Available online at www.sciencedirect.com
ScienceDirect
Automatica 44 (2008) 278–281
automatica
www.elsevier.com/locate/automatica

Technical communique

Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models☆

Berit Floor Lund[a,b,*], Bjarne A. Foss[b]

[a] *Sintef ICT, Applied Cybernetics, 7465 Trondheim, Norway*
[b] *Norwegian University of Science and Technology, NTNU, Institute of Engineering Cybernetics, 7491 Trondheim, Norway*

Received 18 January 2006; received in revised form 10 November 2006; accepted 23 April 2007
Available online 20 August 2007

Abstract

The paper addresses methods for parameter sensitivity analysis in a large, nonlinear, mechanistic model which is to be run in an on-line estimation scheme. The parameter sensitivity has been obtained by numeric approximation. The paper proposes and applies successive orthogonalization of the sensitivity derivative for parameter ranking. The method is easy to implement and the results are easily interpreted. Orthogonalization of the sensitivity matrix gives a triangular form of the squared sensitivity. The paper shows how the triangular form of the sensitivity derivative gives a particularly easy form of the variance contribution of individual parameters, provided the model error can be assumed Gaussian. This information has been used to decide how many parameters from the ranked set are to be selected for on-line estimation.

Keywords: Sensitivity analysis; Parameter sensitivity; Decomposition methods

1. Introduction

This paper addresses the problem where a high fidelity, simulation model of a system is to be used on-line for some supervision or control purpose. The model gives a mechanistic representation (i.e. based on first principles modelling) of the represented process. On-line use implies more specifically that the model's outputs, $\hat{y}$, are to be adapted to a set of real-time measured outputs, y, by adjusting a subset of the model parameters, θ. The considered model class is given by

$$\hat{y} = g(\varphi, v; x_0),$$

where $\hat{y} \in \mathbb{R}^{n_y}$ is the model output vector, g is a differentiable operator representing an underlying nonlinear, dynamic system described by a model on differential algebraic equation (DAE) form, $\varphi \in \mathbb{R}^{n_\varphi}$ is the model parameter vector, $v \in \mathbb{R}^{n_v}$ is a measured input vector which may be present, and $x_0 \in \mathbb{R}^x$ is the initial state vector.

Mechanistic simulation models normally have a large parameter vector. Many of the parameters can be determined and fixed during modelling. Candidates for on-line updating are found among uncertain or time varying parameters. Often it is necessary to further reduce the candidate set due to insufficient identifiability through the available outputs or more practical reasons such as limited computation time. An assumption in this work is that the reduced parameter vector $\theta \in \mathbb{R}^{n_\theta}$ should have a physical interpretation. This is accomplished by selecting $n_\theta < n_\varphi$ elements of φ and placing these elements into θ. Hence the problem is to select θ to obtain a reasonable balance between the need for a small prediction error, and the need for a fast convergence rate of the on-line optimization algorithms used in the estimation scheme.

This paper suggests to first rank the candidate parameters based on successive orthogonalization of the parameter sensitivity matrix, see Section 3. This method can be used without

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Karl Henrik Johansson under the direction of Editor André Tits. This work has been supported by NTNU, Elkem and SINTEF.

* Corresponding author. Sintef ICT, Applied Cybernetics, 7465 Trondheim, Norway. Tel.: +47 7359 4395; fax: +47 7359 4399.
E-mail addresses: berit.floor.lund@sintef.no (B.F. Lund), Bjarne.Foss@itk.ntnu.no (B.A. Foss).

0005-1098/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2007.04.006

considering the probability distribution of the model error. As a second step, the parameters' individual variance contribution is used to determine how many parameters from the ranked set are to be included for estimation. If model error has a Gaussian probability distribution, then orthogonalization of the sensitivity derivative gives a particularly simple form of the variance contribution of each parameter, see Section 4.

