Multiobjective parameter estimation for nonlinear systems: Affine information and least-squares formulation

Article *in* International Journal of Control · June 2007

DOI: 10.1080/00207170601185053

3 authors:

Erivelton Nepomuceno
Maynooth University

Ricardo H. C. Takahashi
Federal University of Minas Gerais

Luis Antonio Aguirre
Federal University of Minas Gerais

International Journal of Control
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713393989

Multiobjective parameter estimation for non-linear systems: Affine information and least-squares formulation

First Published on: 16 May 2007
To cite this Article: Nepomuceno, E. G., Takahashi, R. H. C. and Aguirre, L. A. , 'Multiobjective parameter estimation for non-linear systems: Affine information and least-squares formulation', International Journal of Control, 80:6, 1 - 9
To link to this article: DOI: 10.1080/00207170601185053
URL: http://dx.doi.org/10.1080/00207170601185053

International Journal of Control
2007, 1–9, iFirst

Multiobjective parameter estimation for non-linear systems: Affine information and least-squares formulation

E. G. NEPOMUCENO†, R. H. C. TAKAHASHI*‡ and L. A. AGUIRRE§

†Dep. de Eng. Elétrica, Univ. Federal de São João Del-Rei, Pça. Frei Orlando, 170, São João Del-Rei, 36307-352, MG, Brasil
‡Dep. de Matemática, Univ. Federal de Minas Gerais, Av. Antonio Carlos, 6627, 30123-970, Belo Horizonte, MG, Brasil
§Dep. de Eng. Eletrônica, Univ. Federal de Minas Gerais, Av. Antônio Carlos, 6627, 31270-901, Belo Horizonte, MG, Brasil

(*Received 19 September 2006; in final form 20 December 2006*)

This paper defines a class of system information—affine information—that includes both the dynamic residuals and some types of auxiliary information that can be used in system parameter estimation as special cases. The types of information that can be cast under the affine information format give rise to quadratic functions that measure the extent to which a model fits such information, and that can be aggregated in a single weighted quadratic cost functional. This allows the definition of a multiobjective methodology for parameter estimation in non-linear system identification, which allows taking into account any type of affine information. The results are presented in terms of a set of efficient solutions of the multiobjective estimation problem—such a solution set is more meaningful than a single model. Since any affine information leads to a convex (quadratic) functional, the whole set of efficient solutions is exactly accessible via the minimization of the quadratic functional with different weightings, via a least-squares minimization (a non-iterative, computationally inexpensive procedure). The decision stage, in which a single model is chosen from the Pareto-set, becomes well-defined with a single global solution. Residual variance, fixed point location, static function and static gain are shown to fit in the class of affine information. A buck DC-DC converter is used as example.

1. Introduction

Black-box identification uses input and output data, usually acquired from dynamical tests, as the only source of information about the system (Söderström and Stoica 1989). In such a framework, other features of the system are not directly considered, as for instance, static gain, number and location of fixed points (equilibria) and the static function. Those neglected features could play a substantial role in the construction of a suitable model if they were taken into account to some extent (Eskinat *et al*. 1993, Tulleken 1993, Johansen 1996, 2000, Pearson and Pottmann 2000). Any information about the system, besides the measured data, is called *a priori* knowledge (Sjöberg *et al*. 1995) or auxiliary information (Eskinat *et al*. 1993). Gray-box identification techniques employ such kind of information for the purpose of model building.

This paper is concerned with the use of auxiliary information in parameter estimation. By auxiliary information it is meant information apart from the set of dynamical data. Classical estimation algorithms perform a particular kind of optimization, which minimizes the sum of squared one-step-ahead prediction errors, ξ. Such algorithms minimize the quadratic functional $\|\xi\|_2^2$, using a least-squares (LS) algorithm. Auxiliary information can be taken into account in two basic

*Corresponding author. Email: taka@mat.ufmg.br

International Journal of Control
ISSN 0020–7179 print/ISSN 1366–5820 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00207170601185053

ways: either defining a composite cost function as a weighted sum of the original objective plus new objectives related to the auxiliary information, or using the auxiliary information not to build objectives, but to define constraints that will represent this knowledge (Johansen 1996; Aguirre and Corrêa 2002). It should be noticed that, in the preceding formulations for using auxiliary information, either defining constrained problems or weighted augmented objective functions, the resulting optimization problem usually becomes a generic problem of minimization of a non-linear function: the numeric solution is obtained from iterative non-linear optimization tools. This is fundamentally different from the classical minimization of the squared prediction errors only, via least-squares procedure, because: (i) while the generic non-linear optimization algorithms can find local solutions, the LS procedure leads to the (analytical) global optimum; and (ii) the generic non-linear optimization is necessarily iterative, while the LS procedure is non-iterative, what results in some orders of magnitude of difference in terms of computational effort.

Johansen (1996) presented a rather general study of how to formulate model estimation procedures with auxiliary information as optimization problems. Johansen defined the optimization problem with the objective function given by a weighted sum of the squared dynamic error plus penalty terms that account for non-smoothness of the model, mismatch of the model and a default model (a non-linear model that does not have polynomial NARX structure), and violation of some soft constraints (for instance, deviation from known linear models that are valid on some operating conditions). The optimization problem could have also some hard constraints, associated, for instance, to open-loop model stability. After defining this general problem, Johansen presented a "taxonomy" of the possible types of optimization problems: (i) quadratic criterion (in the model parameters) and no constraints; (ii) quadratic criterion and linear finite-dimensional constraints; (iii) non-quadratic criterion and no constraints; (iv) non-quadratic criterion and finite-dimensional constraints; (v) any criterion and infinite-dimensional constraints. Each problem class can be solved with a different class of optimization methods. In addition, Johansen noticed that the class (i) of problems would be computationally the easiest one, since such problems could be solved by non-iterative algebraic methods.

