Regularization aspects in continuous-time model identification

Saïd Moussaoui, David Brie, Alain Richard

Regularization aspects in continuous-time model identification

Saïd Moussaoui*, David Brie, Alain Richard

Université Henri Poincaré, Nancy 1, Centre de Recherche en Automatique de Nancy (CRAN), UMR 7039 CNRS-UHP-INPL, B.P. 239, F-54506 Vandœuvre-lès-Nancy Cedex, France

This paper presents an analysis of some regularization aspects in continuous-time model identification. The study particulary focuses on linear filter methods and shows that filtering the data before estimating their derivatives corresponds to a regularized signal derivative estimation by minimizing a compound criterion whose expression is given explicitly. A new structure based on a null phase filter corresponding to a true regularization filter is proposed and allows to discuss the filter phase effects on parameter estimation by comparing its performances with those of the Poisson filter-based methods. Based on this analysis, a formulation of continuous-time model identification as a joint system input–output signal and model parameter estimation is suggested. In this framework, two linear filter methods are interpreted and a compound criterion is proposed in which the regularization is ensured by a model fitting measure, resulting in a new regularization filter structure for signal estimation.

Keywords: System identification; Continuous-time model; Derivative estimation; Regularization

1. Introduction

System identification and, in particular, direct approaches for identifying a system represented by continuous-time models has been the subject of many works—see for example the surveys of Young (1981), Unbehauen and Rao (1990), Unbehauen and Rao (1998), Nielson, Madson, and Young (2000), Young, Pedregal, and Tych (1999) and the references therein. Several methods were developed, allowing to reach a good level of performances and even better than an indirect approach (Rao & Garnier, 2002; Ljung, 2003). Most of the available algorithms are gathered into the Matlab toolboxes CAPTAIN[1] (Young, 2002) and CONTSID[2] (Garnier, Gilson, & Huselstein, 2003a). A comparative performance evaluation of seventeen different continuous-time identification methods and their numerical implementation issues has been performed in Mensler (1999), Garnier, Mensler, and Richard (2003b) from which it turns out that the linear filter methods present good estimation performances. This motivates the focus of this paper on these methods.

The main difficulty in continuous-time model identification, is the need to estimate, from the measured data the non-measurable time derivatives of the system input and output signals before evaluating the model parameters by a parametric estimation method. A naïve approximation of these time derivatives by usual numerical differentiation methods causes a noise amplification, that will affect model parameter estimation. To handle this problem, a first possible approach, including linear filter methods, consists in applying a linear transformation to the input and output signals in order to avoid an explicit calculation of the measured signal

* Corresponding author.
E-mail address: said.moussaoui@cran.uhp-nancy.fr (S. Moussaoui).

doi:10.1016/j.automatica.2004.10.008

[1] see http://www.es.lancs.ac.uk/cres/captain/.
[2] see http://www.cran.uhp-nancy.fr/contsid/.

derivatives. The model parameters are then estimated using the time derivatives calculated from these transformed signals. Another approach consists in developing methods that directly give estimates of the time derivatives. In particular, Söderström, Carlsson, and Bigi (1997) presented an analysis of the effect of the derivative approximation by standard finite difference methods on the bias and variance of the parameter estimators and, to reduce them, proposed some particular numerical differentiators. The theoretical formulation of time-derivative estimation as an ill-posed inverse problem has been addressed in many papers (Cullum, 1971; Tikhonov & Arsenin, 1977; Surova, 1979; Jakeman & Young, 1984) and Young and Foster (1993) applied the fixed interval smoothing approach (FIS)—which is equivalent to using explicitly a regularization technique (Young & Pedregal, 1999)—to continuous-time model identification. This work aims at linking these apparently distinct classes of methods by interpreting the linear filter transformation as a regularization filter for derivative estimation.

This paper is organized as follows: Section 2 recalls the concepts of ill-posed inverse problem, regularization and addresses the formulation of derivative estimation as an ill-posed inverse problem and gives its Tikhonov regularized estimate. In Section 3 the principle of continuous-time model identification by linear filter methods is presented. Then, Section 4 provides a link between this method and regularization techniques by formalizing the regularization properties of the linear filters. From this interpretation, Section 5 proposes a new structure of filter that allows to give some insights into the filter phase effect on model parameter estimation. Finally, Section 6 addresses the problem of the optimal filter design by formulating the continuous-time model identification problem as a joint system input–output signal and model parameter estimation.

2. Inverse problem and regularization

To illustrate these concepts, let us consider a dynamic system with an input signal $x_0(t)$, output signal $y(t)$ and some disturbances modelled by an additive noise $v(t)$. The direct problem consists in defining the model that expresses the relation between these signals

$$y(t) = \mathscr{H}[x_0](t) + v(t), \tag{1}$$

where $\mathscr{H}$ is an operator representing the transformation induced by the system to the input signal $x_0(t)$. In the case of linear time-invariant system of impulse response $h(t)$ (case considered in this study), $\mathscr{H}$ is a convolution operator, i.e. $\mathscr{H}[x_0](t) = [x_0 \star h](t)$. Two inverse problems are associated to this formulation (Fig. 1). The first is the determination of $\mathscr{H}$ from $x_0(t)$ and $y(t)$ (identification) and the second is the estimation of $x_0(t)$ from $\mathscr{H}$ and $y(t)$ (deconvolution). This section only considers the problem of deconvolution. An inverse problem is said to be ill-posed, if the solution does not exist, or is not unique or if a small disturbance on data

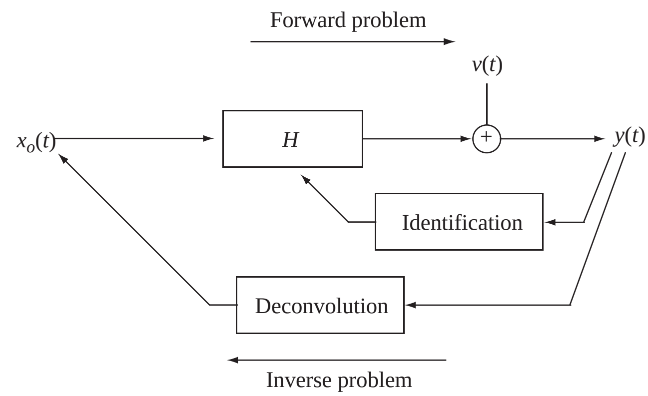

Fig. 1. Direct and inverse problems.

induces a large variation of the solution (Hadamard, 1923). The regularization aims at solving an ill-posed problem by searching a solution meaningful and stable with respect to data variations.

Earlier works on the explicit use of regularization techniques in system identification can be found in Sjöberg, McKelvey, and Ljung (1993), Young and Foster (1993) and more recently Johansen (1996), Johansen (1997), Ninness and Henriksen (2003). A tutorial on inverse problems in control has been presented in the 15th IFAC world congress (Goodwin, 2002).

2.1. Tikhonov regularization

A particular method of regularization consists in minimizing a compound criterion. Instead of minimizing only a data fitting measure, a new term is added in order to make the solution faithful to some a priori knowledge or to some specified constraints. Tikhonov regularization (Tikhonov & Arsenin, 1977) is one of the methods suggested from this point of view. This technique tends to minimize a compound criterion given by

$$J(x, y) = \|y - \mathscr{H}[x]\|^2 + \sum_{d=0}^{p} \alpha_d \|\mathscr{D}_d[x]\|^2, \tag{2}$$

where $\{\alpha_d\}_{d=0}^{p}$ are constant regularization parameters, $\mathscr{D}_d$ the dth time-derivative operator and $\|\cdot\|$ the $\mathscr{L}_2$ norm. The Tikhonov stabilizers of order p used in the second part of criterion (2) imposes to the solution to be the smoothest in the class of p times derivable functions. The resulting regularized solution is given by

$$\hat{x}_{\text{reg}}(t) = \arg\min_x J(x, y) \tag{3}$$

and leads to

$$\hat{x}_{\text{reg}}(t) = \frac{\mathscr{H}^*}{\mathscr{H}^*\mathscr{H} + \sum_{d=0}^{p} \alpha_d \mathscr{D}_d^* \mathscr{D}_d}[y](t), \tag{4}$$

where $\mathscr{H}^*$ and $\mathscr{D}_d^*$ are the adjoint operators of $\mathscr{H}$ and $\mathscr{D}_d$, respectively.