2. Parameter ranking

The sensitivity of the model outputs to the parameter vector is defined as

$$S(\theta, v, W; x_0) = W^{-1/2}\frac{\mathrm{d}\hat{y}(\theta)}{\mathrm{d}\theta} \in \mathbb{R}^{n_y \times n_\theta}, \tag{1}$$

where $W^{-1/2}$ is a diagonal weighting matrix. In the following S will be used for $S(\theta, v, W; x_0)$. It is paramount that S has been sensibly scaled since the scaling directly affects the result of the parameter ranking. Here, the model outputs $\hat{y}$ and θ are variables with a physical interpretation. The scaling has therefore been chosen according to their natural or permissible variation range. Further, S is calculated for a given θ and a given operating point given by v and x_0. A numerical approximation of S will be used in a later example, in which case g needs not be differentiable.

Since the model is nonlinear, global identifiability can generally not be proven. To increase the probability that the model is identifiable over the whole parameter space, the sensitivity must be checked for multiple parameter vector values and operating points. This requires efficient automated analysis methods.

Each column of the scaled sensitivity S will express all outputs' sensitivity to one parameter, and is therefore a sensitivity direction for this parameter. Assuming that S has been properly scaled it is of interest to find out if a column contains several large elements, indicating high sensitivity to the parameter in these outputs. The size of the sensitivity to the parameters can be compared through the norm of the columns. In addition to this, the degree of linear dependence between the columns is of interest. A high linear dependence between two columns would indicate that altering these parameters have a similar effect on the outputs.

Suitable linear transformations of S or $S'S$ can reveal properties such as norm and linear dependence. Li, Henson, and Kurtz (2004) determine parameter ranking by principal component analysis and compute the minimum distance between each sensitivity vector and the eigenvectors (principal direction). Ranking by condition number has been used by Weijers and Vanrolleghem (1997). These methods require the design of a selection algorithm to combine the sensitivity vectors before comparing the properties of different subsets. (Brun, Reichert, & Kunsch, 2001) and (Belsley, 1991) rank sensitivity vector subsets according to the condition number, or "collinearity index", of $C.C$ is given by in $S'S = DCD$ where $D = \mathrm{diag}(\|s_1\|^2, \ldots, \|s_{n_\theta}\|^2)$ and $c_{i,j} = \langle s_i, s_j\rangle/\|s_i\|\|s_j\|$, recognized as $\cos\angle(s_1, s_2)$ in $\mathbb{R}^2$. This is combined with inspection of different norms of the columns of the sensitivity derivative in order to decide which parameters are the most dominant. The method is however largely manual, and not very well suited for analysis of a large number of sensitivity derivative matrices.

Successive orthogonalization as demonstrated in Section 3, is in effect QR transformation with column permutation (QRcp) of S. The advantage of this linear transformation lies in its simplicity and directness. It utilizes the original directions S, making the result easy to interpret. Also, no algorithm for combination of the columns needs to be designed as the selection order is an inherent part of the transformation itself. Although QRcp is described as a type of linear matrix transformation in text books (Golub & Van Loan, 1996), it is not reported applied to parameter ranking for estimation using nonlinear, mechanistic models. The closest application is reported by Kanjilal, Ballav, and Saha (1995) who applied QRcp for finding the most dominant directions in a linear regression matrix in a multilayer neural network.

3. Ranking by successive orthogonalization of S

Successive orthogonalization of S involves ranking the columns of S according to their norm and linear independence, simultaneously. The order of selection is stored in a permutation matrix. The method can be used without assumptions about the probability distribution of the model error. The selection procedure is demonstrated by example using the matrix $S = [a \quad b \quad c]$.

Example 1. Assume column vector a in S has the largest norm and therefore corresponds to the parameter with the highest individual average (normed) sensitivity. This direction is selected, a unit vector, $q_1 = a/\|a\|$, is formed and removed from b and c by subtracting the projection of b and c onto q_1,

$$\widetilde{b} = b - (q_1'b)q_1 \quad \text{and} \quad \widetilde{c} = c - (q_1'c)q_1$$

