lower computational cost than the method LR itself. The computational costs are reported in Tables 3 and 4. The method CL is very cheap with only one evaluation of the nonlinear model and the evaluations of the sensitivities with respect to the two parameters, but its quality is not satisfactory. The SANCR method uses 59 function evaluations and 42 ellipses. The latter correspond to 84 sensitivity evaluations according to the number of two parameters. The LR and the MCMC methods have been used here with  $10^4$ , respectively  $5 \cdot 10^6$ , model evaluations.

**Table 1.** Parameters and variance

| $\sigma^2$ | $\theta_{\text{true}}$ | $\hat{\theta}$   |
|------------|------------------------|------------------|
| 0.1        | (0.725, 4)             | (0.7279, 3.9974) |

**Table 2.** Position of measurement points

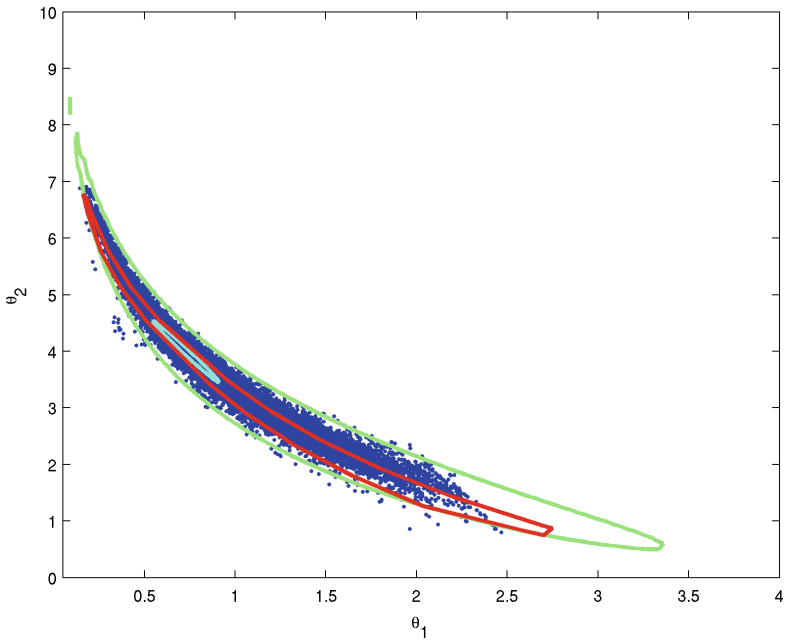
| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|-------|-------|-------|-------|-------|-------|
| 1.309 | 1.471 | 1.490 | 1.565 | 1.611 | 1.680 |

**Table 3.** Model evaluations of the four methods

| SANCR | CL | LR     | MCMC           |
|-------|----|--------|----------------|
| 59    | 1  | $10^4$ | $5 \cdot 10^6$ |

**Table 4.** Derivatives computations of the four methods

| SANCR | CL | LR | MCMC |
|-------|----|----|------|
| 84    | 2  | 0  | 0    |



**Fig. 4.** Confidence region approximated by the four methods.



## References

1. Bard, Y.: *Nonlinear Parameter Estimation*. Academic Press, a subsidiary of Harcourt Brace Jovanovich, Publishers, New York (1974)
2. Bates, D.M., Watts, D.G.: Relative curvature measures of nonlinearity. *J. Royal Stat. Soc. Ser. B* **42**, 1–25 (1980)
3. Bates, D.M., Watts, D.G.: *Nonlinear regression analysis and its applications*. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, Chichester (1988)
4. Box, G.E.P., Tiao, G.C.: *Bayesian Inference in Statistical Analysis*. Wiley, New York (1973)
5. Conrad, P., Marzouk, Y.M., Pillai, N., Smith, A.: Accelerating asymptotically exact MCMC for computationally intensive models via local approximations. *Journal of the American Statistical Association* (2015, in press)
6. Cook, R., Witmer, J.: A note on parameter-effects curvature. *J. Am. Stat. Assoc.* **80**, 872–878 (1985)
7. Cox, D., Hinkley, D.: *Theoretical Statistics*. Chapman and Hall, London (1974)
8. Gallant, A.R.: Confidence regions for the parameters of a nonlinear regression model. Mimeograph Series No. 1077, pp. 633–643. North Carolina University, Institute of Statistics (1976)
9. Haario, H., Laine, M., Mira, A., Saksman, E.: Dram: efficient adaptive MCMC. *Stat. Comput.* **16**(4), 339–354 (2006)
10. Hamilton, D.C., Watts, D.G., Bates, D.M.: Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. *Ann. Stat.* **10**(2), 386–393 (1982)
11. Jennrich, R.I.: Asymptotic properties of non-linear least squares estimators. *Ann. Math. Stat.* **40**(2), 633–643 (1969)
12. Kostina, E., Nattermann, M.: Second-order sensitivity analysis of parameter estimation problems. *Int. J. Uncertain. Quantif.* **5**(3), 209–231 (2015)
13. Marsili-Libelli, S., Guerrizio, S., Checchi, N.: Confidence regions of estimated parameters for ecological systems. *Ecol. Model.* **165**(23), 127–146 (2003)
14. Meeker, W.Q., Escobar, L.A.: Teaching about approximate confidence regions based on maximum likelihood estimation. *Am. Stat.* **49**(1), 48–53 (1995)
15. Oberkampf, W.L., Barone, M.F.: Measures of agreement between computation and experiment: validation metrics. *J. Comput. Phys.* **217**(1), 5–36 (2006)
16. Pázman, A.: *Nonlinear Statistical Models*. Kluwer Academic Publishers, Dordrecht (1994)
17. Potocký, R., To, V.B.: Confidence regions in nonlinear regression models. *Appl. Math.* **37**(1), 29–39 (1992)
18. Rooney, W.C., Biegler, L.T.: Design for model parameter uncertainty using nonlinear confidence regions. *AIChE J.* **47**(8), 1794–1804 (2001)
19. Seber, G., Wild, C.: *Nonlinear Regression*. Wiley, New York (1989)
20. Vugrin, K.W., Swiler, L.P., Roberts, R.M., Stucky-Mack, N.J., Sullivan, S.P.: Confidence region estimation techniques for nonlinear regression in groundwater flow: three case studies. *Water Resour. Res.* **43**(3) (2007). <http://onlinelibrary.wiley.com/doi/10.1029/2005WR004804/full>