This paper further studies the class (i) of problems that can be solved by non-iterative least-squares methods (problems with quadratic-in-the-parameter criterion and no hard constraints). It is pointed out here that the class of system information (called here the system affine information) that gives rise to errors that are affine in the model parameters presents the property that the ℓ_2-norm of such errors become quadratic functions that measure the extent to which a model fits such information. This kind of information, therefore, is the basic building block for problems of type (i) as defined in Johansen (1996). As pointed out by Johansen (1996), the resulting objective function, composed as a weighted sum of several such quadratic functionals, can be minimized via the LS procedure, in this way recovering both the features of non-iterative computation and guaranteed global solution.

Johansen (1996) noticed that additional work was needed in order to develop ways of coding prior information into the model by means of objectives and/or constraints. It is believed that the present paper contributes in such a direction by showing that residual variance, fixed point location, static function and static gain fit into the class of affine information. This was not realized in previous related works (Johansen 1996, Aguirre and Corrêa 2002).

This paper further exploits the idea presented in Johansen (2000) and Nepomuceno *et al.* (2003), of defining the problem of parameter estimation with auxiliary information as a multiobjective optimization problem. Johansen (2000) has proposed the multi-objective identification in the context of linear FIR models, using a progressive decision-making scheme (Chankong and Haimes 1983), in which the trade-off between all objectives is examined at each step, progressively leading to the final model. This procedure can be considered a further step after Johansen (1996), since in that work the main difficulty was exactly to define a procedure for finding the suitable weighting factors for the different auxiliary informations. Nepomuceno *et al.* (2003) have posed the problem of parameter estimation of NARX models in another multiobjective optimization setting: an *a posteriori* decision scheme (Chankong and Haimes 1983) has been used, by defining two sub-problems: first, find a set of samples of the non-dominated solutions (or the Pareto-set solutions); after, choose a final solution from this set, using some decision criterion. The work (Nepomuceno *et al.* 2003), however, has performed this task with two objectives only: the time-series fitting error and the fixed-point fitting error.

The approach that has been sketched in Nepomuceno *et al.* (2003) is further developed here: the *Pareto-set* is estimated for an arbitrary number of objectives originated from affine information sets. The reasoning behind the proposed approach is: (i) since any affine information leads to a convex (quadratic) functional, the whole set of efficient solutions is exactly accessible via the minimization of the weighted quadratic functional (Chankong and Haimes 1983), through a least squares minimization (notice that this is not true for functionals composed by weighted sums of generic non-linear, possibly non-convex, functions, as would be the case in Johansen (2000)); (ii) since the least squares procedure has low computational cost, it seems reasonable to perform a large number of model estimations in order to find a representative sample of the Pareto-set (this would not be practical in the context of models estimated via iterative optimization procedures); (iii) once the Pareto-set samples become available, a decision procedure takes place, picking up a single model from this set. It should be noticed that the Pareto-optimal set is more meaningful than a single model, and can reveal the relationship among the several data sets that are employed as auxiliary information.

The computational framework proposed in this paper, which yields the Pareto-set, seems to be very welcome in the context of gray-box modeling which can easily turn out to be very computationally demanding and require from the user a sophisticated tuning stage (Johansen 1996). Johansen (1996) discussed five types of auxiliary information. This paper discusses other two types plus steady-state data (also considered by Johansen, which is used in a different way in this paper). Moreover, one of the great challenges is how to code prior information in the form of constraints or criteria penalty terms on the model parameters. Coding the auxiliary information for the class of system affine information is one of the contributions of this paper.

This paper is organized as follows: § 2 provides a general view of multiobjective estimation; in § 3, the class of affine information is defined, a single-step LS algorithm that yields the set of efficient solutions is provided, and the decision stage is discussed; § 4 shows that fixed points (equilibria) location, static function and static gain of NARX polynomials (Leontaritis and Billings 1985) can be cast in the framework of affine information. Consequently, such type of auxiliary information can be effectively used during parameter estimation. An example is given in § 5 and the main conclusions of the paper appear in § 6.

2. Affine information and multiobjective formulation

Definition 1 (Affine Information): Consider the parameter vector $\hat{\theta} \in \Re^n$, a vector $\mathbf{v} \in \Re^p$ and a matrix $G \in \Re^{p\times n}$. Both $\mathbf{v}$ and G are assumed to be accessible. Moreover, suppose $G\hat{\boldsymbol{\theta}}$ constitutes an estimate of $\mathbf{v}$, such that $\mathbf{v} = G\hat{\boldsymbol{\theta}} + \epsilon$, where $\epsilon \in \Re^p$ is an error vector. Then $[\mathbf{v}, G]$ is said to be an affine information pair of the system.

Consider the NARX model described by the following equation:

$$y(k) = F^{\ell}\begin{bmatrix} y(k-1),\ldots,y(k-n_y), \\ u(k-d),\ldots,u(k-d-n_u+1) \end{bmatrix} + e(k), \quad (1)$$

where n_y, n_u and n_e are the maximum lags considered for the output $y(k)$, input $u(k)$ and noise $e(k)$, respectively, and d is the delay. In this paper the non-linear function $F^{\ell}[\cdot]$ is taken to be a polynomial with degree $\ell \in \mathbf{Z}^+$. Because $F^{\ell}[\cdot]$ is linear in the parameters, (1) can be expressed as follows:

$$y(k) = \psi^T(k-1)\hat{\boldsymbol{\theta}} + \xi(k), \quad (2)$$

where $\psi(k-1)$ contains linear and non-linear combinations of output, input, and noise terms up to and including time $k-1$.

Example 1: Taking (2) over a set of data yields

$$\mathbf{y} = \Psi\hat{\boldsymbol{\theta}} + \boldsymbol{\xi}. \quad (3)$$

According to the definition above, $[\mathbf{y}, \Psi]$ is an affine information pair, where $\mathbf{y} \in \Re^N$, $\Psi \in \Re^{N\times n}$, $\epsilon = \boldsymbol{\xi}$ and $p = N$. In § 3, other types of affine information will be discussed.

