IEEE CDC 92, pp. 3692-3697 C69

Data filtering, reparametrization, and the numerical accuracy of parameter estimators

Gang Li* and Michel Gevers†

* School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 2263
†Centre for Systems Engineering and Applied Mechanics, Université Catholique de Louvain, Bâtiment Maxwell, B-1348 Louvain-La-Neuve, Belgium

Abstract

We establish that a reparametrization of a discrete transfer function in terms of polynomials other than the classical powers of the shift operator z is equivalent to a filtering of the data. In adaptive parameter estimation problems, such reparametrization can yield significantly better accuracy for the estimate and convergence speed for the parameter estimator. We present one such generalized parametrization that yields nice numerical results independently of the spectral properties of the data that enter into the unknown system.

1 Introduction and motivation

Most of the results on the minimization of the effects of Finite Word Length (FWL) errors - due to FWL encoding and/or to arithmetic roundoff - have been for applications in filter design or, more recently, in controller design. Many of these analyses have led to the formulation and the solution of optimal state space design problems, in which the search is performed over the set of all equivalent (i.e. similar) state space realizations of a filter or a controller: see e.g. [4], [9], [5], [2] and many others.

Less attention has been paid to the minimization of FWL error effects on identification or parameter estimation problems. The existing results in this area are for the most part on the selection of algorithms having good numerical properties: [6], [10] and [11]. A notable exception is [3] where Goodwin addressed the numerical issue of parameter estimation and filtering problems, and showed that better accuracy can be obtained if a δ-operator parametrization is used instead of the usual shift-operator parametrization.

Here we generalize this idea: we show that the numerical accuracy with which Least Squares (LS) parameter estimation problems are solved are intimately connected to the parametrization that is used to represent the input-output relation. In fact, we first connect the choice of a parametrization with that of a data prefilter. This establishes the link between the choice of parametrization and the classical techniques that are used by numerical analysts to solve LS problems in a numerically efficient way. These techniques come down to a transformation of the information matrix using QR or other transformation techniques. Therefore they can be seen as both a prefiltering of the regressor vector and a transformation of the parameter vector.

A transformation leading to an information matrix that is optimally conditioned is therefore possible only if the spectral properties of the input and output process (or, equivalently, of the input process and the system being estimated) are known. This is a reasonable assumption in an off-line parameter estimation application, but not in adaptive identification or filtering applications. Our major contribution is to show that some parametrizations, including a particular one based on the bilinear transformation that we call γ-operator parametrization, have some robustness properties in that with these parametrizations the information matrix has desirable properties irrespective of the data processes and of the unknown system.

The paper is organized as follows. In Section 2 we briefly recall the importance of the information matrix in parameter estimation problems. In Section 3 we establish the connections between the choice of parametrization, a filtering of the regressor vector and a filtering of the data processes. Optimal and suboptimal parametrization design games are formulated and discussed in Section 4, while in Section 5 we present our γ-operator parametrization and exhibit the properties it gives to the information matrix in a LS problem. A comparison is also made with the by now more classical δ-operator parametrization. Finally, the advantage of the γ-operator parametrization in adaptive estimation is suggestively evidenced in Section 6 by a simulation.

2 Parameter estimation and the information matrix

Consider that it is desired to estimate the parameters $\{a_i\}$ and $\{b_i\}$ of the transfer function

$$H(z) = \frac{b_0 z^n + b_1 z^{n-1} + \ldots + b_n}{z^n + a_1 z^{n-1} + \ldots + a_i z^{n-i} + \ldots + a_n} \tag{1}$$

from input and output data, $\{u(t)\}$ and $\{y(t)\}$, using an *equation error* formulation as we used to say in the old days, or an ARX model structure as we more fashionably say today. It is then common to set up the following regression model:

$$\begin{aligned} y(t) &= -\sum_{i=1}^{n} a_i y(t-i) + \sum_{i=0}^{n} b_i u(t-i) + v(t) \\ &= \theta_z^T \Psi_z(t) + v(t), \end{aligned} \tag{2}$$

where

$$\Psi_z(t) \triangleq \begin{pmatrix} -y(t-1) & \cdots & -y(t-n) & u(t) & \cdots & u(t-n) \end{pmatrix}^T \tag{3}$$

is the regressor vector, $\theta_z^T \triangleq (a^T \;\; \bar{b}^T)$ is the parameter vector, a and $\bar{b}$ are defined as