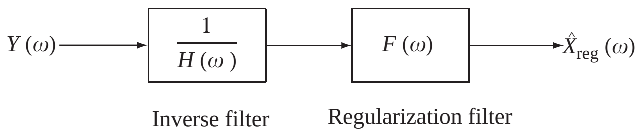

Fig. 2. Inverse filtering and regularization.

2.2. Frequency-domain interpretation

By applying the Fourier transform to (4), the regularized solution expressed in the frequency domain is obtained

$$\hat{X}_{\text{reg}}(\omega) = \frac{H^*(\omega)}{H^*(\omega)H(\omega) + \sum_{d=0}^{p}\alpha_d\omega^{2d}}\mathrm{Y}(\omega). \tag{5}$$

This solution corresponds to applying a regularization filter to the naïve solution obtained by inverse filtering (Fig. 2) or equivalently filtering the output signal before inverse filtering. Therefore, the regularization filter is expressed by

$$F(\omega) = \frac{H^*(\omega)H(\omega)}{H^*(\omega)H(\omega) + \sum_{d=0}^{p}\alpha_d\omega^{2d}}. \tag{6}$$

As introduced, a particular example of ill-posed inverse problem is the nth derivative estimation where the corresponding operator $\mathscr{H}$ is an nth order integrator. In this case, according to (6) the expression of the Tikhonov regularization filter is

$$\begin{aligned} F(\omega) &= \frac{(1/\omega)^{2n}}{(1/\omega)^{2n} + \sum_{k=0}^{p}\alpha_d\omega^{2d}} \\ &= \frac{1}{1 + \sum_{d=n}^{p+n}\alpha_{d-n}\omega^{2d}}. \end{aligned} \tag{7}$$

This is a low-pass filter which attenuates the high-frequency part where the noise amplification problem appears. Note that both its shape and bandwidth depend on the regularization parameters $\{\alpha_d\}_{d=0}^{p}$.

2.3. Bayesian formulation

In this section, a Bayesian interpretation of Tikhonov regularization criterion is given. A more general discussion about the synthesis of regularization criteria can be found in Jakeman and Young (1984), Demoment (1989), Young et al. (1999), and Idier (2001). Note that the probabilistic formulation imposes to consider discrete-time data, however the interpretation still remains valid for the continuous-time case. The model defined in (1) is considered and the noise signal sequence $\{v(t_k)\}_{k=1}^{N}$ is assumed independent and identically distributed (i.i.d.), zero mean and Gaussian with variance σ_v^2, i.e. $v(t_k) \sim \mathscr{N}(0, \sigma_v^2)$. The likelihood function associated to the problem of estimating the signal $x(t_k)$ is then expressed as

$$\begin{aligned} &p(\boldsymbol{y} = \{y(t_k)\}_{k=1}^{N} \,|\, \boldsymbol{x} = \{x(t_k)\}_{k=1}^{N}, \sigma_v^2) \\ &\quad = \prod_{k=1}^{N}\left(\frac{1}{2\pi\sigma_v^2}\right)^{1/2} \exp\left[-\frac{[y(t_k) - \mathscr{H}[x](t_k)]^2}{2\sigma_v^2}\right]. \end{aligned} \tag{8}$$

The dth time derivative of the input signal x_o is assumed i.i.d. zero mean and Gaussian with variance σ_d^2, defines the prior probability density function (pdf) of $\boldsymbol{x}$ as

$$p(\boldsymbol{x} \,|\, \sigma_d^2) = \prod_{k=1}^{N}\left(\frac{1}{2\pi\sigma_d^2}\right)^{1/2} \exp\left[-\frac{[\mathscr{D}_d[x](t_k)]^2}{2\sigma_d^2}\right]. \tag{9}$$

Using Bayes' theorem

$$p(\boldsymbol{x} \,|\, \boldsymbol{y}, \sigma_v^2, \sigma_d^2) = \frac{p(\boldsymbol{y} \,|\, \boldsymbol{x}, \sigma_v^2)\,p(\boldsymbol{x} \,|\, \sigma_d^2)}{p(\boldsymbol{y})}, \tag{10}$$

where $p(\boldsymbol{x} \,|\, \boldsymbol{y}, \sigma_v^2, \sigma_d^2)$ is the posterior pdf of $\boldsymbol{x}$ and $p(\boldsymbol{y})$ is a normalization constant. Using the proportionality symbol $\propto$, to omit all the constants, one can write

$$\begin{aligned} &p(\boldsymbol{x} \,|\, \boldsymbol{y}, \sigma_v^2, \sigma_d^2) \\ &\quad \propto p(\boldsymbol{y} \,|\, \boldsymbol{x}, \sigma_v^2)\,p(\boldsymbol{x} \,|\, \sigma_d^2), \\ &\quad \propto \prod_{k=1}^{N} \exp\left[-\frac{[y(t_k) - \mathscr{H}[x](t_k)]^2}{2\sigma_v^2} - \frac{[\mathscr{D}_d[x](t_k)]^2}{2\sigma_d^2}\right], \\ &\quad \propto \exp\left[-\sum_{k=1}^{N}\left(\frac{[y(t_k) - \mathscr{H}[x](t_k)]^2}{2\sigma_v^2} + \frac{[\mathscr{D}_d[x](t_k)]^2}{2\sigma_d^2}\right)\right]. \end{aligned}$$

The maximum a posteriori (MAP) estimate of the input signal corresponds to the value of $\boldsymbol{x}$ that maximizes this posterior pdf $p(\boldsymbol{x} \,|\, \boldsymbol{y}, \sigma_v^2, \sigma_d^2)$ or equivalently, minimizes the resulting criterion $J(x, y)$ obtained from its inverse logarithm

$$\begin{aligned} J(x, y) &= -\log p(\boldsymbol{x} \,|\, \boldsymbol{y}, \sigma_v^2, \sigma_d^2) \\ &= \frac{1}{2\sigma_v^2}\|y - \mathscr{H}[x]\|^2 + \frac{1}{2\sigma_d^2}\|\mathscr{D}_d[x]\|^2, \end{aligned} \tag{11}$$

where the ℓ_2 norm notation is used. The criterion can be rearranged in the form

$$J(x, y) = \|y - \mathscr{H}[x]\|^2 + \alpha_d\|\mathscr{D}_d[x]\|^2, \tag{12}$$

where $\alpha_d = (\sigma_v^2/\sigma_d^2)$ which corresponds to the optimal of the regularization parameter, as mentioned by Jakeman and Young (1984). A more general form of this criterion takes into account all the derivatives of the input signal up to a maximal order p, resulting in criterion (2).

2.4. Implementation issues

To optimize criterion (2), several techniques are possible. Direct batch implementation, after discretization of the problem (Phillips, 1962; Twomey, 1963), yields an explicit solution but requires the inversion of a matrix of size equivalent to the data length, so this method is not used for large data sets. To avoid this direct large matrix inversion, iterative optimization by a gradient, or Newton-based algorithms may be used as well as a frequency-domain implementation proposed by Hunt (1973), for the case of periodic signals. Another alternative for the inversion consists in a recursive implementation using Kalman filtering techniques (Mendel, 1983). Because these methods are causal, a time lag is introduced on the solution. By adding a backward smoothing, the fixed interval smoothing (FIS) method (Jakeman & Young, 1984) is an en-bloc approach which produces a zero-lag solution. All these optimization procedures will produce almost the same regularized estimates but practical considerations, such as computational burden and real-time implementation, will play an important role in the final selection of a particular method.