$\widetilde{b}$ and $\widetilde{c}$ are now orthogonal to q_1. Assuming b originally was pointing in nearly the same direction as a, then a large component would have been subtracted from b when forming $\widetilde{b}$. Assume therefore that $\widetilde{c}$ has the larger norm, and is selected to form the second unit vector, $q_2 = \widetilde{c}/\|\widetilde{c}\|$. The q_2 direction is removed from vectors of the remaining set, which is now only $\widetilde{b}$, to form the new vector $\bar{b}$

$$\bar{b} = \widetilde{b} - (q_2'\widetilde{b})q_2$$

and set $q_3 = \bar{b}/\|\bar{b}\|$. The selection order and decomposition can be summarized into $SE = QR$, where

$$S\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix} = \begin{bmatrix}q_1 & q_2 & q_3\end{bmatrix}\begin{bmatrix}q_1'a & q_1'c & q_1'b\\ 0 & q_2'c & q_2'b\\ 0 & 0 & q_3'b\end{bmatrix}. \tag{2}$$

The result (2) can be further refined by extracting the diagonal, $R = D\bar{R}$ which gives the cross product $E'S'SE=$

$\overline{R}' D' Q' Q D \overline{R} = \overline{R}' D' D \overline{R}$, or

$$E'S'SE = \begin{bmatrix} 1 & 0 & 0 \\ \frac{q_1'c}{q_1'a} & 1 & 0 \\ \frac{q_1'b}{q_1'a} & \frac{q_2'b}{q_2'c} & 1 \end{bmatrix} D'D \begin{bmatrix} 1 & \frac{q_1'c}{q_1'a} & \frac{q_1'b}{q_1'a} \\ 0 & 1 & \frac{q_2'b}{q_2'c} \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

with $D'D = D^2 = \mathrm{diag}\,[\,(q_1'a)^2 \quad (q_2'c)^2 \quad (q_3'b)^2\,]$.

The QR algorithm above can easily be implemented from scratch. The QR algorithm in Matlab can also be used, run in "economy" mode, and with a permutation matrix as one of the return variables. Possible negative values on the diagonal of R are corrected by multiplication with -1 in R and Q, and extraction of the diagonal of R into D must be made.

4. Individual variance contribution

The triangular form of $S'S$ in (3) is a Cholesky form. This connection between the results of Gram–Schmidt orthogonalization of a matrix S and the Cholesky form of $S'S$ is pointed out in Strang (1988, p. 196). Berntsen (1977) pointed out that the triangular form, $S'S = LDL'$ gives a particularly simple expression for the variance contribution of individual parameters in the set. This is explained in the following.

The Cramer Rao inequality $cov(\hat{\theta}) \geqslant M^{-1}$, (Ljung, 1999) states that the inverse Fisher information matrix, M^{-1}, is a lower limit to the covariance of any unbiased estimator of θ. We assume that this is a reasonable measure although prior information about the parameters' value range exists. If the model error, $\varepsilon = y - \hat{y}(\hat{\theta})$, with ε, $y \in \mathbb{R}^{n_y}$, is assumed stochastic with Gaussian properties and covariance W, then

$$M = \frac{\mathrm{d}\hat{y}(\hat{\theta})}{\mathrm{d}\theta}' W^{-1} \frac{\mathrm{d}\hat{y}(\hat{\theta})}{\mathrm{d}\theta} = S'S.$$

The trace, or the sum of the diagonal elements in a covariance matrix, $var(\hat{\theta}) = \mathrm{tr}(cov(\hat{\theta}))$, can be regarded as a measure of the variance of the parameter set. The Cramer Rao inequality is therefore used as a lower limit to the variance of the set, and if the model error can be considered Gaussian, then

$$var\ \hat{\theta} \geqslant \mathrm{tr}(S'S)^{-1}.$$

Berntsen (1977) demonstrated that if $(S'S)^{-1} = UD^{-2}U'$ with $U = L^{-1}$ unit upper diagonal, then the individual variance contribution of each parameter, $\Delta var\ \hat{\theta}_i$, is