$$a = (a_1 \; \ldots \; a_n)^T \quad \bar{b} = (b_0 \; b_1 \; \ldots \; b_n)^T, \tag{4}$$

and $v(t)$ is a noise process. It is a model in the shift operator; hence the notation θ_z. For future use, we also introduce the following expression for $\Psi_z(t)$:

$$\Psi_z(t) = z^{-n} \begin{pmatrix} \nu_n(z) & 0 \\ 0 & \nu_{n+1}(z) \end{pmatrix} \begin{pmatrix} -y(t) \\ u(t) \end{pmatrix}, \tag{5}$$

where $\nu_k(z) \triangleq \left(z^{k-1} \; \ldots \; z \; 1\right)^T$.

The LS estimate of θ_z is obtained as the solution of the following equations:

$$\left(\sum_t \Psi_z(t)\Psi_z^T(t)\right)\hat{\theta}_z = \sum_t \Psi_z(t) y(t), \tag{6}$$

or equivalently, $R_z \hat{\theta}_z = p_z$, where

$$R_z \triangleq \sum_t \Psi_z(t)\Psi_z^T(t) \quad \text{and} \quad p_z \triangleq \sum_t \Psi_z(t) y(t). \tag{7}$$

The matrix R_z is called the *information matrix*. The solution $\hat{\theta}_z$ is unique if R_z is nonsingular.

The information matrix is constructed from the data, and hence its properties depend on these data. Here we formulate some observations about the properties of the information matrix, and their connection with the accuracy of the LS solution. In the next section we show how these properties can be changed by the use of transformations on the regression vector which correspond to changing the basis for the representation of $H(z)$.

- In practice the elements in R_z and p_z are incorrect because of the FWL encoding of the signals and the arithmetic roundoff in the computation of these elements. The solution of the equation (6) will then have an error.
- It can be shown that the relative error of the solution is proportional to the condition number of the information matrix. If fast sampling is used, as is often the case in modern control and estimation problems, the corresponding information matrix is poorly conditioned because the input and output signals are both highly correlated. The condition number actually tends towards infinity when the sampling period goes to zero [8].
- Consider the problem of solving (6) by using the off-line Gradient Algorithm

$$\hat{\theta}_z(k+1) = \hat{\theta}_z(k) + \mu[p_z - R_z\hat{\theta}_z(k)], \tag{8}$$

  where $\hat{\theta}_z(k)$ is the estimate of $\hat{\theta}_z$ at the k-th iteration and μ is the adaptation step size. The speed of convergence of the algorithm can be measured by the maximum of the absolute eigenvalues of $(I - \mu R_z)$, and is influenced by the step size μ and the condition number of R_z.
- A more subtle phenomenon, related to the convergence issue, involves the "coloration" of the regressor vector. If the regressor vector is colored, then R_z will have significant off-diagonal elements and this introduces couplings between different parameters in the parameter vector estimate. This means that, even if some component, $\hat{\theta}_{z,i}$, has already reached its optimal value, it could well be moved away from it if some other component, $\hat{\theta}_{z,j}$, has converged, as long as the corresponding correlation coefficient $R_z(i,j)$ between the two parameters is nonzero.

We conclude that to improve the accuracy of a parameter estimate and to accelerate the convergence of the estimation algorithm, it is desirable to have an information matrix that is

- well-conditioned
- of diagonal form, or with many zero off-diagonal elements.

Our pursuit in this paper is to replace the shift-operator parametrizations by appropriately chosen polynomial-operator parametrizations so that the corresponding information matrices are better conditioned and closer to diagonal matrices. The use of polynomial-operator parametrizations *in lieu of* the classical shift-operator parametrization is effectively a way of filtering the data (and hence the regressor vector), and as such it could be seen as nothing more than a variation on the classical methods that numerical analysts have developed for solving LS problems in a numerically efficient way. Indeed, we shall formally display this link and show how the parametrization can be optimally selected on the basis of data information. In addition, we shall show that, even when no information is available about the statistical properties of the data, *parametrizations can be chosen that have some robustness properties, in that they yield information matrices that have nice properties whatever the data.*

3 Connection between parametrization and data filtering

In this section we show how a filtering of the regression vector can improve the properties of the information matrix, and we establish the connection between such regression filtering and the parametrization of the transfer function.

