CHEMICAL ENGINEERING TRANSACTIONS Volume 17, 2009
Editor Sauro Pierucci
 ISBN 978-88-95608-01-3 ISSN 1974-9791
DOI: 10.3303/CET0917223

A New Strategy to Improve Parameter Estimation

Flavio Manenti and Guido Buzzi-Ferraris

Politecnico di Milano, CMIC dept. "Giulio Natta"
Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy

A new strategy to improve the parameters estimation is formulated analytically and validate on a specific test case. Benefits coming from the novel strategy are then discussed and compared to the existing approaches.
The proposed strategy deals with the problem of the optimal design of experiments for nonlinear models and it is based on the detection of possible correlations among the new planned experimental points and the existing data points on the overall experimental domain.

1. Introduction

An optimal selection of experimental points is a critical issue for the parameter estimation as well as for the discrimination among rival models. The relevant interest in this topic is testified by a large amount of works in the scientific literature. Actually, some strategies to improve parameter assessment and model discrimination were proposed in the scientific literature by long time (Box, 1949; Box and Lucas, 1959; Box and Draper, 1965), whereas some others were more recently proposed (Buzzi-Ferraris and Forzatti, 1983; Buzzi-Ferraris et al., 1984; Forzatti et al., 1987; Ponce de Leon and Atkinson, 1991; Buzzi-Ferraris, 1999).
Notwithstanding, optimal design of experiments, parameters estimation, algorithm robustness, and outliers detection are well-known hard problems and they are still open issues in the scientific community, even for their spread in many fields involving experimental activity.
Nowadays, the diffusion of the object-oriented programming codes, the parallel computing, and the significant increase in available computational power lead to the possibility to formulate and implement new methods to effectively face the problem of parameters estimation. On the other hand, some of these approaches were practically infeasible even only some years ago for their hard implementability and the computational time required for their solution.
A new strategy based on Gram-Schmidt transformation for vectors orthonormalization is therefore analytically proposed, and validated. A brief survey of the existing techniques is given in the section 2. The new strategy for estimating the model parameters is then analytically proposed in section 3. At last, a simple test case is solved and discussed in section 4.

Please cite this article as: Manenti F. and Buzzi-Ferraris G., (2009), A new strategy to improve the parameters estimation, Chemical Engineering Transactions, 17, 1335-1340 DOI: 10.3303/CET0917223

2. Survey of Existing Techniques

To analyze these criteria, some definitions are given here below. For the sake of conciseness, suppose we have only one dependent variable $y = g(\mathbf{x}, \mathbf{b})$. Such an assumption shall be maintained even for the numerical validation. Let $\mathbf{F}_n \in \mathbb{R}^{n \times p}$ and $\mathbf{F}_{n+1} \in \mathbb{R}^{(n+1) \times p}$ be the matrices of the linearized model:

$$\mathbf{F}_n = \begin{bmatrix} f_1(\mathbf{x}_1) & f_2(\mathbf{x}_1) & \cdots & f_p(\mathbf{x}_1) \\ f_1(\mathbf{x}_2) & f_2(\mathbf{x}_2) & \cdots & f_p(\mathbf{x}_2) \\ \cdots & \cdots & \cdots & \cdots \\ f_1(\mathbf{x}_n) & f_2(\mathbf{x}_n) & \cdots & f_p(\mathbf{x}_n) \end{bmatrix} \tag{1}$$

$$\mathbf{F}_{n+1} = \begin{bmatrix} f_1(\mathbf{x}_1) & f_2(\mathbf{x}_1) & \cdots & f_p(\mathbf{x}_1) \\ f_1(\mathbf{x}_2) & f_2(\mathbf{x}_2) & \cdots & f_p(\mathbf{x}_2) \\ \cdots & \cdots & \cdots & \cdots \\ f_1(\mathbf{x}_n) & f_2(\mathbf{x}_n) & \cdots & f_p(\mathbf{x}_n) \\ f_1(\mathbf{x}_{n+1}) & f_2(\mathbf{x}_{n+1}) & \cdots & f_p(\mathbf{x}_{n+1}) \end{bmatrix} \tag{2}$$

where the coefficients $f_k(\mathbf{x}_i)$ are the derivatives of the function g against the parameters b_k calculated in $\mathbf{x}_i$ and for the optimal $\mathbf{b}$.