The vector $\hat{\boldsymbol{\theta}}$ is usually estimated by minimizing convex functionals of the form

$$\left.\begin{aligned} J(\hat{\boldsymbol{\theta}}) &= \|\epsilon\|_2^2 = (\mathbf{v} - G\hat{\theta})^{\mathrm{T}}(\mathbf{v} - G\hat{\theta}) \\ J_{\mathrm{LS}}(\hat{\boldsymbol{\theta}}) &= \|\boldsymbol{\xi}\|_2^2 = (\mathbf{y} - \Psi\hat{\theta})^{\mathrm{T}}(\mathbf{y} - \Psi\hat{\theta}), \end{aligned}\right\} \quad (4)$$

where the last functional is minimized by the least-squares estimator. In a multiobjective approach, the problem is to minimize

$$\mathbf{J}(\hat{\theta}) = \begin{bmatrix} J_1(\hat{\theta}) & \ldots & J_m(\hat{\theta}) \end{bmatrix}^{\mathrm{T}}, \quad (6)$$

where $\mathbf{J}(\cdot)\colon \Re^n \mapsto \Re^m$. The outcome is a set of solutions—called the Pareto-set—that describes the trade-off among these objectives, namely the minimization of each cost function. In this paper, the cost functions $J_2(\hat{\theta})\ldots J_m(\hat{\theta})$ take into account auxiliary information about the system.

In general, there is not a unique solution (model) that simultaneously minimizes all the different cost functions $J_i(\cdot)$. Rather, several solutions (models) are found with the property that the improvement of any objective necessarily implies the loss in some other objective. These are the efficient solutions or the Pareto-set solutions. Any parameter vector which is an efficient solution will be referred to as a Pareto-model. (Because the model structure is assumed to be known when it comes to parameter estimation, there is a one-to-one correspondence between each parameter vector on the Pareto-set and a model. Thus, Pareto-models are "best" in the sense that there is no ordering among them, and that there is always some Pareto-model that is better than any non-efficient solution, when compared in all optimization objectives. In the case of all functionals J_i being convex (notice that this is not true if any functional J_i is not convex), the Pareto-set can be found by defining (Chankong and Haimes 1983)

$$W = \left\{ \mathbf{w} | \mathbf{w} \in \Re^m, w_j \geq 0 \text{ and } \sum_{j=1}^{m} w_j = 1 \right\} \tag{6}$$

and solving the convex optimization problem

$$\hat{\theta}^* = \arg\min_{\hat{\theta}} \langle \mathbf{w}, \mathbf{J}(\hat{\theta}) \rangle. \tag{7}$$

For each vector $\mathbf{w}$, which defines a particular combination of weights to the various cost functions involved, a solution $\hat{\theta}^*$ belonging to the Pareto-set $\hat{\Theta}^*$ is found. The entire Pareto-set is associated to the set of all realizations of $\mathbf{w} \in W$.

An effective single-step computational strategy for solving the multiobjective problem (7) by means of an LS formulation is provided by the following theorem that is stated without proof.

Theorem 1: *Let $[\mathbf{v}_i, G_i]$ with $i = 1, \ldots, m$ be m affine information pairs related to a system, where $\mathbf{v}_i \in \Re^{p_i}$ and $G_i \in \Re^{p_i \times n}$. Assume that at least one of the matrices G_i is full column rank. Let $\mathcal{M}$ be a given model structure which is linear in the parameter vector $\hat{\theta} \in \Re^n$. Then the m affine information pairs can be simultaneously taken into account while estimating the parameters of model $\mathcal{M}$, by solving*

$$\hat{\theta}^* = \arg\min_{\hat{\theta}} \sum_{i=1}^{m} w_i (\mathbf{v}_i - G_i\hat{\theta})^{\mathrm{T}} (\mathbf{v}_i - G_i\hat{\theta}), \tag{8}$$

with $\mathbf{w} = [\, w_1 \ldots w_m \,]^{\mathrm{T}} \in W$. The unique solution of (8) is given by

$$\hat{\theta}^* = \left[\sum_{i=1}^{m} w_i G_i^{\mathrm{T}} G_i \right]^{-1} \left[\sum_{i=1}^{m} w_i G_i^{\mathrm{T}} \mathbf{v}_i \right]. \tag{9}$$

Example 2: If the only information available is the usual input/output data (stored in the affine information pair $[\mathbf{y}, \Psi]$), then $m = 1$, $w_1 = 1$, $[\mathbf{v}, G] = [\mathbf{y}, \Psi]$ and Theorem 1 reduces to the conventional (mono-objective) LS solution. The full rank assumption on Ψ can be interpreted as the usual requirement that the system dynamics has been "fully excited" by the input signal. Therefore, this information, being employed jointly with any other affine information for parameter estimation, satisfies the assumption of Theorem 1 that at least one of the matrices G_i is full column rank. On the other hand, there are situations in which Ψ is approximately rank-deficient. In such cases, the parameter estimation by the conventional approach would be difficult or impossible. By using auxiliary information, such rank deficiency could be avoided, and the estimation problem could become numerically well-behaved.

2.1 *The decision stage*

The Pareto-models are the outcome of the multiobjective optimization problem. It is believed that such a set of models, which combines different sources of information with different weights, is in itself a rich picture of the system. The whole set of Pareto-models gives a much better and wider view of the system in so far as it takes into account auxiliary information. Nevertheless, in a particular application, a single model might be required. In that case, in order to choose one Pareto-model the following general procedures can be followed, depending on the kind of decision information that is available.

In this way, the final model(s) will somehow include information of several sources, namely from all the cost functions used in the multiobjective optimization procedure (parameter estimation) and from the decision stage (model selection). Because the decision stage is bound to be strongly application-dependent, below only some general concepts are given. It is stressed that this stage is also an interesting feature of the present procedure, not included in previous methods.

2.1.1 Qualitative information decision. In most of the applications of model identification, the final judgement of the end user is the definitive criterion that must be fulfilled by any identified model, in order to be accepted. This judgement is based on the joint analysis of several data, like frequency response, similarity with the real process under specific conditions, qualitative behaviour (for instance: Poincaré maps), and so forth. These data are filtered by the user "expert" evaluation that tries to take into account the effect of each model feature in its intended use. This kind of "expert" information processing is unlikely to be mapped into an explicit mathematical model, and must be extracted through a human-machine interactive procedure.