We consider again the model (1)-(2). Now let $T \in \mathbb{R}^{(2n+1)\times(2n+1)}$ be any nonsingular matrix and let $p \in \mathbb{R}^n$ and $q \in \mathbb{R}^{n+1}$ be any two known vectors. We can then define a new vector θ_p as the solution of the following set of equations:

$$\theta_z = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} p \\ q \end{pmatrix} + T^T \begin{pmatrix} \alpha \\ \bar{\beta} \end{pmatrix} \triangleq \theta_z^0 + T^T \theta_p \tag{9}$$

where $\theta_z^0 \triangleq \begin{pmatrix} p^T & q^T \end{pmatrix}^T$ and $\theta_p \triangleq \begin{pmatrix} \alpha^T & \bar{\beta}^T \end{pmatrix}^T$ with $\alpha \in \mathbb{R}^n$ and $\bar{\beta} \in \mathbb{R}^{n+1}$. It then follows from (2) that

$$y(t) = (\theta_z^0)^T \Psi_z(t) + \theta_p^T T \Psi_z(t) + v(t),$$

or in short

$$y^*(t) = \theta_p^T \Psi_p(t) + v(t), \tag{10}$$

where

$$\begin{aligned} y^*(t) &\triangleq [1 + p_1 z^{-1} + \cdots + p_n z^{-n}] y(t) \\ &\quad -[q_0 + \cdots + q_n z^{-n}] u(t) \\ &\triangleq p(z) z^{-n} y(t) - q(z) z^{-n} u(t) && (11) \\ \Psi_p(t) &\triangleq T\Psi_z(t) = T \begin{pmatrix} \nu_n(z) & 0 \\ 0 & \nu_{n+1}(z) \end{pmatrix} z^{-n} \begin{pmatrix} -y(t) \\ u(t) \end{pmatrix} \\ &\triangleq F(z) \begin{pmatrix} -y(t) \\ u(t) \end{pmatrix} && (12) \end{aligned}$$

with

$$p(z) \triangleq z^n + \sum_{i=1}^{n} p_i z^{n-i} \quad \text{and} \quad q(z) \triangleq \sum_{i=0}^{n} q_i z^{n-i}. \tag{13}$$

We note that $F(z)$ is a $(2n+1) \times 2$ proper rational filter. For given T and θ_z^0, $y^*(t)$ and $\Psi_p(t)$ can be computed from the data $\{u(t)\}$ and $\{y(t)\}$. The problem of estimating θ_z from $\{u(t)\}$ and $\{y(t)\}$ is then converted into one of estimating θ_p using $\{y^*(t)\}$ and $\{\Psi(t)\}$ which are filtered versions of the original signals. The information matrix of $\Psi_p(t)$ will be denoted by R_p:

$$R_p = \sum_t \Psi_p(t) \Psi_p^T(t). \tag{14}$$

It follows immediately from (12) that the information matrices of $\Psi_p(t)$ and $\Psi_z(t)$ are related by

$$R_p = T R_z T^T. \tag{15}$$

These manipulations allow us to make following observations.

1. Filtering the data and the regressor is effectively equivalent to a reparametrization of the initial model: θ_z is replaced by θ_p.

2. The vectors p and q in (9), and hence the polynomials $p(z)$ and $q(z)$, are arbitrary subject to $p(z)$ having stable roots. The filter $\frac{q(z)}{p(z)}$ can thus be set as an (a priori) estimate of the transfer function $H(z)$ of the system. The effect of the filtering operation (11) is to center the parameter vector θ_z on $[p^T \;\; q^T]^T$.

3. The key observation is that by proper filtering - that is by proper choice of T or, equivalently, $F(z)$ - the new regression vector $\Psi_p(t)$ can be made to have nice properties, such as a better conditioned information matrix and, possibly, a diagonal or quasi-diagonal structure, leading to decoupled or quasi-decoupled parameters.

We now establish some further connections between the information matrix of the reparametrized model, the filter applied to the data processes, and the spectral or covariance properties of these data processes. Let $\{u(t)\}$ be a stationary discrete time process. Its autocorrelation function $R_u(\tau)$ and spectral density function $\Phi_u(e^{j\omega})$ are defined as

$$R_u(\tau) \triangleq E[u(t+\tau)u^T(t)], \quad \Phi_u(z) \triangleq \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} R_u(\tau)z^{-\tau}. \tag{16}$$

$R_u(\tau)$ is obtained from $\Phi_u(z)$ as follows:

$$R_u(\tau) = \int_{-\pi}^{\pi} \Phi_u(e^{j\omega})e^{j\omega\tau}d\omega. \tag{17}$$

In particular,

$$R_u \triangleq R_u(0) = \int_{-\pi}^{\pi} \Phi_u(e^{j\omega})d\omega. \tag{18}$$

In the sequel we also replace the information matrices, defined in (7) and (14) on the basis of a particular realization, by their expected values. For example, $R_z \triangleq E[\Psi_z(t)\Psi_z^T(t)]$.