The methods described above require the knowledge of the regularization hyperparameters, which are concatenated in a vector $\boldsymbol{\theta} = [\sigma_v^2, \{\sigma_d^2\}_{d=0}^{p}]^{\mathrm{T}}$. The maximum likelihood estimator may be applied to assess the hyper-parameters from estimates during a recursive/iterative optimization method (Young et al., 1999). They can also be estimated in a Bayesian framework by a joint maximization of the posterior $p(\boldsymbol{x}, \boldsymbol{\theta} \,|\, \boldsymbol{y})$ with respect to $\boldsymbol{x}$ and $\boldsymbol{\theta}$ by assigning appropriate prior distributions to the hyperparameters (Mohammad-Djafari, 1996).

3. Continuous-time model identification

3.1. Problem statement

Let us now consider a system represented by a continuous-time, time-invariant model, linear, stable and causal whose input–output relationship is given by

$$\sum_{i=0}^{na} a_i y_{\mathrm{o}}^{(i)}(t) = \sum_{i=0}^{nb} b_i u_{\mathrm{o}}^{(i)}(t), \tag{13}$$

where

- $u_{\mathrm{o}}(t)$ and $y_{\mathrm{o}}(t)$ are, respectively, the noise-free input and output signals;
- $x^{(i)}(t)$ is the ith time derivative of $x(t)$. The initial conditions are supposed to be null;
- $\{a_i\}_{i=0}^{na}$ and $\{b_j\}_{j=0}^{nb}$ are the model parameters, ($na \geqslant nb$ and $a_{na} = 1$).

The input and output signals are sampled at a constant frequency $f_s = T_s^{-1}$. The available data used for the identification are noted $\{u(t_k), y(t_k)\}_{k=0}^{N-1}$, with $t_k = kT_s$, and represent the samples of the measured system of input and output signals. The output error model is considered which assumes that only the measured output signal is prone to disturbances, modeled by an additive noise $v(t)$, independent of the input signal $u_{\mathrm{o}}(t)$, that is $y(t_k) = y_{\mathrm{o}}(t_k) + v(t_k)$ while

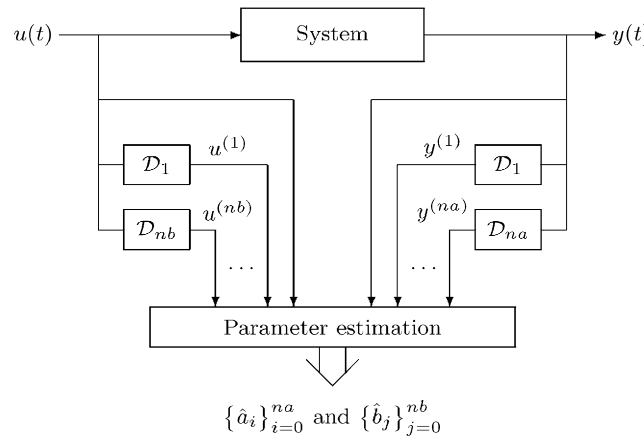

Fig. 3. Continuous-time model identification problem.

the input signal is noise-free, that is $u(t_k) = u_{\mathrm{o}}(t_k)$. The case of error-in-variables model, where both input and output signals are prone to disturbances, will be discussed in Section 6. The problem of continuous-time model identification may be stated as follows: knowing the orders *na*, *nb* and having the samples $\{u(t_k), y(t_k)\}_{k=0}^{N-1}$,determine the coefficients $\{a_i\}_{i=0}^{na-1}$ and $\{b_j\}_{j=0}^{nb}$ of the differential equation (13).

To estimate the model parameters, the differential equation should be evaluated at each time $\{t_k\}_{k=0}^{N-1}$ and written as a linear regression model (Fig. 3) in order to apply a parametric estimation method. This operation requires an estimation of both input and output signal time derivatives. This is the main difficulty because of the noise amplification problem.

3.2. Linear filter methods

One of the solutions used to circumvent time-derivative estimation difficulty consists in applying a linear transformation to Eq. (13). This transformation corresponds to filtering by a linear filter of impulse response $f(t)$

$$\sum_{i=0}^{na} a_i \mathscr{T}[y_{\mathrm{o}}^{(i)}(t)] = \sum_{j=0}^{nb} b_j \mathscr{T}[u_{\mathrm{o}}^{(j)}(t)] + \mathscr{T}[v(t)], \tag{14}$$

where

$$\mathscr{T}[x^{(i)}(t)] = \left[f \star \frac{\mathrm{d}^i x}{\mathrm{d}t^i} \right](t). \tag{15}$$

The evaluation of $x_f^{(i)}(t_k) = \mathscr{T}[x^{(i)}(t_k)]$ is achieved by an adequate discretization technique (Garnier et al., 2003b) and the resulting linear regressor model resolution is carried out by a parametric estimation method (Fig. 4).

Many linear filtering methods were developed, only differing on the form of the filter. Within the framework of this paper, we will focus mainly a particular form of *state variable filter* (SVF) method which uses a cascade of identical first-order filters. This method originates in the works of Young (1964, 1965) and was introduced under the name

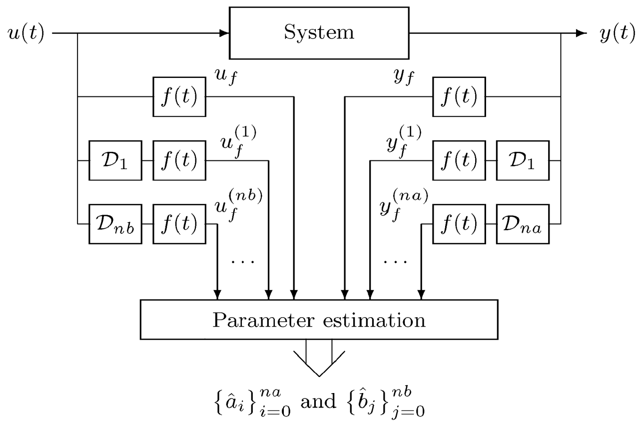

Fig. 4. Continuous-time model identification by linear filter methods.

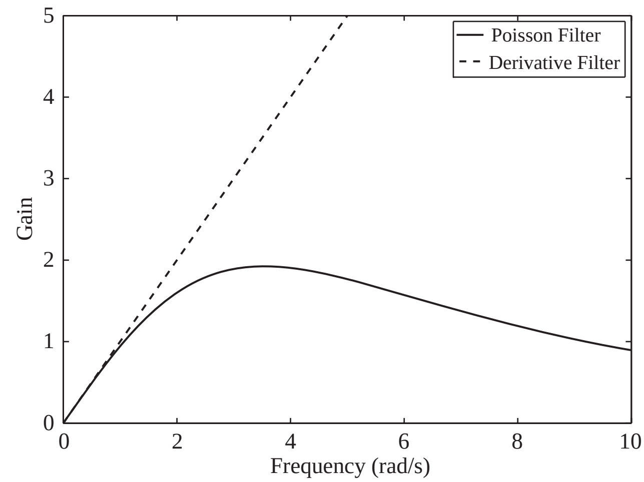

Fig. 5. First derivative Poisson moment functional filter, $\kappa = \lambda = 5\,\text{rad/s}$ and $l = 2$.

of MMF for *method of multiple filters*. It is based on the use of a filter in the form of

$$F_l(s) = \left(\frac{\kappa}{s+\lambda}\right)^{l+1}, \tag{16}$$

where s is the Laplace variable, κ and λ are design parameters of this filter. Its impulse response being

$$f_l(t) = \kappa^{(l+1)} \frac{t^l}{l!} \exp(-\lambda t) \tag{17}$$

which corresponds to the generalized Poisson pulse function, the method is also termed as GPMF for *generalized Poisson moment functionals* method (Saha & Rao, 1983). The choice of the SVF system representation is an exact system representation that allows to achieve a linear in-the-parameters formulation. Often the parameter κ is set equal to λ to get a unitary filter, but the main difficulty with this type of filter still remains the determination of the optimal value of the parameter λ.