The $\mathbf{UDV}^{\mathrm{T}}$ factorization (also known as Singular Value Decomposition, SVD) returns the coefficients of the matrix $\mathbf{D}$ (*i.e.* the eigenvalues of $\mathbf{F}^{\mathrm{T}}\mathbf{F}$) and the principal axes $\mathbf{P}_n$ of the matrix $\mathbf{F}_n$. Accordingly, the projection of the matrix $\mathbf{F}_n$ into the space of principal axes can be obtained from the product $\mathbf{F}_n\mathbf{P}_n^{\mathrm{T}}$. The distance between one row of the matrix $\mathbf{F}_n\mathbf{P}_n^{\mathrm{T}}$ and the vector $\mathbf{P}_n\mathbf{f}_{n+1}$ represents the distance between two points in this space.

Let $\mathbf{X}_n$ be the following matrix:

$$\mathbf{X}_n^{\mathrm{T}} = \begin{bmatrix} x_1, x_2, \ldots, x_n \end{bmatrix} \tag{3}$$

The distance between one row of this matrix and the vector $\mathbf{x}_{n+1}$ stays for the distance between the two points estimated in the $\mathbf{x}$-space. Even this matrix can be factorized SVD to get principal axes $\mathbf{P}_x$.

As in the previous case, it is possible to project the matrix $\mathbf{X}_n$ into the space of its principal axes by the product $\mathbf{X}_n\mathbf{P}_x^{\mathrm{T}}$. By doing so, the distance between one row of this matrix and the vector $\mathbf{P}_x\mathbf{x}_{n+1}$ is the distance between the two points in this space.

It is now possible to introduce some of the most common criteria adopted in the design of experiments.

2.1 Criterion no. 1

The new point has to maximize the product $\prod_{i=1}^{P} d_i$ *, i.e. the determinant of the matrix* $\mathbf{F}_{n+1}{}^{\mathrm{T}}\mathbf{F}_{n+1}$.

For each new point $\mathbf{x}_{n+1}$, the factorization of the matrix $\mathbf{F}_{n+1}$ is required. This criterion tries to minimize the confidence volume of parameters (Box and Lucas, 1959; Box, 1971). The volume of hyper-ellipsoid W is $Vol(\mathrm{W}) \approx \dfrac{1}{\sqrt{\det\left(\mathbf{F}_{n+1}{}^{\mathrm{T}}\mathbf{F}_{n+1}\right)}}$ .where no multiplicative factor is taken into account.

2.2 Criterion no. 2

The new point has to maximize the minimum d_i .

This method requires the factorization of the matrix $\mathbf{F}_{n+1}$ for each new point $\mathbf{x}_{n+1}$ and tries to minimize the maximum diameter of the confidence volume of parameters (Hosten, 1974).

2.3 Criterion no. 3

The new point has to maximize the function $f_3 = \mathbf{f}\left(\mathbf{x}_{n+1}\right)^{\mathrm{T}}\left(\mathbf{F}_n{}^{\mathrm{T}}\mathbf{F}_n\right)^{-1}\mathbf{f}\left(\mathbf{x}_{n+1}\right)$.

This criterion adopts a simplified approach against the criterion #4. A single factorization of the matrix $\mathbf{F}_n$ is required, but the solution of $\mathbf{F}_n{}^{\mathrm{T}}\mathbf{w} = \mathbf{f}_{n+1}$ is required for each new point.

2.4 Criterion no. 4

It minimizes the maximum of the function $f_4 = \mathbf{f}^{\mathrm{T}}\left(\mathbf{x}_i\right)^{\mathrm{T}}\left(\mathbf{F}_{n+1}{}^{\mathrm{T}}\mathbf{F}_{n+1}\right)^{-1}\mathbf{f}\left(\mathbf{x}_i\right)$ *against* $\mathbf{x}_i$ *for each new point* $\mathbf{x}_{n+1}$.

Smith (1918) proposed this formulation for polynomial models with a single dependent variable. For each new point, the method needs the factorization of the matrix $\mathbf{F}_{n+1}$, the solution of the system $\mathbf{F}_{n+1}{}^{\mathrm{T}}\mathbf{w} = \mathbf{f}_i$ and the maximization of the function $\mathbf{f}^{\mathrm{T}}\left(\mathbf{x}_i\right)\left(\mathbf{F}_{n+1}{}^{\mathrm{T}}\mathbf{F}_{n+1}\right)^{-1}\mathbf{f}\left(\mathbf{x}_i\right)$. The aim is to minimize the maximum variance for the model prevision.