We can then relate the spectrum and the information matrix of the filtered regression vector $\Psi_p(t)$ to that of the shift-operator regressor $\Psi_z(t)$ and to the signal spectra :

$$\begin{aligned}\Phi_p(z) &= F(z)\begin{pmatrix} \Phi_y(z) & -\Phi_{yu}(z) \\ -\Phi_{uy}(z) & \Phi_u(z)\end{pmatrix}F^T(z^{-1}) & (19)\\ &= T\Phi_z(z)T^T, & (20)\end{aligned}$$

and

$$\begin{aligned}R_p &= TR_zT^T & (21)\\ &= \int_{-\pi}^{\pi} F(e^{j\omega})\begin{pmatrix} \Phi_y(e^{j\omega}) & -\Phi_{yu}(e^{j\omega}) \\ -\Phi_{uy}(e^{j\omega}) & \Phi_u(e^{j\omega})\end{pmatrix}F^T(e^{-j\omega})d\omega. & (22)\end{aligned}$$

Having established the equivalence between a reparametrization and a filtering of the regressor $\Psi_z(t)$, we can now turn to optimal and suboptimal design choices for the transformation matrix T or, equivalently, for the matrix filter $F(z)$, given that we know what properties are desirable for the information matrix of $\Psi_p(t)$.

4 Optimal and suboptimal design choices

It is clear from (19)-(22) that if one wants to give some special properties to the information matrix R_p, then the filter $F(z)$ that achieves this aim depends on the data through Φ_u, Φ_y and Φ_{yu} or, equivalently, through Φ_z. In adaptive filtering, such information is not available, and one wants to design a filter that does not explicitly depend on the data. We therefore examine two different strategies. We first discuss the design of optimal (but data-dependent) filters or parametrizations; we then show that some robust parametrizations yield information matrices with desirable properties independently of the spectrum of the data.

4.1 Manipulating the information matrix with data information

Suppose that it is desired to design a $(2n+1)\times 2$ filter $F(z)$ such that the filtered regression vector $\Psi_p(t)$ has a prescribed information matrix R_p. If the spectral density functions $\Phi_u(e^{j\omega})$, $\Phi_y(e^{j\omega})$, and $\Phi_{yu}(e^{j\omega})$ are given, then the $(2n+1)\times 2$ optimal filter $F(z)$ can be obtained by solving (22).

For example, if $R_p = I$ is desired, then the optimal transformation matrix T is the inverse of any square root of R_z. We note that R_z is uniquely defined by the signal spectra Φ_u, Φ_y and Φ_{yu}. Since R_z is a symmetric positive definite matrix, it is always possible to factor it as $R_z = DD^T$. So, if a unit matrix is desired for R_p, then the optimal transformation matrix T_{opt} is given by $T_{opt} = D^{-1}$. The filter $F(z)$ is then uniquely determined from T by (12). The idea of orthogonalizing the regression vector has been used in many estimation and signal processing applications [1], [7] and [3].

4.2 A robustness property

The optimal filter $F(z)$ depends on the spectral properties of the data processes. We now show that, by proper reparametrization - or, equivalently, by proper data filtering - a new regressor can be obtained whose information matrix has desirable properties independently of the spectral properties of the data processes. This new parametrization is thus 'robust'. We first need the following lemma.

Lemma 1 : *Let $H(z)$ and $G(z)$ be two scalar filters driven by the same scalar input $u(t)$, with $y(t) \triangleq H(z)u(t)$ and $w(t) \triangleq G(z)u(t)$. Denote by* $\arg(c)$ *the argument of a complex number c and let $R_{yw}(\tau) \triangleq E[y(t+\tau)w^T(t)]$. Then*

$$R_{yw}(0) = 0 \tag{23}$$

if there exists an integer function $m(\omega)$ such that, for all $\omega \in [0, \pi]$,

$$\phi(\omega) \triangleq \arg(H(e^{j\omega})) - \arg(G(e^{j\omega})) = m(\omega)\pi + \pi/2. \tag{24}$$