To overcome that problem and to improve the statistical efficiency of parameter estimation, Jakeman and Young (1980) proposed to use a filter of the form

$$F(s) = \frac{1}{A(s)}, \tag{18}$$

where $A(s)$ is the denominator of the transfer function to identify. This filter structure is attractive, because when associated with an instrumental variable parameter estimation method, it yields a (quasi-) maximum likelihood optimality property in the case of additive white measurement noise (Young, 2002). In practice, the actual value of $A(s)$ is unknown and has to be replaced by an estimate $\hat{A}(s)$, resulting in an iterative method referred to as SRIVC for *simplified refined instrumental variable for continuous-time model*.

3.3. *Linear filter frequency response analysis*

The analysis of the frequency response of the filters used to estimate the Poisson moments of the successive derivatives, shows that they behave as derivators in the low-frequency part but attenuate the high-frequency band (Fig. 5). Thus, they have a regularizing effect which is studied in the next section.

4. Regularization aspects

This section first aims at formalizing the regularization aspects of the Poisson filter. Then the result is generalized to the SRIVC filter.

4.1. *First-order approximation of the Poisson filter*

Consider the form of the Poisson filter

$$F_l(\omega) = \left(\frac{\kappa/\lambda}{1+(1/\lambda)\mathrm{j}\omega}\right)^{l+1}. \tag{19}$$

By setting $\gamma = 1/\lambda$ and $\kappa = \lambda$, the expression of the filter becomes

$$F_l(\omega) = \left(\frac{1}{1+\gamma^2\omega^2}\right)^{(l+1)/2} \exp[-\mathrm{j}(l+1)\arctan(\gamma\omega)].$$

A first-order approximation of the filter phase around $\omega = 0$ (low-frequencies) gives

$$\arctan(\gamma\omega) = \gamma\omega + \mathcal{O}((\gamma\omega)^3) \tag{20}$$

and a series expansion of the filter gain yields

$$(1+\gamma^2\omega^2)^{(l+1)/2} = 1 + \sum_{m=1}^{M} S_{l+1}^m \gamma^{2m}\omega^{2m}, \tag{21}$$

where

$$M = \begin{cases} \frac{l+1}{2} & \text{if } (l+1) \text{ is even,} \\ \infty & \text{if } (l+1) \text{ is odd,} \end{cases}$$

and $S_{l+1}^m = \prod_{v=0}^{m-1}(\frac{(l+1)/2-v}{m!})$. Consequently, by noting $\beta_m(\lambda) = S_{l+1}^m \gamma^{2m}$, the expression of the filter becomes

$$F_l(\omega) \approx \frac{1}{1+\sum_{m=1}^{M}\beta_m(\lambda)\omega^{2m}} \exp[-\mathrm{j}(l+1)\gamma\omega]. \qquad (22)$$

4.2. Link with a regularization filter

By comparing (22) to the Tikhonov regularization filter (7), we conclude that the Poisson filter module exactly corresponds to a first derivative regularization filter, which minimizes the criterion

$$J_\lambda(x, y) = \|y - \mathscr{H}[x]\|^2 + \sum_{d=0}^{M-1} \alpha_d(\lambda)\|\mathscr{D}_d[x]\|^2, \qquad (23)$$

where

$$\mathscr{H}[x](t) = \int_0^t x(\tau)\,\mathrm{d}\tau,$$
$$\alpha_d(\lambda) = \beta_{d+1}(\lambda), \quad \text{for } d = 0, \ldots, M-1.$$

The smoothness constraint imposes that the required solution should be infinitely derivable (in the case where $(l+1)$ is even, the constraint is that the solution is $(M-1)$ time derivable). The regularization parameters depend explicitly on λ and the filter order l. The choice of this constraint is important because it ensures that choosing $l \geqslant na$, the estimation of the needed high order derivatives of the signals, using this regularized first derivative, will not yield a noise amplification.

Eq. (22) shows that the Poisson filter has approximately a linear phase in the low frequencies. This linear phase can be interpreted using the following theorem.

Theorem 1. *Consider the regularization filter whose frequency response is expressed by*

$$F(\omega, t_0) = \frac{|H(\omega)|^2}{|H(\omega)|^2 + \sum_{d=0}^{p}\alpha_d\omega^{2d}} \exp[-\mathrm{j}\omega t_0]. \qquad (24)$$

This filter corresponds to the minimization of a criterion

$$J(x, y, t_0) = \|y - \mathscr{H}[\tilde{x}]\|^2 + \sum_{d=0}^{p} \alpha_d\|\mathscr{D}_d[\tilde{x}]\|^2, \qquad (25)$$

where $\mathscr{H}$ is a convolution operator, $\tilde{x}(t) = x(t+t_0)$ and t_0 is the time delay introduced in the solution.

Proof 1. The application of the Fourier transform to Eq. (25) yields

$$\begin{aligned} J(X, Y, t_0) &= \|Y(\omega) - H(\omega)X(\omega)\exp[\mathrm{j}\omega t_0]\|^2 \\ &\quad + \sum_{d=0}^{p}\alpha_d\|(\mathrm{j}\omega)^d X(\omega)\exp[\mathrm{j}\omega t_0]\|^2, \\ &= \int \Big\{ |Y(\omega)|^2 - H(\omega)X(\omega)Y^*(\omega)\exp[\mathrm{j}\omega t_0] \\ &\quad - H^*(\omega)X^*(\omega)Y(\omega)\exp[-\mathrm{j}\omega t_0] \\ &\quad + |H(\omega)X(\omega)|^2 + \sum_{d=0}^{N}\alpha_d\omega^{2d}|X(\omega)|^2 \Big\}\,\mathrm{d}\omega. \end{aligned}$$

The minimization of $J(X, Y, t_0)$ with respect to $X(\omega)$ leads to

$$\forall\omega, \ \frac{\mathrm{d}}{\mathrm{d}X}J(X, Y, t_0)\bigg|_{X(\omega)=\hat{X}_{\mathrm{reg}}(\omega)} = 0,$$
$$\Rightarrow \hat{X}_{\mathrm{reg}}(\omega) = \frac{H^*(\omega)Y(\omega)}{|H(\omega)|^2 + \sum_{d=0}^{p}\alpha_d\omega^{2d}} \exp[-\mathrm{j}\omega t_0].$$

Recalling that $\hat{X}_{\mathrm{reg}}(\omega) = F(\omega, t_0)Y(\omega)/H(\omega)$, we obtain

$$F(\omega, t_0) = \frac{|H(\omega)|^2}{|H(\omega)|^2 + \sum_{d=0}^{p}\alpha_d\omega^{2k}} \exp[-\mathrm{j}\omega t_0], \qquad (26)$$

which corresponds to the regularization filter (24). □

From that result, it appears that applying a regularization filter with a linear phase corresponds to estimating a delayed solution. In the case of the Poisson filters, as shown in Fig. 6, this delay increases as the filter order increases or as the filter cut-off frequency decreases ($t_0 = (l+1)/\lambda$).

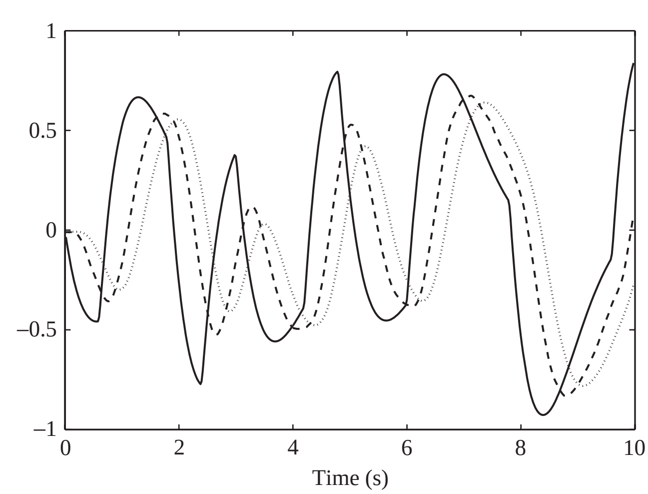

Fig. 6. Comparison of the exact derivative (full) and the filtered derivative by GPMF filter, $l = 2$ (dashed), $l = 5$ (dotted) with $\kappa = \lambda = 5\,\mathrm{rad/s}$.

