Chapter 37
Parameter Estimation of Nonlinear Response Surface Models by Using Genetic Algorithm and Unscented Kalman Filter

Özlem Türkşen and Esin Köksal Babacan

Abstract Some of the real world problems are characterized by using nonlinear functions in the parameters. In this case, optimization of nonlinear response surface models become challenging with derivative-based optimization methods. In this study, two of the derivative free methods, Genetic Algorithm (GA) and Unscented Kalman Filter (UKF), are used for parameter estimation of complex nonlinear response surface model. A numerical example in chemical science is given to illustrate the performance of the methods.

37.1 Introduction

One of the main stage to solve a real world problem is development of an adequate functional relationship between a response of interest and a number of associated input variables. In general, such a relationship is unknown and may have uncertainty related to the structure of model parameters which characterize the model. The most common preferred basic modeling tool is regression analysis. The regression analysis investigates the response models which are linear in the parameters such as low degree polinomial models in many response surface modeling studies [2, 7]. Parameter estimation of these linear response models is achieved by using Ordinary Least Squares (OLS) method. However, many mathematical models used in scientific research contain parameters that are not expressed linearly. In this case, nonlinear regression models are preferred.

A regression model is called nonlinear in the parameters, if the derivatives of the model with respect to the model parameters depends on one or more parameters. The most commonly used assumptions for nonlinear regression is the same as assumptions for linear regression. The only exception being that the regression function is a nonlinear function of the unknown parameters instead of a linear function of the parameters. The nonlinear models have been used in many fields

Ö. Türkşen (✉) • E.K. Babacan
Faculty of Science, Statistics Department, Ankara University, Tandoğan, Ankara 06100, Turkey
e-mail: ozlemturksen@hotmail.com

Ş.Ş. Erçetin (ed.), *Chaos, Complexity and Leadership 2014*,
Springer Proceedings in Complexity, DOI 10.1007/978-3-319-18693-1_37

particularly in guidance or navigational systems, target tracking, biological and chemical sciences, economic and growth models. The models, which are nonlinear in parameters, can be made linear in parameters by suitable transformations. Even so, the transformations seem to destroy the model assumptions, e.g. assumption of variance homogeneity. A formal application of the OLS to the transformed model will not produce a model with correct statistical properties. According to this, the OLS method will be biased for parameter estimation of the transformed models. On the other hand, calculating derivatives with respect to the parameters will be resulted with normal equations which are nonlinear in the parameters. Therefore, derivative-free optimization methods will be very useful for obtaining parameter estimates of complex response problems.

The parameter estimation procedure for nonlinear response models is based on minimizing the quadratic function of difference between observed and predicted response values which is called error function. This error function is considered as objective function. There have been several derivative free methods for optimizing this complex objective function, e.g. Genetic Algorithm (GA), [4] Nelder-Mead simplex method [11], Simulated Annealing [8], Unscented Kalman Filter (UKF) [5]. In this study, two of the derivative free methods, GA and UKF, are preferred to achieve the parameter estimation procedure for nonlinear response models. In the next section, brief description about nonlinear response model is given. In Sect. 37.3, parameter estimation procedures of nonlinear response problems with GA and UKF are explained and algorithmic steps are presented. An application study is performed in Sect. 37.4. In Sect. 37.5, the conclusion is given with the obtained estimation results through performance metric.

37.2 Nonlinear Responses Model

A nonlinear response surface model is a continuous nonlinear multivariate approximation to real form of the response. Suppose an unknown response model given as the form below

$$Y_i = \eta\left(\mathbf{X}_i; \boldsymbol{\theta}\right) + \varepsilon_i, i = 1, 2, \ldots, n \tag{37.1}$$