$$\Delta var\ \hat{\theta}_i = \frac{\|u_i\|^2}{d_i^2}, \tag{4}$$

where u_i is a column in U. In the following, the results from Berntsen (1977) have been formulated as a proposition based on successive orthogonalization of S instead of LU factorization of $S'S$ used by Berntsen (1977). The basic idea is that if one parameter is appended to a set of $i-1$ parameters, the new sensitivity derivative is formed by appending the i'th parameter's sensitivity derivative, $S_i = [S_{i-1}|s_i]$. Performing successive orthogonalization of S_i in the column order 1 to i gives

$$S_i = Q_i \begin{bmatrix} D_{i-1} & 0 \\ 0 & d_i \end{bmatrix} \begin{bmatrix} \bar{R}_{i-1} & \bar{r}_i \\ 0 & 1 \end{bmatrix}.$$

D_{i-1} and $\bar{R}_{i-1}$ are given by $S_{i-1} = Q_{i-1} D_{i-1} \bar{R}_{i-1}$. The appended i'th column will only appear in the i'th column of $\bar{R}_i^{-1}$. This property of the inverse is also used in the Levinson–Durbin algorithm (Söderström & Stoica, 1989).

Proposition 2. *Define the matrix $S_i = [s_1|\ldots|s_i]$, $S_i \in \mathbb{R}^{n_y \times i}$, n_y is the output vector dimension $n_y \geqslant i$ and $rank(S_i) = i$. $(S_i'S_i)$ has the Cholesky factorization*

$$S_i'S_i = \bar{R}_i' D_i^2 \bar{R}_i,$$

where $D_i = \mathrm{diag}[d_1, \ldots, d_i]$ and $\bar{R}_i$ is upper unit triangular. Then the inverse

$$(S_i'S_i)^{-1} = U_i' D_i^{-2} U_i,$$

where $\bar{R}_i^{-1} = U_i = [u_1, \ldots, u_i]$ and $i \geqslant 1$ has the property

$$\mathrm{tr}(S_i'S_i)^{-1} = \mathrm{tr}(S_{i-1}'S_{i-1})^{-1} + \frac{\|u_i^2\|}{d_i^2}.$$

Proof. $i = 1 : S_1 = s_1$ gives $R_1 = 1$, $U_1 = 1$ then

$$\mathrm{tr}(S_1'S_1)^{-1} = \frac{1}{d_1^2}$$

$i = n : S_n = [S_{n-1}|s_n]$. Assuming there exists an orthogonal form $S_{n-1} = Q_{n-1} D_{n-1} \bar{R}_{n-1}$, then

$$S_n = Q_n \begin{bmatrix} D_{n-1} & 0 \\ 0 & d_n \end{bmatrix} \begin{bmatrix} \bar{R}_{n-1} & \bar{r}_n \\ 0 & 1 \end{bmatrix}$$

and $U_{n-1} = \bar{R}_{n-1}^{-1}$ then U_n given by $U_n \bar{R}_n = I$ equals

$$U_n \begin{bmatrix} U_{n-1} & -U_{n-1}\bar{r}_n \\ 0 & 1 \end{bmatrix},$$

where $\bar{r}_n$ only affects the n'th column of U_n. Setting $u_x = -U_{n-1}\bar{r}_n$ then

$$\begin{aligned} \mathrm{tr}(S_n'S_n)^{-1} &= \mathrm{tr}(U_n D_n^{-2} U_n') \\ &= \mathrm{tr} \begin{bmatrix} U_{n-1} D_{n-1}^{-2} U_{n-1}' + u_x d_n^{-2} u_x' & u_x d_n^{-2} \\ d_n^{-2} u_x' & d_n^{-2} \end{bmatrix} \\ &= \mathrm{tr}(U_{n-1} D_{n-1}^{-2} U_{n-1}') + u_x d_n^{-2} u_x' + d_n^{-2} \\ &= \mathrm{tr}(S_{n-1}'S_{n-1})^{-1} + \frac{\|u_n\|^2}{d_n^2} \end{aligned}$$

where $u_n = [\,-U_{n-1}\bar{r}_n \quad 1\,]$.