4.3. Generalization

At the beginning of this section, we concentrated on the Poisson filter. The same analysis can be carried out for all linear filters of the following form

$$F(\omega) = \frac{1}{A(\omega)}, \tag{27}$$

where $A(\omega)$ is polynomial of order na. This filter may be written as

$$F(\omega) = \frac{1}{1 + \sum_{m=1}^{\infty} \beta_m \omega^{2m}} \exp[-\mathrm{j}\varphi_A(\omega)], \tag{28}$$

where the parameters β_m, and the phase $\varphi_A(\omega)$ depend on $A(\omega)$ polynomial coefficients. Structure (28) corresponds to the minimization of the criterion given by

$$J_A(x, y) = \|y - \mathscr{H}[\tilde{x}]\|^2 + \sum_{d=0}^{\infty} \alpha_d \|\mathscr{D}_d[\tilde{x}]\|^2, \tag{29}$$

where

$$\tilde{x}(t) = [x \star \mathscr{F}^{-1}\{\exp[\mathrm{j}\varphi_A(\omega)]\}](t),$$
$$\alpha_d = \beta_{d+1}, \quad \text{for } d = 0, \ldots, \infty.$$

and $\mathscr{F}^{-1}$ stands for the inverse Fourier transform operator.

5. Phase effect analysis

5.1. Phase analysis

From the previous section, it appears that linear filters only differ from true Tikhonov regularization filters by the phase term. Thus, the effect of the filter phase on continuous-time model identification needs to be discussed. In the framework of linear systems, if a linear phase filter is used, a time delay will be introduced on the filtered output signal. Consequently, to achieve parameter estimation correctly the same delay has to be introduced on the input signal. This explains the need of applying the same filter to both input and output signals even if the input signal is noise free. However, there is no objection to use instead a null phase filter and this has already been done for continuous-time model identification in Young and Foster (1993) where the derivatives are estimated using the fixed interval smoothing approach. Note that the use of such a filter is necessary in the case of nonlinear systems since the commutation operation is not possible (Young, 1993, Young, 1998).

We were not able to address the phase effect on parameter estimation using analytical developments. So, to get some insights into this phase effect, we perform some numerical simulations. In that respect, we propose to compare the performances of two filters having the same module and differing only by their phases: the first has a null phase while the second a non-linear phase. In the case of the SRIVC method, the implementation of the strictly equivalent null phase regularization filter is not possible, because of the infinite summation appearing in its module. However, the synthesis of the null phase filter having the same module as the Poisson filter is possible by taking

$$F_l^{\#}(\omega) = F_l(\mathrm{j}\omega)F_l(-\mathrm{j}\omega), \tag{30}$$

$$= \left(\frac{\kappa^2}{\omega^2 + \lambda^2}\right)^{l+1}. \tag{31}$$

Note that the module of this filter is polynomial on ω and of order $2(l+1)$, therefore the Poisson filter that has the same module is of order $(2l+1)$. Let us note $\gamma = 1/\lambda$, and consider $\kappa = \lambda$. By expanding (31), we get

$$F_l^{\#}(\omega) = \frac{1}{1 + \sum_{m=1}^{l+1} \beta_m(\lambda)\omega^{2m}}, \tag{32}$$

where $\beta_m(\lambda) = C_{l+1}^m \gamma^{2m}$. This form corresponds exactly to the Tikhonov regularization filter minimizing the criterion

$$J_\lambda(x, y) = \|y - \mathscr{H}[x]\|^2 + \sum_{d=0}^{l} \alpha_d(\lambda)\|\mathscr{D}_d[x]\|^2, \tag{33}$$

where the regularization parameters

$$\alpha_d(\lambda) = \beta_{d+1}(\lambda), \quad \text{for } d = 0, \ldots, l,$$

depend on λ and on the filter order. The smoothness constraint imposes a non-delayed (see Fig. 7) and l times derivable solutions.

5.2. Simulation example

A simulation example is used to assess the effect of the filter phase on continuous-time model identification.

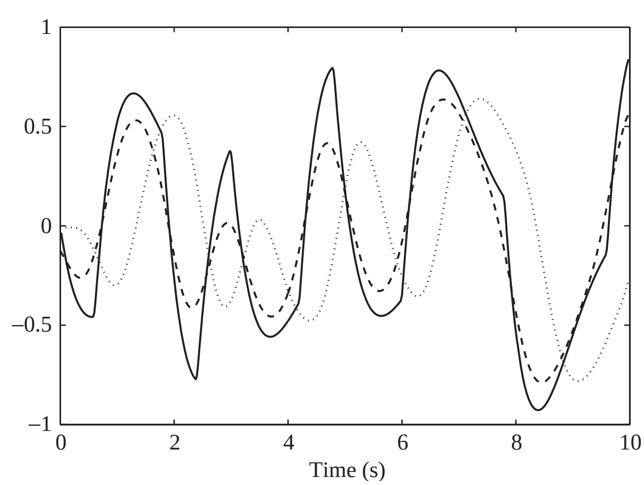

Fig. 7. Exact derivative (full) and regularized derivative (dashed) with $l = 2$ and filtered derivative by Poisson filter (dotted) with $l = 5$, for $\kappa = \lambda = 5$ rad/s.

The example concerns the identification of the following continuous-time transfer function (Wang & Gawthrop, 2001)

$$G(s) = \frac{-2s+1}{s^3+1.6s^2+1.6s+1}. \tag{34}$$

The input signal is a pseudo-random binary sequence of maximum length, chosen to excite the system over all its dynamic range. The sampling period is taken equal to 0.02 s and the number of samples is fixed to $N = 1260$. The results are obtained for a Monte Carlo simulation of $S = 1000$ trials, with a signal-to-noise ratio (SNR) equal to 5 dB. The SNR is defined by

$$\mathrm{SNR}\ (\mathrm{dB}) = 10\log\left(\frac{P_{y_o}}{\sigma_v^2}\right), \tag{35}$$

where P_{y_o} represents the power of the noise free output signal $y_o(t)$ and σ_v^2 is the variance of the additive noise. In order to have the same simulation conditions, the filters (Poisson and null phase filter) are applied to both input and output signals. The null phase filter is implemented using a forward–backward filtering and due to the initial/final conditions, the first and last samples corresponding to the setting time of the filter are removed. The parameters are estimated using the instrumental variable method with an auxiliary model (Young, 1970; Söderström & Stoica, 1989; Johansson, 1993) obtained after an initial estimation by least squares technique. Empirical mean ($\hat{m}_{\hat{\theta}_j}$), standard deviation ($\hat{\sigma}_{\hat{\theta}_j}$)and mean square error ($\widehat{\mathrm{MSE}}_{\hat{\theta}_j}$) evaluated for each parameter, are used to discuss the statistical performances of parameter estimation. The normalized mean square error (NMSE) is used to assess the global identification performances using the two filters. The NMSE is defined by

$$\mathrm{NMSE}\ (\mathrm{dB}) = 10\log\left(\frac{\sigma_\varepsilon^2}{\sigma_y^2}\right), \tag{36}$$

where σ_ε represents the standard deviation of the output error ($\varepsilon(t) = y(t) - \hat{y}_o(t)$), and σ_y^2 is the variance of the noisy output signal.