The systematic procedure for performing such interaction (the decision-making) is based on the hypothesis that there is a utility function underlying the user's preferences, that must be quasi-concave, in order to define a well-posed problem (Chankong and Haimes 1983). In conjunction with the fact that the multiobjective criterion functions are all convex in the context of affine information, this allows to state that a well-defined solution set exists, and can be found via some simple decision procedures. Chankong and Haimes (1983) describe some of such procedures, that involve queries that ask for binary comparisons (the user is asked to compare two candidate solutions, quantifying the extent to which one is preferred in relation to the other one) between Pareto points, leading to a single preferred solution. The procedure is structured in order to minimize the number of queries, while leading to the global solution.

2.1.2 Quantitative information decision. There are also some instances of quantitative information that can be used for the purpose of decision. These are of two natures. (i) Test data, that has been kept apart from the estimation data, can be used in the decision stage. For instance, it is possible to search for the model inside the Pareto-set that minimizes the sum of squared residuals of the model in relation to some time series data that has not been used in the estimation stage. This procedure can be useful for the purpose of avoiding overfitting of estimation data. (ii) It has been shown in Nepomuceno *et al.* (2003) that the weighting parameter (in that case, in the context of a bi-objective formulation) could be used to find models with different dynamic behaviours, resembling a "bifurcation parameter".

3. Auxiliary information as affine information

It will be shown that fixed points, static function and static gain are instances of affine information that can be used in parameter estimation (see Theorem 1).

3.1 *Fixed points*

The deterministic part of a polynomial NARX model can be expanded as the summation of terms with degrees of non-linearity in the range $1 \le m \le \ell$. Each mth-order term is multiplied by a coefficient $c_{p,m-p}(n_1, \ldots, n_m)$ as follows:

$$y(k) = \sum_{m=0}^{\ell} \sum_{p=0}^{m} \sum_{n_1, n_m}^{n_y, n_u} c_{p,m-p}(n_1, \ldots, n_m) \times \prod_{i=1}^{p} y(k-n_i) \prod_{i=p+1}^{m} u(k-n_i), \tag{10}$$

where

$$\sum_{n_1, n_m}^{n_y, n_u} \equiv \sum_{n_1=1}^{n_y} \cdots \sum_{n_m=1}^{n_u}$$

and the upper limit is n_y if the summation refers to factor in $y(k-n_i)$ or n_u for factors in $u(k-n_i)$.

Considering a locally asymptotically stable model in steady-state excited by a constant input, equation (10) can be written as

$$y(k) = \sum_{n_1, n_m}^{n_y, n_u} c_{p,m-p}(n_1, \ldots, n_m) \sum_{m=0}^{\ell} y(k-1)^p u(k-1)^{m-p}, \tag{11}$$

and the following definition can be presented.

Definition 2 (Cluster coefficients)**:** In equation (11), $\sum_{n_1, n_m}^{n_y, n_u} c_{p,m-p}(n_1, \ldots, n_m)$ are the coefficients of the term clusters $\Omega_{y^p u^{m-p}}$, which contain terms of the form $y(k-i)^p u(k-j)^{m-p}$ for $m = 0, \ldots, \ell$ and $p = 0, \ldots, m$. Such coefficients are called cluster coefficients and are represented by capital sigmas: $\sum_{y^p u^{m-p}}$ (Aguirre and Billings 1995).

The fixed points of NAR(X) polynomial models with degree of non-linearity ℓ in the output are the solutions of

$$\Sigma_{y^\ell} \bar{y}^\ell + \cdots + \Sigma_{y^2} \bar{y}^2 + (\Sigma_y - 1)\bar{y} + \Sigma_0 = 0, \tag{12}$$

where the known constants Σ_0, $\Sigma_{y^i}, i = 1, \ldots, \ell$ are the model cluster coefficients (Aguirre and Mendes 1996). Fixed points are defined for deterministic autonomous models, hence the X part of the model will not have any effect on the results of § 3.1. The computation of the cluster coefficients of a model with fixed points $[\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_\ell]$, is performed by equating the terms with same degree of non-linearity in

$$\zeta \prod_{i=1}^{\ell} (\bar{y} - \bar{y}_i) = \Sigma_{y^\ell} \bar{y}^\ell + \cdots + \Sigma_{y^2} \bar{y}^2 + (\Sigma_y - 1)\bar{y} + \Sigma_0 \tag{13}$$

in which $\zeta = \Sigma_{y^\ell}$. Consider the set Σ of cluster coefficients

$$\Sigma = \left[\Sigma_{y^\ell}, \ldots, \Sigma_{y^2}, (\Sigma_y - 1), \Sigma_0\right]^{\mathrm{T}}. \tag{14}$$

Because (12) or (13) times a constant will yield the same set of fixed points, some kind of normalization should be used to avoid a large variance in the estimated parameters. Let

$$\Sigma_{\mathrm{LS}} = \left[\Sigma_{\mathrm{LS}y^\ell}, \ldots, \Sigma_{\mathrm{LS}y^2}, (\Sigma_{\mathrm{LS}y} - 1), \Sigma_{\mathrm{LS}0}\right]^{\mathrm{T}} \tag{15}$$

be the cluster coefficients of a model with parameters estimated by standard LS. Hence the normalized set of cluster coefficients is

$$\left.\begin{aligned} \tilde{\boldsymbol{\sigma}} &= \frac{\|\Sigma_{\mathrm{LS}}\|}{\|\Sigma\|}[\Sigma_{y^\ell}, \ldots, \Sigma_{y^2}, (\Sigma_y - 1), \Sigma_0]^{\mathrm{T}} \\ &= [\sigma_\ell, \ldots, \sigma_2, (\sigma_1 - 1), \sigma_0]^{\mathrm{T}} \\ \boldsymbol{\sigma} &= [\sigma_\ell, \ldots, \sigma_2, \sigma_1, \sigma_0]^{\mathrm{T}}, \end{aligned}\right\} \tag{16}$$

where $\|\cdot\|$ is the Euclidean norm.