in which Y_i is the response variable; $\mathbf{X}_i$ is a vector of input variables, $\mathbf{X}_i = [X_{1i}, X_{2i}, \ldots, X_{ki}]$; $\boldsymbol{\theta}$, is a vector of parameters, $\boldsymbol{\theta} = \left[\theta_1, \theta_2, \ldots, \theta_p\right]$; η is a nonlinear function in the parameters (at least partially nonlinear function of a η dimensional parameter vector $\boldsymbol{\theta}$) and ε_i is the error term, $i = 1, 2, \ldots, n$. In order to obtain parameter estimates of the model given in Eq. (37.1), the sum of squares for error is defined as

$$\phi\left(\boldsymbol{\theta}\right) = \sum_{i=1}^{n} \left[Y_i - \eta\left(\mathbf{X}_i; \boldsymbol{\theta}\right)\right]^2 = \sum_{i=1}^{n} \varepsilon_i^2 \tag{37.2}$$

which wanted to be minimized. The quadratic function given in Eq. (37.2) is considered as objective function for parameter estimation procedure. In order to minimize the Eq. (37.2) a well-known methodology is calculating derivatives with respect to parameter vector $\boldsymbol{\theta}$. However, the obtained normal equations can be nonlinear in the parameters. In this case, derivative-free optimization methods should be used to achieve the parameter estimates.

37.3 Parameter Estimation Procedure

In this section, two derivative-free optimization methods, GA and UKF, are explained briefly. The algorithmic steps of the methods are given in detail for parameter estimation of nonlinear response surface problems.

37.3.1 Parameter Estimation with Genetic Algorithm

Genetic Algorithm (GA) is a metaheuristic method based on natural selection and genetic mechanism. The basic principle of it is the Darwinian "survival of the fittest approach", introduced by Holland [4]. GAs search from a population points, not a single point; use objective function information, not derivatives; use probabilistic rules, not deterministic rules; can produce the solution without requiring initial solutions by searching from many search points simultaneously [6]. The algorithmic steps of the GA are given below:

Step 1: Create an initial population

A population of chromosomes is created initially. The chromosomes, represents the parameters, are candidate solutions of the problem. Each chromosome composed with genes as a string of binary digits which is called encoding. The population size, N_{pop}, is defined as the number of chromosomes in the population. Set the generation number $generation = 0$.

Step 2: Determine the fitness value of each individual

Fitness function value, which represents the objective function value, is calculated for each individual chromosome.

Step 3: Select next generation (parent population)

According to the fitness function, the strings with high fitness are selected so the best chromosomes are included in the new population by using selection functions, e.g. with replacement, roulette wheel, stochastic uniform, for reproduction.

Step 4: Perform reproduction using crossover (mating)

A locus is chosen randomly and exchanges the subsequences before and after that locus between two chromosomes with crossover probability, $\Pr_c$, by using crossover functions e.g. single point, 2-points, uniform, for creating two offsprings.

Step 5: Perform mutation

Some of the bits in a chromosome is chosen and altered from 0 to 1 or 1 to 0 with mutation probability, $\Pr_m$, for increasing the variability of the population.

Step 6: Replace the current population with the new population

New solutions are replaced with the current solution set. Set $generation = generation + 1$. If $generation < maxgen$ then go to *Step 2* else display results.

Each iteration of this procedure is called generation. The entire set of generations is called a run. Since randomness plays a great role in each run, different runs produce different results. So, it will be better to report statistics of results [16]. The main disadvantages of the GA stem from its computational complexity because of the N_{pop} size of the different estimated parameters in each run [3, 10]. And also, the GA does not guarantee the optimum solution, but leads to solutions acceptably close to the optimal solution [9].