5.3. Discussion

Fig. 8 shows the evolution of the NMSE versus the filter bandwidth and the estimation results for the optimal value of λ, for both filters, are summarized in Table 1. The use of the null phase filter gives a slightly smaller mean square error than the Poisson filter, but the statistical performances of the two methods are very similar and both methods give asymptotically unbiased estimates. It is clear that no general rule may be inferred from that particular example, but this has been also verified on other simulation examples (Moussaoui, 2002). This point would need to be comforted by a theoretical analysis of the filter phase effect on parameter estimation. These methods have also been compared to the SRIVC method which provides the best results, confirming the very

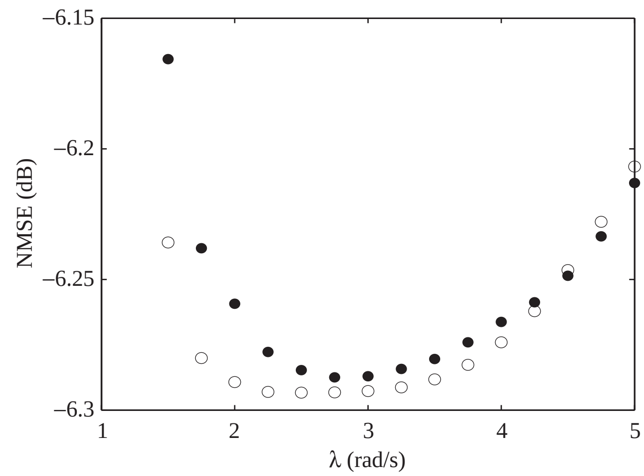

Fig. 8. NMSE versus filter bandwidth, null phase filter (circles) and Poisson filter (dots).

good behavior of this method as mentioned by Young (2002). These results are not reported here because only the phase effect is under investigation and the null phase filter, strictly equivalent to the SRIVC, has not been implemented.

As a conclusion, this experiment validates our interpretation of the Poisson filter as a delayed derivative estimation since the two filters yields nearly the same results. So, the main interest of the proposed approach is to give an explicit formulation of the filter design in terms of derivative estimation by minimizing a compound criterion that takes into account some prior information. Such a formulation may serve as a starting point to the design of new methods aiming at better accounting available knowledge coming from the characteristics of either the signals to restore or the system to identify.

6. Towards a joint signal and model parameter estimation

In the previous sections, an interpretation of regularization properties of the linear filters for signal derivative estimation has been given. However, the ultimate goal of continuous-time identification is to determine the model parameters. Based upon this interpretation, this problem may be stated as a joint signal and model parameter estimation from which the optimal regularization filter design for parameter estimation can be addressed. Linear filter methods (Poisson Filter and SRIVC) are discussed in this framework and a new formulation is proposed.

6.1. Input and output signal denoising

First of all, it should be noted that the derivative estimation using regularization filter corresponds to the denoising of the signal before applying the derivative operator. The denoised signal noted y_o is the solution that minimizes a

Table 1
Monte Carlo simulations results

	Poisson filter			Regularization filter		
λ_{opt} (rad/s)	2.75			2.5		
Parameters	$\hat{m}_{\hat{\theta}_j}$	$\hat{\sigma}_{\hat{\theta}_j}$	$\widehat{\text{MSE}}_{\hat{\theta}_j}$	$\hat{m}_{\hat{\theta}_j}$	$\hat{\sigma}_{\hat{\theta}_j}$	$\widehat{\text{MSE}}_{\hat{\theta}_j}$
$\hat{b}_0(1)$	1.0010	0.0602	0.0036	0.9980	0.0644	0.0042
$\hat{b}_1(-2)$	−2.0058	0.1161	0.0135	−2.0018	0.1105	0.0122
$\hat{a}_0(1)$	1.0016	0.0713	0.0051	1.0019	0.0656	0.0043
$\hat{a}_1(1.6)$	1.6011	0.0448	0.0020	1.6006	0.0463	0.0021
$\hat{a}_2(1.6)$	1.6015	0.0941	0.0089	1.5990	0.0866	0.0075

particular form of criterion (2), where $\mathscr{H}$ is the identity operator. In particular, the first derivative regularization filter (32) interpreted as a denoising filter corresponds to the minimization of the following cost function:

$$J(y, y_o) = \|y - y_o\|^2 + \sum_{d=1}^{l+1} \alpha_d(\lambda)\|\mathscr{D}_d[y_o]\|^2. \quad (37)$$

Note that the case of Poisson filter and SRIVC just corresponds to the search of a "delayed" denoised solution (see Sections 4.2 and 4.3). As mentioned in Section 5.1 the same filter is applied to the input and output data, which corresponds to the joint minimization of

$$C(u_o, y_o) = J(u, u_o) + J(y, y_o), \quad (38)$$

where the same criterion J, defined by (37), is used for both input and output signal estimation. The need of a noise free-input signal filtering ("denoising") can be questioned when a null phase filter is used but it is properly stated by considering the problem of error-in-variables model identification (Söderström, Soverini, & Mahata, 2002) for which errors affect both input and output data. In this framework, the question that naturally arises is: should we use the same null phase filter to the input and output data? We will come back to that point at the end of the section, but before, let us try to formalize what is an optimal continuous-time model identification by Poisson filter (including the proposed null phase filter) and the SRIVC methods.

6.2. Joint signal and parameter estimation via Poisson filter

The whole identification procedure, including the determination of the optimal value of λ, corresponds to the following criterion minimization

$$G(\mathscr{A}, \mathscr{B}, u_o, y_o) = C(u_o, y_o) + Q(u_o, y_o, \mathscr{A}, \mathscr{B}) + R(\mathscr{A}, \mathscr{B}), \quad (39)$$

where

$$(\hat{\mathscr{A}}, \hat{\mathscr{B}}, \hat{u}_o, \hat{y}_o) = \arg \min_{\mathscr{A}, \mathscr{B}, u_o, y_o} G(\mathscr{A}, \mathscr{B}, u_o, y_o), \quad (40)$$

and $\mathscr{A}$, $\mathscr{B}$ are operators representing the numerator and the denominator of the model, respectively. For the Poisson filter and the null phase filter

- $C(u_o, y_o)$ is given by (38), with a criterion $J = J_\lambda$ given by (25) for the Poisson filter and by (33) for the regularization filter. Note that the criterion C also depends on the design parameter λ;
- $Q(u_o, y_o, \mathscr{A}, \mathscr{B}) = \|\mathscr{A}[y_o] - \mathscr{B}[u_o]\|^2$ is the ℓ_2-norm of the equation error, minimized to estimate the model parameters from the filtered data;
- $R(\mathscr{A}, \mathscr{B}) = \|y - \frac{\mathscr{B}}{\mathscr{A}}[u_o]\|^2$ is the ℓ_2-norm of the output error, minimized to estimate the optimal value of λ.

Criterion (40) optimization is achieved in three steps:

1. $(\hat{u}_o^{(\lambda)}, \hat{y}_o^{(\lambda)}) = \arg \min_{u_o, y_o} C(u_o, y_o)$ obtained by filtering the data with the corresponding filter structure for a fixed λ;
2. $(\hat{\mathscr{A}}^{(\lambda)}, \hat{\mathscr{B}}^{(\lambda)}) = \arg \min_{\mathscr{A}, \mathscr{B}} Q(\mathscr{A}, \mathscr{B}, \hat{u}_o^{(\lambda)}, \hat{y}_o^{(\lambda)})$ obtained by a parametric estimation algorithm;
3. repeat the two previews steps to minimize $R(\hat{\mathscr{A}}^{(\lambda)}, \hat{\mathscr{B}}^{(\lambda)})$ with respect to λ by an exhaustive search in a fixed range $[\lambda_{\min}, \lambda_{\max}]$.

This optimization procedure separates the signal extraction problem (step 1) from that of model parameter estimation (step 2) by considering them as independent. However, it is clear that these two problems are strongly coupled. We believe that this decoupling scheme is the main shortcoming of such an approach.