The term $\|u_n^2\|/d_n^2$ expresses the cost in terms of additional variance to the whole parameter set after inclusion of parameter n in the ranked set. Calculation of individual variance contributions may reveal leaps in the variance giving crucial information about how many parameters to include from an already ranked set.

5. Application of the parameter ranking scheme to the Simod model

The method has been applied using Simod, an industrial, dynamic simulation model for submerged arc silicon furnaces. The underlying model is a nonlinear, DAE-model with approximately 100 states (Foss & Wasb, 2001), (Lund, Foss, & Løvåsen, 2006). The most dominant parameters for inclusion in an estimation scheme were found using the following steps:

(1) Based on process and model knowledge, create a candidate parameter set from uncertain and/or time varying members of the parameter vector.
(2) Scale based on the natural or permissible parameter and output ranges. This should be possible in a mechanistic model where the parameters and outputs are physical quantities. Scaling based on nominal values generally would give a different result, but is an alternative when a range cannot be assessed.
(3) Generate the sensitivity derivative for as many values of the parameter vector as possible, emphasizing realistic values. Use actual inputs, or make sure that the inputs span the input space.
(4) Rank the parameters using successive orthogonalization for each generated S. Calculate a mean rank for each parameter using the permutation matrix E in (2). Consider also variations in rank.
(5) Check the ranked parameter set for leaps in additional parameter variance, provided an unbiased parameter estimate and a stochastic model error with Gaussian properites are reasonable assumptions.

The candidate parameter set consisted of seven parameters. Analysis of the individual parameter variance contribution in the ranked set revealed a leap going from 3 to 4 parameters. This, together with considerations regarding processing time, contributed to the conclusion of estimating only three out of the seven candidate parameters.

6. Conclusion

Successive orthogonalization has been used to rank the parameters for application in an estimation scheme using a nonlinear mechanistic model. The method gives a simple implementation since the ranking is an inherent property of the method itself. The method can therefore easily be programmed from scratch, no selection algorithm needs to be programmed, and no manual interpretation of complex plots is required. Hence, the method is particularly suited for automated analysis of a large number of sensitivity derivatives calculated for many values of parameter values and model inputs. Trials have shown that the described parameter ranking scheme is both efficient and effective despite the assumptions made about the parameter and prediction error properties.

References

Belsley, D. A. (1991). *Conditioning diagnostics. Collinearity and weak data in regression*. New York: Wiley.

Berntsen, H.E. (1977). Modelling, parameter estimation and identifiability in electrocadiology. PhD thesis, Norwegian Institute of Technology.

Brun, R., Reichert, P., & Kunsch, H. R. (2001). Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, *37*(4), 1015–1030.

Foss, B. A., & Wasb, S. O. (2001). An integration scheme for stiff solid-gas reactor models. *Computer Methods in Applied Mechanics and Engineering*, *190/45*, 6009–6021.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations*. (3rd ed.), Baltimore, MD: John Hopkins University Press.

Kanjilal, P. P., Ballav, P., & Saha, G. (1995). Fast successive selection of variables in linear models using modified QR factorization. *Electronics Letters*, *31*(14), 1204–1205.

Li, R., Henson, M. A., & Kurtz, M. J. (2004). Selection of model parameters for off-line parameter estimation. *IEEE Transactions on Control Systems and Technology*, *12*(3), 402–412.

Ljung, L. (1999). *System identification. Theory for the user*. Englewood Cliffs, NJ: Prentice-Hall.

Lund, B. F., Foss, B. A., & Løvåsen, K. R. (2006). Analysis and characterization of complex reactor behaviour. A case study. *Journal of Process Control*, *16*, 711–718.

Söderström, T., & Stoica, P. (1989). *System identification*. Englewood Cliffs, NJ: Prentice-Hall.

Strang, G. (1988). *Linear algebra and its applications*. (3rd ed.), Philadelphia: Saunders HBJC Publishers.

Weijers, S. R., & Vanrolleghem, P. A. (1997). Procedure for selecting best identifiable parameters in calibrating activated slugde model no. 1 to full-scale plant data. *Water Science and Technology*, *36*(5), 69–79.