2.5 Criterion no. 5

The new point has to maximize the sum $\sum_{i=1}^{P} d_i$.

Contrarily to the criterion #1, which tries to minimize the confidence volume of the parameters through the volume of hyper-ellipsoid W, this criterion operates on the perimeter of W. It requires either the product $\mathbf{F}_{n+1}{}^{\mathrm{T}}\mathbf{F}_{n+1}$ or the factorization of the matrix $\mathbf{F}_{n+1}$ for each new point $\mathbf{x}_{n+1}$.

2.6 Criterion no. 6

The new point has to maximize the inverse of $\mathbf{F}_{n+1}$ *condition number* $k = \dfrac{d_{\min}}{d_{\max}}$.

Purpose of this criterion is the minimization of the condition number. For each new point $\mathbf{x}_{n+1}$, it requires the factorization of the matrix $\mathbf{F}_{n+1}$.

2.7 Criterion no. 7

The new point has to maximize the quotient $\dfrac{\sum_{i=1}^{P} d_i}{\prod_{i=1}^{P} d_i}$.

This criterion is maximizes the sphericity of the confidence volume of parameters. It factorizes the matrix $\mathbf{F}_{n+1}$ for each new point $\mathbf{x}_{n+1}$.

All the aforementioned criteria deal with the reduction of possible parameter correlations for the selected model and, on this purpose, some consequences already discussed elsewhere (Buzzi-Ferraris and Manenti, 2009a, b) have to be taken into account.

3. New Strategy

It is also possible to see the problem from another point of view to detect possible correlations among parameters. Exploiting the fact that non-orthonormal vectors can be transformed into orthonormal ones through different techniques (Golub, Van Loan, 1983), vectors orthonormality can be considered as a good index to detect possible parameters correlations and improving the parameters correlation as well.

According to Gram-Schmidt transformation, it is possible to orthonormalize vectors $\mathbf{f}_{i,n+1}$ among each others as follows, in order to select the best $n+1$ additional experimental point:

$$\begin{aligned} \mathbf{g}_1 &= \mathbf{f}_{1,n+1} \\ \mathbf{v}_1 &= \mathbf{g}_1 / \left\| \mathbf{g}_1 \right\|_2 \end{aligned} \tag{4}$$

$$\begin{aligned} \mathbf{g}_2 &= \mathbf{f}_{2,n+1} - \left(\mathbf{f}_{2,n+1}{}^{\mathrm{T}} \mathbf{v}_1 \right) \mathbf{v}_1 \\ \mathbf{v}_2 &= \mathbf{g}_2 / \left\| \mathbf{g}_2 \right\|_2 \end{aligned} \tag{5}$$

$$\begin{aligned} \mathbf{g}_3 &= \mathbf{f}_{3,n+1} - \left(\mathbf{f}_{3,n+1}{}^{\mathrm{T}} \mathbf{v}_2 \right) \mathbf{v}_2 - \left(\mathbf{f}_{3,n+1}{}^{\mathrm{T}} \mathbf{v}_1 \right) \mathbf{v}_1 \\ \mathbf{v}_3 &= \mathbf{g}_3 / \left\| \mathbf{g}_3 \right\|_2 \end{aligned} \tag{6}$$

$$\mathbf{g}_4 = \mathbf{f}_{4,n+1} - \left(\mathbf{f}_{4,n+1}{}^{\mathrm{T}}\mathbf{v}_3\right)\mathbf{v}_3 - \left(\mathbf{f}_{4,n+1}{}^{\mathrm{T}}\mathbf{v}_2\right)\mathbf{v}_2 - \left(\mathbf{f}_{4,n+1}{}^{\mathrm{T}}\mathbf{v}_1\right)\mathbf{v}_1$$
$$\mathbf{v}_4 = \mathbf{g}_4 / \left\|\mathbf{g}_4\right\|_2 \qquad (7)$$
...

where $\mathbf{v}_i$, where $i = 1,...,n$ with n denoting columns $\mathbf{f}_{i,n+1}$ of the matrix $\mathbf{F}_{i,n+1}$, is the set of orthonormalized vectors obtained by Gram-Schmidt process. At this point, it possible to define the new strategy in according to the aforementioned criteria:

"The new point has to minimize the maximum absolute value of cosine of the angles among vectors $\mathbf{v}_i$*".*

In other words, the maximum absolute value of the inner product among vectors $\mathbf{v}_i$ has to be minimized: $\min\left\{\max\left\{\mathbf{v}_i{}^{\mathrm{T}}\mathbf{v}_{j\neq i}\right\}\right\}$.