Now suppose a specific set of fixed points, $\mathcal{P}$, is given and it is desired that the set, $\hat{\mathcal{P}}$, of fixed points of a model to be identified should approximate $\mathcal{P}$ in some sense. This can be achieved minimizing the following cost function (Nepomunceno *et al.* 2003)

$$J_{\mathrm{FP}}(\hat{\theta}) = (\boldsymbol{\sigma} - \hat{\boldsymbol{\sigma}})^{\mathrm{T}}(\boldsymbol{\sigma} - \hat{\boldsymbol{\sigma}}) = (\boldsymbol{\sigma} - S\hat{\boldsymbol{\theta}})^{\mathrm{T}}(\boldsymbol{\sigma} - S\hat{\boldsymbol{\theta}}), \tag{17}$$

in which $S \in \Re^{\ell+1\times n}$ is a constant matrix, that maps parameters to the cluster coefficients, that is $\hat{\boldsymbol{\sigma}} = S\hat{\boldsymbol{\theta}}$. Therefore, $[\boldsymbol{\sigma}, S]$ is an affine information pair.

Example 3: Suppose it is desired to obtain a model that will simultaneously provide a good fit to a given set of data and approximate a given set of fixed points. The parameters of such a model can be estimated minimizing a linear combination of J_{LS} (4) and J_{FP} (17). According to Theorem 1, the required solution is

$$\hat{\theta}^* = [w_1\Psi^{\mathrm{t}}\Psi + w_2S^{\mathrm{t}}S]^{-1}[w_1\Psi^{\mathrm{t}}\mathbf{y} + w_2S^{\mathrm{t}}\boldsymbol{\sigma}]. \tag{18}$$

Varying w_1 and w_2 such that $w_1, w_2 \geq 0$ and $w_1 + w_2 = 1$, it is possible to obtain the entire Pareto-set. For $[w_1, w_2] = [1, 0]$, equation (18) yields the standard LS solution and for $[w_1, w_2] = [0, 1]$, conversely, equation (18) yields the solution that takes into account exclusively the location of fixed points. Clearly, such solutions are the extreme points of the Pareto-set. In between such extreme solutions there are infinite Pareto-models that incorporate the system information available in the vectors $\mathbf{y}$ and $\boldsymbol{\sigma}$. For instance, the solution obtained for $[w_1, w_2] = [0.5, 0.5]$ would be the result of giving the same weight to both information sources.

3.2 *Static function*

The static or steady-state behaviour of an asymptotically stable system can be obtained by fixing the input to a constant value $\bar{u}$ in which case the output will achieve the steady-state $\bar{y}$. For a general system $\bar{u}$ and $\bar{y}$ are related by $\bar{y} = f(\bar{y}, \bar{u})$, where $f(\cdot)$ is the static function. Hammerstein models are special cases for which $\bar{y} = f(\bar{u})$. For a general NARX polynomial model, due to linearity in the parameters, the following expression holds

$$\bar{y}_i = f(\bar{y}_i, \bar{u}_i) = \mathbf{q}_i^{\mathrm{T}} R\boldsymbol{\theta}, \tag{19}$$

$$\mathbf{q}_i^T = \begin{bmatrix} 1 & \bar{y}_i, \bar{y}_i^2 \ldots \bar{y}_i^\ell & \bar{u}_i, \bar{u}_i^2 \ldots \bar{u}_i^\ell & F_{yu} \end{bmatrix}, \tag{20}$$

$$R\boldsymbol{\theta} = \begin{bmatrix} \Sigma_0 & \Sigma_y, \Sigma_{y^2} \ldots \Sigma_{y^\ell} & \Sigma_u, \Sigma_{u^2} \ldots \Sigma_{y^\ell} & F_\Sigma \end{bmatrix}^{\mathrm{T}}, \tag{21}$$

where F_{yu} stands for all non-linear monomials in the model that involve $y(k)$ and $u(k)$, ℓ is the largest non-linearity in the model and need not be the same for input and output terms. F_Σ stands for all the cluster coefficients corresponding to all the term clusters in F_{yu} and R is a constant matrix of ones and zeros that maps the parameter vector to the cluster coefficients. Thus, given a NARX polynomial model with estimated parameter vector $\hat{\theta}$, a full picture of the static function can be described in matrix form as $\hat{\bar{\mathbf{y}}} = QR\hat{\theta}$, where $Q = \left[\mathbf{q}_1 \ldots \mathbf{q}_{n_{\mathrm{sf}}}\right]$, in which n_{sf} different steady-state points $(\bar{u}_i, \bar{y}_i)$ were considered. Now suppose that n_{sf} different steady-state operating points $(\bar{u}_i, \bar{y}_i)$ are specified and it is desired that the estimated model should approximate such a static function. This can be achieved by minimizing the following cost function

$$J_{\mathrm{SF}}(\hat{\theta}) = (\bar{\mathbf{y}} - \hat{\bar{\mathbf{y}}})^{\mathrm{T}}(\bar{\mathbf{y}} - \hat{\bar{\mathbf{y}}}) = (\bar{\mathbf{y}} - QR\hat{\theta})^{\mathrm{T}}(\bar{\mathbf{y}} - QR\hat{\theta}). \tag{22}$$

Clearly, $[\bar{\mathbf{y}}, QR]$ is an affine information pair.