6.3. Joint signal and parameter estimation via the SRIVC method

The criterion to minimize can be expressed as

$$G(\mathscr{A}, \mathscr{B}, u_o, y_o) = C(u_o, y_o, \mathscr{A}) + Q(u_o, y_o, \mathscr{A}, \mathscr{B}), \quad (41)$$

where

- $C(u_o, y_o, \mathscr{A})$ is expressed as in Eq. (38) with a criterion $J = J_A$ expressed by (29) and regularization parameters that depend on the model parameters, as given in Section 4.3. Note that the dependence of the criterion C with respect to $\mathscr{A}$ has been made explicit.
- $Q(u_o, y_o, \mathscr{A}, \mathscr{B}) = \|\mathscr{A}[y_o] - \mathscr{B}[u_o]\|^2$ is the ℓ_2-norm of the equation error, minimized to estimate the model parameters from the filtered data.

The criterion is optimized with an iterative procedure. At each iteration r

1. $(\hat{u}_o^{(r+1)}, \hat{y}_o^{(r+1)}) = \arg\min_{u_o, y_o} C(u_o, y_o, \hat{A}^{(r)})$ obtained by filtering the data with the corresponding filter obtained from the previous iteration;
2. $(\hat{\mathscr{A}}^{(r+1)}, \hat{\mathscr{B}}^{(r+1)})$=arg $\min_{\mathscr{A},\mathscr{B}} Q(\mathscr{A}, \mathscr{B}, \hat{u}_o^{(r+1)}, \hat{y}_o^{(r+1)})$ obtained by an instrumental variable estimation algorithm;
3. Repeat steps 1 and 2 until convergence.

The algorithm is initialized using any other estimation method, for example the Poisson filter-based method with a not necessary optimal value of λ. The criterion minimization procedure corresponds to a relaxation method which may preclude the global minimum to be reached. But the particular choice of a regularization filter, depending on the model parameters, ensures a coupling between signal estimation and model identification. In addition, it makes the minimization of the second part of criterion (41) equivalent to the minimization of the output mean square error, expressed by the third part of criterion (39), and results in an optimal parameter estimation in the maximum likelihood sense.

6.4. Joint signal and parameter estimation via regularization

Considering the more general case of error-in-variables model, the joint estimation of input–output signals and model parameters, can be formulated as finding the values of $(\mathscr{A}, \mathscr{B}, u_o, y_o)$ that minimize the following compound criterion:

$$G(\mathscr{A}, \mathscr{B}, u_o, y_o) = \frac{1}{\sigma_1^2}\|u - u_o\|^2 + \frac{1}{\sigma_2^2}\|y - y_o\|^2 + \frac{1}{\sigma_3^2}\|\mathscr{A}[y_o] - \mathscr{B}[u_o]\|^2. \quad (42)$$

The first part of the criterion is a data fitting measure, while the second part is a model fitting measure that regularize the solutions $u_o(t)$ and $y_o(t)$. $\{\sigma_k\}_{k=1}^3$ are regularization parameters that can be interpreted in a Bayesian framework as the standard deviation of the input noise, output noises and the model error, that are assumed Gaussian, respectively. This optimization problem may be solved for example by a joint optimization approach

$$\hat{u}_o^{(r+1)} = \arg\min_{u_o} G(\hat{\mathscr{A}}^{(r)}, \hat{\mathscr{B}}^{(r)}, u_o, \hat{y}_o^{(r)}), \quad (43)$$

$$\hat{y}_o^{(r+1)} = \arg\min_{y_o} G(\hat{\mathscr{A}}^{(r)}, \hat{\mathscr{B}}^{(r)}, \hat{u}_o^{(r+1)}, y_o), \quad (44)$$

$$(\hat{\mathscr{A}}^{(r+1)}, \hat{\mathscr{B}}^{(r+1)}) = \arg\min_{\mathscr{A},\mathscr{B}} G(\mathscr{A}, \mathscr{B}, \hat{u}_o^{(r+1)}, \hat{y}_o^{(r+1)}), \quad (45)$$

where index r denotes the estimation obtained in the iteration r. Concerning problems (43) and (44), $\mathscr{A}$ and $\mathscr{B}$ being fixed to $\hat{\mathscr{A}}^{(r)}$ and $\hat{\mathscr{B}}^{(r)}$, the explicit form of the solution is obtained as

$$\hat{u}_o^{(r+1)}(t) = \frac{1}{1+\alpha_1\hat{\mathscr{B}}^{*(r)}\hat{\mathscr{B}}^{(r)}}[u](t) + \alpha_1\frac{\hat{\mathscr{B}}^{*(r)}\hat{\mathscr{A}}^{(r)}}{1+\alpha_1\hat{\mathscr{B}}^{*(r)}\hat{\mathscr{B}}^{(r)}}[\hat{y}_o^{(r)}](t), \quad (46)$$

$$\hat{y}_o^{(r+1)}(t) = \frac{1}{1+\alpha_2\hat{\mathscr{A}}^{*(r)}\hat{\mathscr{A}}^{(r)}}[y](t) - \alpha_2\frac{\hat{\mathscr{A}}^{*(r)}\hat{\mathscr{B}}^{(r)}}{1+\alpha_2\hat{\mathscr{A}}^{*(r)}\hat{\mathscr{A}}^{(r)}}[\hat{u}_o^{(r+1)}](t), \quad (47)$$

where $\alpha_1 = (\sigma_1/\sigma_3)^2$, $\alpha_2 = (\sigma_2/\sigma_3)^2$. $\hat{\mathscr{A}}^{*(r)}$ and $\hat{\mathscr{B}}^{*(r)}$ correspond to the adjoint operators of $\hat{\mathscr{A}}^{(r)}$ and $\hat{\mathscr{B}}^{(r)}$, respectively. The signals u_o and y_o being fixed to $\hat{u}_o^{(r+1)}$ and $\hat{y}_o^{(r+1)}$,the optimisation problem (45) can be solved by a least-squares method in the ARX model case or an instrumental variable approach. Concerning the hyperparameters $\{\alpha_j\}_{j=1}^2$, because of the noise ambiguity, the problem will be significantly simplified if the noise variances or their ratios are known (Söderström et al., 2002). When the noise variances are not known exactly, a statistical approach reported in Kavetski, Franks, and Kuczera (2002), consists in addressing the problem in a Bayesian framework by assigning an informative prior to either the input or output noise variance.

To conclude this section we note that, similarly to the SRIVC approach, such a joint signal and parameter estimation approach yields filters depending explicitly on the model parameters. The input and output signals are estimated using different filters and these estimations depend not only on the measured signals, but also on the signals and model parameters estimated at the previous iteration. Note that when $\alpha_1 = 0$, which corresponds to the output error model, no filtering needs to be applied to the input signal. Future works will be directed at implementing and investigating more deeply this approach.

7. Conclusion

Time-derivative estimation is an ill-posed inverse problem which is encountered in continuous-time model identification. This paper has shown that applying a linear transformation to the data corresponds to a regularization of this problem to reduce the sensitivity of the estimated time-derivatives to measurement errors. The minimized criterion is expressed explicitly and corresponds to an estimation of delayed signal time-derivatives, justifying the need of filtering the input signal even if it is noise-free. From this interpretation, a null phase filter regularization filters is proposed and applied to continuous-time linear systems identification. This allows to discuss the filter phase effects on parameter estimation. As a result, it is confirmed that regularization filters can be applied to the identification of continuous-time models. But unlike the case of nonlinear systems, the zero phase propriety of the filters is not necessary for the identification, if the same filter is applied to both input and output signals. The main advantage of the synthesis of filters using the regularization point of view is the ability to incorporate additional a priori knowledge on the signals and/or the system. Particularly, the non-stationarity of the signals to restore can be handled by local regularization techniques and the Bayesian formulation associated to stochastic optimization algorithms such as Markov chain Monte Carlo methods (Fitzgerald, 2001; Ninness & Henriksen, 2003) give an attractive framework to address the problem of jointly estimating signals, model parameters and hyperparameters. We also believe that such techniques might be helpful in solving the more difficult error-in-variables model cases.