37.3.2 Parameter Estimation with Unscented Kalman Filter

Kalman Filter (KF) and its derivations have been extensively used for linear and nonlinear state estimation problems [1]. The KF is an optimal estimator for linear dynamic systems. However, in real world problems, the systems are generally formulated as complex and nonlinear. For nonlinear stochastic systems and nonlinear deterministic systems KF has also been utilized but it is not optimal. The Extended Kalman Filter (EKF) is commonly used method in the field of nonlinear estimation. The EKF uses the standard KF equations to the first order approximation of the nonlinear model about the last estimate. It is very sensitive to initialization and if the arbitrary noise matrices have not been chosen appropriately filter divergence is inevitable [12]. The UKF is an another nonlinear estimation method for nonlinear state space models. The performance of the UKF estimator is equivalent to the KF for linear systems yet generalizes elegantly to nonlinear systems without the linearization steps required by the EKF. The UKF is based on the Unscented Transformation (UT). The UT is a method for calculating the statistics of a random variable which undergoes a nonlinear transformation [5, 13]. Suppose that x is a random variable which has been transformed a y by using a nonlinear function h ($y = h(x)$) and suppose that x has mean $\bar{x}$ and covariance P_{xx}. To calculate the statistics of y, the UT is used deterministically chosen $2n+1$ sample

points which are called sigma points. The sigma points are chosen to guarantee that the sample mean $\bar{x}$ and sample covariance P_{xx} by the following algorithm

$$\begin{cases} \chi_0 = \bar{x} & , W_0 = \kappa/(n+\kappa) \\ \chi_i = \bar{x} + \left(\sqrt{(n+\kappa)P_{xx}}\right)_i & , W_i = 1/2(n+\kappa), i = 1,\ldots,n \\ \chi_i = \bar{x} - \left(\sqrt{(n+\kappa)P_{xx}}\right)_{i-n} & , W_{i+n} = 1/2(n+\kappa), i = n+1,\ldots,2n \end{cases}$$

in which $\kappa \in R$. $\left(\sqrt{(n+\kappa)P_{xx}}\right)_i$ is the i.th row or column of the matrix square root of $(n+\kappa)P_{xx}$. W_i's are weights of i. sample point with satisfying $\sum_{i=0}^{2n} W_i = 1$. The sigma points are instantiated through the process model, $\gamma_i = h(\chi_i), i = 1,2,\ldots,n$, and the mean and covariance of y are computed by using a weighted sample mean and covariance of the posterior sigma points

$$\bar{y} = \sum_{i=0}^{2n} W_i \gamma_i$$

$$P_{yy} = \sum_{i=0}^{2n} W_i \{\gamma_i - \bar{y}\}\{\gamma_i - \bar{y}\}^T.$$

Consider the following nonlinear discrete-time stochastic system

$$\begin{aligned} x(k) &= f(x(k-1)) + w(k) \\ y(k) &= h(x(k)) + v(k) \end{aligned} \tag{37.3}$$

where $x(k)$ is the $n \times 1$ state vector, $y(k)$ is the $m \times 1$ measurement vector at time instant k. The vector valued nonlinear functions, f and h are state transition and observation functions, respectively. $w(k)$ and $v(k)$ are uncorrelated zero-mean white noise processes with covariance

$$E\left(w(k)w^T(k)\right) = Q(k), E\left(v(k)v^T(k)\right) = R(k). \tag{37.4}$$

The UKF equations for nonlinear system given by Eq. 37.3 are summarized as follows [13, 15]

Step 0: Initialize

$$\widehat{x}(0) = E(x(0))$$

$$P(0) = E(x(0) - \widehat{x}(0))(x(0) - \widehat{x}(0))^T$$

Step 1: Calculate the sigma points

$$\begin{cases} \chi_i(k-1) = \widehat{x}(k-1) & , i = 0 \\ \chi_i(k-1) = \widehat{x}(k-1) + \left(\sqrt{(n+\kappa)P(k-1)}\right)_i & , i = 1, 2, \ldots, n \\ \chi_i(k-1) = \widehat{x}(k-1) + \left(\sqrt{(n+\kappa)P(k-1)}\right)_i & , i = n+1, n+2, \ldots, 2n \end{cases}$$

Step 2: Prediction

$$\chi_i(k\,|k-1) = f(\chi_i(k-1))$$

The predicted mean and covariance are computed as

$$\widehat{x}(k\,|k-1) = \sum_{i=0}^{2n} W_i \chi_i(k\,|k-1)$$

$$P(k\,|k-1) = \sum_{i=0}^{2n} W_i \left(\chi_i(k\,|k-1) - \widehat{x}(k\,|k-1)\right)\left(\chi_i(k\,|k-1) - \widehat{x}(k\,|k-1)\right)^T + Q(k)$$