Example 4: Consider the following model (Aguirre *et al.* 2000)

$$\begin{aligned} y(k) &= \hat{\theta}_1 y(k-1) + \hat{\theta}_2 y(k-2) + \hat{\theta}_3 + \hat{\theta}_4 u(k-1)^3 \\ &\quad + \hat{\theta}_5 y(k-3) + \hat{\theta}_6 u(k-1)^2 u(k-3) + \hat{\theta}_7 u(k-3)^3 \\ &\quad + \hat{\theta}_8 u(k-1)u(k-3) + \hat{\theta}_9 u(k-1)u(k-3)^2. \end{aligned} \tag{23}$$

Assume the static function $f(\cdot)$ of the system is known or is measured at n_{sf} points. Suppose that model (23) is to be fit to a set of dynamical data $[u(k), y(k)]_{k=1}^{N}$ and should simultaneously approximate the system static function, represented by the n_{sf} pairs $(\bar{u}_i, \bar{y}_i)$. Applying Theorem 1, using J_{LS} (4) and J_{SF} (4), the parameter vector can be estimated by

$$\hat{\theta} = [w_1\Psi^{\mathrm{T}}\Psi + w_2(QR)^{\mathrm{T}}(QR)]^{-1}[w_1\Psi^{\mathrm{T}}\mathbf{y} + w_2(QR)^{\mathrm{T}}\bar{\mathbf{y}}],$$

for any values of w_1 and w_2 such that $w_1, w_2 \geq 0$ and $w_1 + w_2 = 1$. It is instructive to notice that in this example $\mathbf{q}_i^{\mathrm{T}} = \begin{bmatrix} 1 & \bar{y}_i & \bar{u}_i^2 & \bar{u}_i^3 \end{bmatrix}$ and

$$R = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

3.3 *Static gain*

Consider now that, in addition to the information about the location of some points of the system static curve, the information about the static gain at the same points is also available. This means that triples $(\bar{u}_i, \bar{y}_i, \bar{g}_i)$ are available, with

$$\bar{g}_i = \left.\frac{\mathrm{d}\bar{y}}{\mathrm{d}\bar{u}}\right|_{(\bar{u}_i, \bar{y}_i)}. \tag{24}$$

The equation $\bar{y} = f(\bar{y}, \bar{u})$ implicitly defines $\bar{y}$ as a function of $\bar{u}$. Taking the derivative of such function leads to

$$\frac{\mathrm{d}\bar{y}}{\mathrm{d}\bar{u}} = -\frac{\partial[f(\bar{y}, \bar{u}) - \bar{y}]}{\partial\bar{u}}\left(\frac{\partial[f(\bar{y}, \bar{u}) - \bar{y}]}{\partial\bar{y}}\right)^{-1}. \tag{25}$$

Rearranging this equation and instantiating the variables for the available triples $(\bar{u}_i, \bar{y}_i, \bar{g}_i)$ leads to

$$\bar{g}_i\left.\frac{\partial f}{\partial\bar{y}}\right|_{(\bar{y}_i, \bar{u}_i)} - \bar{g}_i = -\left.\frac{\partial f}{\partial\bar{u}}\right|_{(\bar{y}_i, \bar{u}_i)}. \tag{26}$$

From equation (19) comes

$$\frac{\partial f}{\partial\bar{y}} = \frac{\partial\mathbf{q}^{\mathrm{T}}}{\partial\bar{y}}R\boldsymbol{\theta}, \quad \text{and} \quad \frac{\partial f}{\partial\bar{u}} = \frac{\partial\mathbf{q}^{\mathrm{T}}}{\partial\bar{u}}R\boldsymbol{\theta}. \tag{27}$$

Considering equation (20), define matrices Ω_i and Γ_i as

$$\Gamma_i \triangleq \left.\frac{\partial\mathbf{q}}{\partial\bar{y}}\right|_{(\bar{u}_i, \bar{y}_i)} = \begin{bmatrix} 0 & 1 & 2\bar{y}_i & \dots & \ell\bar{y}_i^{\ell-1} & 0 & 0 & \dots & 0 & \dfrac{\partial F_{yu}}{\partial\bar{y}} \end{bmatrix}^{\mathrm{T}},$$

$$\Omega_i \triangleq \left.\frac{\partial\mathbf{q}}{\partial\bar{u}}\right|_{(\bar{u}_i, \bar{y}_i)} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 1 & 2\bar{u}_i & \dots & \ell\bar{u}_i^{\ell-1} & \dfrac{\partial F_{yu}}{\partial\bar{u}} \end{bmatrix}^{\mathrm{T}}.$$

Finally, from equations (26) and (27), comes

$$\left(\bar{g}_i\Gamma_i^{\mathrm{T}} + \Omega_i^{\mathrm{T}}\right)R\boldsymbol{\theta} = \bar{g}_i. \tag{28}$$

Considering that n_{sg} triples $(\bar{u}_i, \bar{y}_i, \bar{g}_i)$ are available, equation (28) can be expressed as $HR\theta = \bar{\mathbf{g}}$, where

$$H = \begin{bmatrix} \bar{g}_1\Gamma_1^{\mathrm{T}} + \Omega_1^{\mathrm{T}} \\ \bar{g}_2\Gamma_2^{\mathrm{T}} + \Omega_2^{\mathrm{T}} \\ \vdots \\ \bar{g}_{n_{sg}}\Gamma_{n_{sg}}^{\mathrm{T}} + \Omega_{n_{sg}}^{\mathrm{T}} \end{bmatrix}. \tag{29}$$

Therefore, a static gain cost functional can be written as

$$J_{\mathrm{SG}} = (\bar{\mathbf{g}} - HR\hat{\theta})^{\mathrm{T}}(\bar{\mathbf{g}} - HR\hat{\theta}), \tag{30}$$

where $[\bar{\mathbf{g}}, HR]$ is an affine information.

One should notice that higher order derivatives of the static function could give rise to other affine information pairs using the same procedure that was employed here. However, these higher-order measurements are more difficult to obtain in practice.