4. Numerical Validation

Such a strategy was validated on different test cases with different degrees of complexity, by obtaining encouraging results. For space reasons, a trivial application only is proposed hereinafter.

According to the formulation of the previous section, let us consider the case with four columns ($y = 1.$, $y = x$, $y = x^2$, and $y = x^3$) and 7 existing experiments already carried out in three distinct points within the domain of our interest $\left[1.;10.\right]$, specifically in correspondence with $\mathbf{x} = \left\{1.,1.,1.,10.,10.,10.,3.5\right\}$.

Adopting the new strategy, the following series of successive optimal experiments is proposed to improve the parameters estimation: $x_8 = 6.14$, $x_9 = 8.07$, $x_{10} = 6.65$, and $x_{11} = 9.50$.

Cleverly, the new strategy does not re-propose any existing point (new points do not overlap the existing ones) and it tends to exploit the overall experimental domain as well. By doing so, it satisfactorily overcomes one of the main shortcomings of the aforementioned criteria no. 1-7, which are usually unable to propose new points far from the existing experiments.

Unfortunately, some other shortcomings affecting the previous criteria no. 1-7 are still present even in the new strategy. One of the most important is the need to have at least $n - 1$ distinct points, where n is the number of columns $\mathbf{f}_{i,n+1}$, to initialize the procedure.

5. Conclusions

This research activity proposed a novel strategy to face the problem of the parameters estimation. The approach was analytically proposed, and validated on specific test cases giving satisfactory results. Actually, the novel strategy does not tend to re-propose the same experimental points (by preventing the overlapping of the new experiments to the existing ones), contrarily to what happens by using one of the traditional criteria mentioned in section 2.

7. References

Box, G.E.P. (1949), A general distribution theory for a class of likelihood criteria, Biometrika, 36, 317-346.

Box, M.J. (1971), An experimental design criterion for precise estimation of a subset of the parameters in a nonlinear model, Biometrika, 58, 149-153.

Box, G.E.P., Draper, N.R. (1965), The Bayesan estimation of common parameters from several responses, Biometrika, 52, 355-365.

Box, G.E.P., Lucas, H.L. (1959), Design of experiments in nonlinear situations, Biometrika, 46, 77-90.

Buzzi Ferraris, G. (1999), Planning of experiments and kinetic analysis, Catalysis Today, 52, 125-132.

Buzzi Ferraris, G., Forzatti, P. (1983), A new sequential experimental design procedure for discriminating among rival models, Chem. Eng. Science, 38, 225-232.

Buzzi Ferraris, G., Forzatti, P., Hemig, G., Hofmann, H. (1984), Sequential experimental design for model discrimination in the case of multiple responses, Chem. Eng. Science, 39, 81-85.

Buzzi-Ferraris, G., Manenti, F. (2009a), Interpolation and Regression Models for the Chemical Engineer Solving Numerical Problems, WILEY-VCH, Weinheim, Germany, 2009 (in press).

Buzzi-Ferraris, G., Manenti, F. (2009b), Kinetic Models Analysis. Chemical Engineering Science, 64(5), 1061-1074.

Forzatti, P., Buzzi Ferraris, G., Canu, P., Tronconi, E. (1987), Sequential design of experiments. A review of the procedures for increasing the precision of the parameter estimates, Ing. Chim. It., 23, 1.

Golub, G., Van Loan, C. (1983), Matrix Computations, John Hopkins Press, Baltimore.

Hosten, L.H. (1974), A sequential experimental design procedure for precise parameter estimation based upon the shape of the joint confidence region, Chem. Eng. Sci., 29, 2247-2252.

Ponce de Leon, A.C., Atkinson, A.C. (1991), Optimum experimental design for discriminating between two rival models in the presence of prior information, Biometrika, 78, 601-608.

Smith, K. (1918), On the Standard Deviations of Adjusted and Interpolated Values of an Observed Polynomial Function and its Constants and the Guidance They Give towards a proper Choice of the Distribution of Observations, Biometrika, 12, 1918, 1-85.