Step 3: Update

$$\gamma_i(k\,|k-1) = h(\chi_i(k\,|k-1))$$

$$\widehat{y}(k) = \sum_{i=0}^{2n} W_i \gamma_i(k\,|k-1)$$

The weighted covariance matrix of the predicted observations is given below

$$P_{yy}(k) = \sum_{i=0}^{2n} W_i \left(\gamma_i(k) - \widehat{y}(k)\right)\left(\gamma_i(k) - \widehat{y}(k)\right)^T + R(k)$$

and the covariance matrix between the state and the measurement is computed as follows

$$P_{xy}(k) = \sum_{i=0}^{2n} W_i \left(\chi_i(k\,|k-1) - \widehat{x}(k\,|k-1)\right)\left(\gamma_i(k) - \widehat{y}(k)\right)^T.$$

Then the state estimate $\widehat{x}(k)$ and the corresponding covariance matrix $P(k)$ can be updated as

$$K(k) = P_{xy}(k)P_{yy}^{-1}(k)$$

$$\widehat{x}(k) = \widehat{x}\,(k\,|k-1\,) + K(k)\,(y(k) - \widehat{y}(k))$$

$$P(k) = P\,(k\,|k-1\,) - K(k)P_{yy}(k)K^{T}(k)$$

where $K(k)$ is a Kalman Gain matrix.

Step 4: Repeat Steps 1–3 for the next sample.

It is clearly seen that there is no necessity to compute the Jacobian matrix in the UKF algorithm whereas it is needed in the EKF. This can be considered as an advantage of UKF against to the EKF.

37.4 Application

In this section, chemical reaction problem is considered as an application. Rational function response is given as

$$\eta\,(\mathbf{X};\boldsymbol{\theta}) = \frac{\theta_1\theta_3X_1}{1+\theta_1X_1+\theta_2X_2},\ X_1,X_2>0;\ 1<\theta_1<3,\ 10<\theta_2<20,\ 0<\theta_3<1$$

which models the chemical reactions of type $R \rightarrow P_1 + P$. Here, η is the speed of the reaction, X_1 is the partial pressure of the sought product P, X_2 is the partial pressure of the sought product P_1, θ_1 is the absorption equilibrium constant for P_1, θ_2 is the effective constant of the speed of reaction, and θ_3 is the absorption equilibrium constant for the reagent R [14]. The data set is given in Table 37.1.

The objective function is defined as

$$\phi\,(\boldsymbol{\theta}) = \sum_{i=1}^{5}\left[Y_i - \frac{\theta_1\theta_3X_1}{1+\theta_1X_1+\theta_2X_2}\right]^2.$$

Table 37.1 The data set for chemical reaction

X_1 :	0.28, 2.9, 3.2, 4.7, 5.5
X_2:	0, 0.1, 0.82, 1.1, 2
η :	0.33, 0.57, 0.33, 0.31, 0.23

Table 37.2 Initial tunable parameters of the GA and the UKF

GA	UKF
popsize = 100 $\text{Pr}_{cr} = 0.90$ $\text{Pr}_{m} = 0.01$ *Roulette Wheel Selection* *Single Po* int *Crossover* *Bit flip mutation* *maxgen* = 100	$x(0) = \begin{bmatrix} 3 \\ 15.3 \\ 0.73 \end{bmatrix}$; $P(0) = \begin{bmatrix} 1\ 0\ 0 \\ 0\ 1\ 0 \\ 0\ 0\ 1 \end{bmatrix} \times 0.0001$ $R = 0.5$; $Q = \begin{bmatrix} 1\ 0\ 0 \\ 0\ 1\ 0 \\ 0\ 0\ 1 \end{bmatrix} \times 0.1$