Example 5: The steady-state relation between output $y(k)$ and input $u(k)$ of model (23) can be written as

$$\bar{y} = \Sigma_0 + \Sigma_y\bar{y} + \Sigma_{u^2}\bar{u}^2 + \Sigma_{u^3}\bar{u}^3. \tag{31}$$

The cluster coefficients of model (23) are $\Sigma_{u^3} = \theta_4 + \theta_6 + \theta_7 + \theta_9$; $\Sigma_{u^2} = \theta_8$; $\Sigma_y = \theta_1 + \theta_2 + \theta_5$ and $\Sigma_0 = \theta_3$. Using (28) yields $(\bar{g}_i[0 \quad 1 \quad 0 \quad 0]^{\mathrm{T}} + \begin{bmatrix} 0 & 0 & 2\bar{u}_i & 3\bar{u}_i^2 \end{bmatrix}^{\mathrm{T}})R\boldsymbol{\theta} = \bar{g}_i$ and

$$\hat{\theta}^* = \left[w_1\Psi^{\mathrm{T}}\Psi + w_2(HR)^{\mathrm{T}}(HR)\right]^{-1}\left[w_1\Psi^{\mathrm{T}}\mathbf{y} + w_2(HR)^{\mathrm{T}}\bar{\mathbf{g}}\right]$$

is obtained applying Theorem 1, with J_{LS} and J_{SG}.

4. Results

An example is considered, of a pilot buck dc-dc converter. The affine information pair to be used is $[\bar{\mathbf{y}}, QR]$

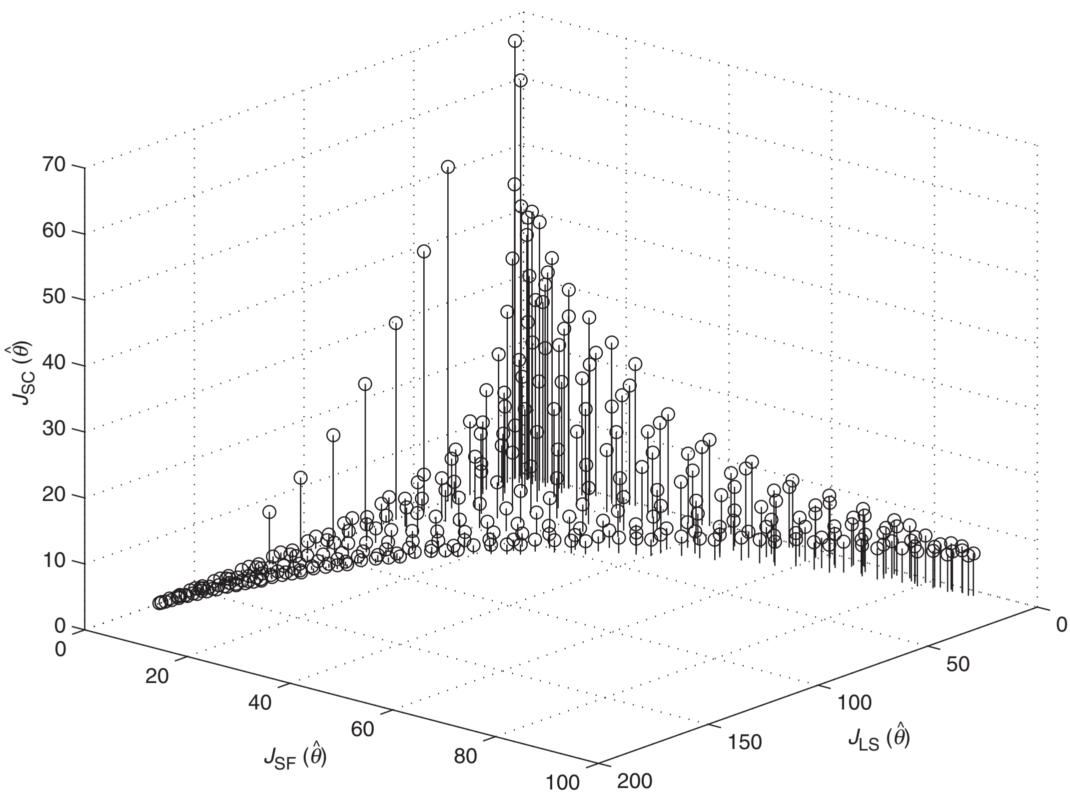

Figure 1. Pareto-set solutions for (33).

and $[\bar{\mathbf{g}}, HR]$. Because this paper concerns parameter estimation, the relevant problem of structure selection is assumed to have been previously tackled by some adequate method (Billings *et al.* 1989, Piroddi and Spinelli 2003). It should be noticed, however, that the model structure should be "consistent" with the type of information to be used. In the words of Definition 1, $G\hat{\theta}$ should be an estimate of $\mathbf{v}$.

The main goal of this example is to estimate the parameters of a dc-dc pilot converter based on a MOSFET IRF840 using three sources of information: dynamical data (measured), static function and static gain (known *a priori* from the theory of power electronics). The model with parameter estimated this way is best, from an overall point of view.

The duty cycle is defined by $\mathrm{D} = T_{\mathrm{on}}/T$ and its complement is $\mathrm{D}' = T_{\mathrm{off}}/T$, where T is the operation cycle. The load voltage V_{o} relates to the source voltage V_{d} as $V_{\mathrm{o}} = \mathrm{D}V_{\mathrm{d}} = (1-\mathrm{D}')V_{\mathrm{d}}$. This converter satisfies $\mathrm{D}' = (\bar{u}-1)/3$ and therefore the static function of this system is known from theory to be

$$V_{\mathrm{o}} = \bar{y} = \frac{4V_{\mathrm{d}}}{3} - \frac{V_{\mathrm{d}}}{3}\bar{u}, \tag{32}$$

where $u(k)$ is the command voltage that determines the duty cycle. The auxiliary information about the static function can be obtained by equation (32) varying values for $\bar{u}$ within $1 < \bar{u} < 4$. The dynamic data was obtained by means of a pseudo random signal (PRBS) (Aguirre *et al.* 2000). The model structure for this system was obtained using the error reduction ratio (ERR) criterion (Billings *et al.* 1989) and is the one considered in Example 4, see structure (23).