Acknowledgements

The authors are indebted to the anonymous reviewers whose comments significantly contributed to improve the first version of this paper. The authors thank gratefully Pr. Hugues Garnier, Dr. Eric Huselstein and Dr. Marion Gilson for insightful suggestions and discussions.

References

Cullum, J. (1971). Numerical differentiation and regularization. *SIAM Journal of Numerical Analysis*, *8*(2), 254–265.

Demoment, G. (1989). Image reconstruction and restoration: overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech and Signal Processing*, *37*(12), 2024–2036.

Fitzgerald, W. J. (2001). Markov chain Monte Carlo methods with applications to signal processing. *Signal Processing*, *81*, 3–18.

Garnier, H., Gilson, M., & Huselstein, E. (2003a). Developments for the Matlab CONTSID toolbox. In *Proceeding of 13th IFAC symposium on system identification*, Rotterdam, The Netherlands.

Garnier, H., Mensler, M., & Richard, A. (2003b). Continuous-time model identification from sampled data: implementation issues and performance evaluation. *International Journal of Control*, *76*(13), 1337–1357.

Goodwin, G.C. (2002). Inverse problems with constraints. In *Proceedings of 15th IFAC world congress*, Barcelona, Spain.

Hadamard, J. (1923). *Lectures on Cauchy's problem in linear partial differential equations*. New Haven, CT: Yale University Press.

Hunt, B. R. (1973). The application of constrained least squares estimation to image restoration. *IEEE Transaction on Computers*, *22*(9), 805–812.

Idier, J. (2001). *Approche bayésienne pour les problèmes inverses*. Hermès.

Jakeman, A. J., & Young, P. C. (1984). Recursive and smoothing procedures for the inversion of ill-posed causal problems. *Utilitas Mathematica*, *25*, 351–376.

Johansson, R. (1993). *System modeling and identification, information and system sciences series*. Englewood Cliffs, NJ: Prentice-Hall.

Johansen, T. A. (1996). Identification of nonlinear systems using empirical data and prior knowledge—an optimization approach. *Automatica*, *32*, 337–356.

Johansen, T. A. (1997). On Tikhonov regularization, bias and variance in nonlinear system identification. *Automatica*, *33*, 441–446.

Kavetski, D., Franks, S.W., & Kuczera, G. (2002). Confronting input uncertainty in enviromental modelling. In Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., Trucotte, R. (Eds.), *Calibration of watershed models. AGU water science and applications series* (pp. 49–68).

Ljung, L. (2003). Initialization aspects for subspace and output-error identification methods. In *Proceedings of European control conference*, University of Cambridge, UK.

Mendel, J. M. (1983). *Optimal seismic deconvolution*. New York: Academic Press.

Mensler, M. (1999). *Analyse et étude comparative de méthodes d'identification des systèmes à représentation continue. Développement d'une boite à outils logicielle*. Ph.D. thesis. Université Henri Poincaré, Nancy 1.

Mohammad-Djafari, A. (1996). Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problem. In *Proceedings of IEEE International conference on image processing*, Lausanne, Swiss (pp. 473–477).

Moussaoui, S. (2002). *Identification de modèles à temps continu par la méthode des filtres linéaires: interprétation en termes de régularisation et extensions*. Master's thesis. Université Henri Poincaré, Nancy 1.

Nielson, J. G., Madson, H., & Young, P. C. (2000). Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control*, *24*, 83–94.

Ninness, B., & Henriksen, S. J. (2003). A Bayesian approach to system identification using Markov chain methods. In *Proceedings of 13th IFAC symposium on system identification*, Rotterdam, The Netherlands.

Phillips, D. L. (1962). A technique for the numerical solution of certain integral equations of the first kind. *Journal of the Association for Computing Machinery*, *9*, 84–97.

Rao, G. P., & Garnier, H. (2002). Numerical illustration of the relevance of direct continuous-time domain identification. In *Proceedings of 15th IFAC world congress*, Barcelona, Spain.

Saha, D. C., & Rao, G. P. (1983). Identification of continuous dynamical systems: the Poisson moment functionals (PMF) approach. *Lecture notes in control and information sciences*. Berlin: Springer.

Sjöberg, J., McKelvey, T., & Ljung, L. (1993). On the use of regularization in system identification. In *Proceedings of 12th IFAC world congress*, Sydney, Australia (pp. 381–386).

Söderström, T., & Stoica, P. (1989). *System identification, series in systems and control engineering*. Englewood Cliffs, NJ: Prentice-Hall.

Söderström, T., Carlsson, B., & Bigi, S. (1997). Least squares parameter estimation of continuous-time ARX models from discrete-time data. *IEEE Transactions on Automatic Control*, *42*(5), 659–672.

Söderström, T., Soverini, U., & Mahata, K. (2002). Perspectives on error-in-variables estimation for dynamic systems. *Signal Processing*, *82*, 1139–1154.

Surova, N. S. (1979). An investigation of the problem of reconstructing a derivative by using an optimal regularizing integral operator. *Numerical Methods and Programming*, *1*, 32–34.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. Washington, DC: Winston.

Twomey, S. (1963). On the numerical solution of fredholm integral equations of the first kind. *Journal of the Association for Computing Machinery*, *10*, 97–101.

Unbehauen, H., & Rao, G. P. (1990). Continuous-time approaches to system identification—a survey. *Automatica*, *26*(1), 23–35.

Unbehauen, H., & Rao, G. P. (1998). A review of identification in continuous-time systems. *Annual Reviews in Control*, *22*, 145–171.

Wang, L., & Gawthrop, P. (2001). On the estimation of continuous-time transfer functions. *International Journal of Control*, *74*(9), 889–904.

Young, P. C. (1964). In flight dynamic checkout. *IEEE Transactions on Aerospace*, *2*, 1106–1111.

Young, P. C. (1965). Process parameter estimation and self-adaptive control. In Hammond, P. H. (Ed.), *Theory of self-adaptive systems* (pp. 118–140). NY: Plenum Press.

Young, P. C. (1970). An instrumental variable method for real time identification of a noisy process. *Automatica*, *6*, 271–281.

Young, P. C. (1981). Parameter estimation for continuous-time models—a survey. *Automatica*, *17*(1), 23–39.

Young, P.C. (1993). Time variable and state dependent modelling of nonstationary and nonlinear time series. In Subba Rao, T. (Ed.), *Developments in time series*, volume in honour of Maurice Priestley (pp. 374–413). London: Chapman & Hall.

Young, P. C. (1998). Data-based mechanistic modelling of engineering systems. *Journal of Vibration and Control*, *4*, 5–28.

Young, P.C. (2002). Optimal IV identification and estimation of continuous-time TF models. In *Proceedings of 15th IFAC world congress*, Barcelona, Spain.

Young, P.C., & Foster, M. (1993). A direct approach to the identification and estimation of continuous-time systems from discrete-time data based on fixed interval smoothing. In *Proceedings of 12th IFAC world congress*, Sydney, Australia (pp. 27–30).

Young, P.C., Garnier, H., & Jarvis, A. (2003). The identification of continuous-time linear and nonlinear models: a tutorial with environmental applications. In *Proceedings of 13th IFAC symposium on system identification*, Rotterdam, The Netherlands.

Young, P. C., & Jakeman, A. J. (1980). Refined instrumental variable methods of recursive time-series analysis: part III, extensions. *International Journal of Control*, *31*, 741–764.

Young, P. C., & Pedregal, D. (1999). Recursive and en-bloc approaches to signal extraction. *Journal of Applied Statistics*, *26*, 103–128.

Young, P. C., Pedregal, D., & Tych, W. (1999). Dynamic harmonic regression. *Journal of Forecasting*, *19*, 369–394.