Table 37.3 Performance metric results for the GA and the UKF

	GA	UKF
RMSE	0.02611	0.02436

The state-space model of chemical reactions for the UKF is defined as

$$x = \begin{bmatrix} 1\ 0\ 0 \\ 0\ 1\ 0 \\ 0\ 0\ 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$$

$$y = \begin{bmatrix} \dfrac{\theta_3 X_1}{1 + \theta_1 X_1 + \theta_2 X_2} & 0\ 0 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}.$$

The initial parameters of the GA and the UKF are given in Table 37.2. These tunable parameters are defined according to the problem structure and expert knowledge.

The Root Mean Square Error (*RMSE*) is used as performance metric for comparison of the GA and the UKF. The *RMSE* is defined as

$$RMSE = \sqrt{\frac{1}{n-p}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2} \tag{37.5}$$

in which Y_i is observed response, $\hat{Y}_i$ is predicted response, n is number of observations, and p is number of parameters. In this study, n and p are 5 and 3, respectively. The performance metric results for the GA and the UKF are given in Table 37.3.

It is seen from the Table 37.3 that the UKF is slightly better than GA for parameter estimates according to the calculated RMSE values. Therefore, it can be said that the UKF can be used as an alternative method for parameter estimation of complex nonlinear response problems.

37.5 Conclusions

Nonlinear response models are most commonly used as well as linear models for modeling stage of the complex problems. In this study, response functions are considered as nonlinear in the parameters. In order to obtain parameter estimates of the nonlinear functions, two of the derivative-free optimization algorithms are used instead of derivative-based optimization algorithms. One of the method, used in the study, is GA which is an efficiently used metaheuristic method for complex optimization problems. The other one is UKF which is the most popular estimation method for nonlinear state-space models. The tunable parameters of the GA and the UKF are chosen according to the problem structure and expert knowledge. The results show that the GA and the UKF have similar performances according to the *RMSE* metric. However, the UKF is slightly better than the GA. It can be said that the UKF can be preferred as an optimization tool for complex nonlinear optimization problems.

References

1. Bar-Shalom Y, Li XR, Kirubarajan T (2001) Estimation with applications to tracking and navigation. Wiley, New York
2. Box GEP, Draper NR (2007) Response surface mixtures and ridge analysis. Wiley, Hoboken
3. Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, New York
4. Holland J (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, MI
5. Julier S, Uhlmann J (1997) A new extension of the Kalman filter to nonlinear systems. In: International symposium on aerospace/defense sensing, simulation, and controls, Orlando
6. Kapanoglu M, Koca İO, Erdogmus Ş (2007) Genetic algorithms in parameter estimation for nonlinear regression models: an experimental approach. J Stat Comput Simul 77(10):851–867
7. Khuri AI, Cornell JA (1996) Response surfaces, 2nd edn. Dekker, New York
8. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220(4598):671–680
9. Melanie M (1998) An introduction to genetic algorithms, a Bradford book. The MIT Press, Cambridge, MA/London
10. Michalewicz Z (1992) Genetic algorithms + data structures = evolutionary programs. AI series. Springer, New York
11. Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313
12. Ozbek L, Babacan EK, Efe M (2010) Stochastic stability of the discrete-time constrained extended Kalman filter. Turk J Electr Eng Comput Sci 18(2), Tubitak. doi:10.3906/elk-0812-17
13. Wan EA, Merwe RV (2001) The unscented Kalman filter. In: Haykin S (ed) Kalman filtering and neural networks, vol 7. Wiley Publishing, New York
14. Wiens DP (2013) Optimla design for nonlinear models: introduction and historical overview. University of Alberta, Canada
15. Xiong K, Liu LD, Zhang HY (2009) Modified unscented Kalman filtering and its application in autonomous satellite navigation. Aerosp Sci Technol 13:238–246
16. Yao L, Sethares WA (1994) Nonlinear parameter estimation via genetic algorithm. IEEE Trans Signal Process 42(4):927–935