Information about static function (22) and static gain (30), together with the usual input/output data (4), have been used as sources of affine information to estimate the parameters of the model (23). The composite cost function is

$$\Upsilon(\hat{\boldsymbol{\theta}}) = w_1 J_{\mathrm{LS}}(\hat{\boldsymbol{\theta}}) + w_2 J_{\mathrm{SF}}(\hat{\boldsymbol{\theta}}) + w_3 J_{\mathrm{SG}}(\hat{\boldsymbol{\theta}}). \tag{33}$$

Each solution (model) in figure 1 has one or two cost function values smaller than any other solution, see table 1 for details. A simple way to choose a single "best" solution, would require defining a criterion such as the solution the minimal Euclidean norm $\|\cdot\|_2$. With $\|J(\hat{\theta})_*\|_2 = 18.14$, the best model is found using $w_1 = 0.25$, $w_2 = 0.25$ and $w_3 = 0.50$, which can be interpreted as a "best compromise" among the objectives.

Table 1. Performance of Pareto-models for the pilot buck dc-dc converter.

w_1	w_2	w_3	$J(\hat{\theta})_{LS}$	$J(\hat{\theta})_{SF}$	$J(\hat{\theta})_{SG}$	$\lVert J(\hat{\theta})_* \rVert_2$
0.98	0.01	0.01	2.19	0.63	56.23	56.27
0.70	0.10	0.20	3.03	5.64	27.89	28.62
0.50	0.30	0.20	3.70	3.02	28.09	28.50
0.35	0.15	0.50	7.81	12.08	11.94	18.69
0.25	0.25	0.50	14.21	5.95	9.59	18.14
0.01	0.98	0.01	26.43	2.35	44.82	52.08
0.15	0.35	0.50	29.83	17.56	19.91	39.93
0.01	0.01	0.98	129.65	23.48	11.26	132.24

5. Conclusion

This work has proposed the definition of affine information. The well-known one-step-ahead prediction error that is usually employed in standard identification problems has been shown to be a special case of such information. It was shown that a multiobjective parameter estimation problem, based on a set of affine information pairs, leads to a convex multiobjective optimization problem. An LS-type non-iterative scheme for finding the Pareto-set solutions for this generic problem has also been proposed in this work. The results hold for any linear-in-the-parameter model structure. The extension to models with MA (moving average) terms is direct because all the constraints discussed in the paper only act on the NARX part of the models.

Three instances of auxiliary information for NARX polynomial models were developed and illustrated, namely fixed-points location, static function and steady state gain. These were shown to fit in the definition of affine information. The main ideas of the paper were illustrated using a pilot buck dc-dc converter. For this converter, the static function is known from the theory and therefore it qualifies as high quality auxiliary information. In fact, the use of such information significantly improves the steady-state performance of the identified models especially when such information is, for some reason, blurred in the dynamical data.

The use of three objectives simultaneously enhances flexibility so that it becomes easier to find a more appropriate model for a specific application. In this case, the static gain increases the quality of the model.

It is believed that an important contribution of this paper is the proposition of a general methodology for dealing with a broad class of auxiliary information, employing a formulation that is an extension of the long-standing standard least squares estimator. Besides, such a methodology yields not a single model, but rather a set of Pareto-models that convey much information about the system overall behaviour. Future research should consider the interpretation of the Pareto-set and the definition of other affine information pairs not only for polynomial NARX models but also for other linear-in-the-parameter model representations.

References

L.A. Aguirre and S.A. Billings, "Improved structure selection for nonlinear models based on term clustering", *Inter. J. Cont.*, 62, pp. 569–587, 1995.

L.A. Aguirre, P.F. Donoso-Garcia and R. Santos-Filho, "Use of *a priori* information in the identification of global nonlinear model-a case study using a buck converter", *IEEE Trans. on Circuits and Syst. Part I: Fundamental Theory and Appl.*, 47, pp. 1081–1085, 2000.

L.A. Aguirre and E.M.A.M. Mendes, "Global nonlinear polynomial models: structure, term cluster and fixed points.", *Inter. J. Bifurcation and Chaos*, 6, pp. 279–294, 1996.

S.A. Billings,, S. Chen and M.J. Korenberg, "Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator", *Inter. J. Cont.*, 49, pp. 2157–2189, 1989.

V. Chankong and Y.Y. Haimes, *Multiobjective decision making: theory and methodology*, New York: North-Holland (Elsevier), 1983.

M.V. Corrêa, L.A. Aguirre and R.R. Saldanha, "Using prior knowledge to constrain parameter estimation in nonlinear system identification", *IEEE Trans. on Circuits and Syst. Part I: Fundamental Theory and Appl.*, 49, pp. 1376–1381, 2002.

E. Eskinat, S.H. Johnson and W.L. Luyben, "Use of auxiliary information in system identification", *Industrial & Engineering Chemistry Research*, 32, pp. 1981–1992, 1993.

T. Johansen, "Identification of non-linear systems using empirical data and prior knowledge – an optimization approach", *Automatica*, 32, pp. 337–356, 1996.

T.A. Johansen, "Multi-objective identification of FIR models", *Proceedings of 12th IFAC Symposium on System Identification 2000*, Santa Barbara, USA, pp. 917–922, 2000.

I.J. Leontaritis and S.A. Billings, "Input–output parametric models for non-linear systems – part i: deterministic non–linear systems", *Inter. J. Cont.*, 41, pp. 303–328, 1985.

E.G. Nepomuceno, R.H.C. Takahashi, G.F.V. Amaral and L.A. Aguirre, "Nonlinear identification using prior knowledge of fixed points: a multiobjective approach", *Inter. J. Bifurcation and Chaos*, 25, pp. 1229–1246, 2003.

R.K. Pearson and M. Pottmann, "Gray-box identification of block-oriented nonlinear models", *J. Process Cont.*, 10, pp. 301–315, 2000.

L. Piroddi and W. Spinelli, "An identification algorithm for polynomial NARX models based on simulation error minimization", *Inter. J. Cont.*, 76, pp. 1767–1781, 2003.

J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson and A. Juditsky, "Nonlinear black-box modeling in system identification: A unified overview", *Automatica*, 31, p. 1691, 1995.

T. Söderström and P. Stoica, "*System Identification*", London: Prentice-Hall, 1989.

H.J.A.F. Tulleken, "Grey-box modelling and identification using physical knowledge and Bayesian techniques", *Automatica*, 29, pp. 285–308, 1993.