Proof: We have $\Phi_{yw}(z) = H(z)\Phi_u(z)G(z^{-1})$. It then follows that

$$\begin{aligned} R_{yw}(0) &= \int_{-\pi}^{\pi} \Phi_{yw}(e^{j\omega})d\omega = \int_{-\pi}^{\pi} F(\omega)e^{j\phi(\omega)}d\omega \\ &= 2\int_0^{\pi} F(\omega)\cos[\phi(\omega)]d\omega \end{aligned} \tag{25}$$

where $F(\omega) \triangleq |H(e^{j\omega})||G(e^{j\omega})|\Phi_u(e^{j\omega})$. The last equality is due to the fact that $F(\omega)$ and the phase difference are even and odd functions of ω, respectively. The lemma then follows from (24). ■

Comment: Note that the condition (24) depends only on the set of filters $H(z)$ and $G(z)$, not on the process~~es $\{y(t)\}$~~ and $\{u(t)\}$. This property enables us to produce filtered regression vectors whose information matrices have special structure. Such regression vectors will be obtained by the introduction of a new γ-operator representation of the transfer function.

5 γ-operator parametrizations

Consider the γ-operator defined as a bilinear transformation of the shift operator as follows:

$$\gamma(z) = \epsilon\frac{z-1}{z+1}, \tag{26}$$

where ϵ is some nonzero real normalizing constant. To stress its dependence on z, we have written $\gamma(z)$ here, and we note that it is a proper and inverse proper function of z. We can rewrite the transfer function (1) as a function of this γ-operator:

$$H(z) = \frac{\beta_0\gamma^n + \beta_1\gamma^{n-1} + \ldots + \beta_n\gamma^0}{\gamma^n + \alpha_1\gamma^{n-1} + \ldots + \alpha_n\gamma^0} \triangleq H_\gamma(\gamma). \tag{27}$$

The relationship between the coefficients $\{a_i, b_i\}$ in (1) and $\{\alpha_i, \beta_i\}$ can be obtained by rewriting (27) in terms of the following polynomial parametrization in z :

$$H(z) = \frac{\beta_0 p_{\gamma_0}(z) + \beta_{\gamma_1} p_{\gamma_1}(z) + \cdots + \beta_n p_{\gamma_n}(z)}{p_{\gamma_0}(z) + \alpha_1 p_{\gamma_1}(z) + \cdots + \alpha_n p_{\gamma_n}(z)} \tag{28}$$

where

$$p_{\gamma_i}(z) \triangleq (z+1)^n \gamma^{n-i} \quad i = 0, 1, \ldots, n. \tag{29}$$

The coefficients $\{a_i, b_i\}$ in (1) and $\{\alpha_i, \beta_i\}$ are related by:

$$\begin{pmatrix} 1 \\ a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix} = k_\gamma T_\gamma^T \begin{pmatrix} 1 \\ \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n \end{pmatrix}, \quad \begin{pmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{pmatrix} = k_\gamma T_\gamma^T \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{pmatrix},$$

where $T_\gamma \in \mathbb{R}^{(n+1)\times(n+1)}$ is given by

$$P_\gamma(z) \triangleq \begin{pmatrix} p_{\gamma_0}(z) \\ p_{\gamma_1}(z) \\ \cdot \\ \cdot \\ \cdot \\ p_{\gamma_n}(z) \end{pmatrix} = T_\gamma \begin{pmatrix} z^n \\ z^{n-1} \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix} = T_\gamma \nu_{n+1}(z) \tag{30}$$

and where k_γ is a normalizing constant that makes $a_0 = \alpha_0 = 1$. This is always possible as long as the system has no pole at $z = -1$. Since T_γ is uniquely determined by the $\{p_{\gamma_i}(z)\}$ defined in (29), the relationship between the two parameter sets $\{a_i, b_i\}$ and $\{\alpha_i, \beta_i\}$ is a one-to-one mapping.

5.1 Computing the information matrix

Using the expression (28) of $H(z)$, the input-output relation can be rewritten as:

$$p_{\gamma_0}(z) z^{-n} y(t) = -\sum_{i=1}^{n} \alpha_i p_{\gamma_i}(z) z^{-n} y(t) + \sum_{i=0}^{n} \beta_i p_{\gamma_i}(z) z^{-n} u(t). \tag{31}$$

The polynomials $p_{\gamma_i}(z)$ are all of degree n. Comparing with (10)-(13), one can see that the γ-operator parametrization (27) corresponds to a data filtering with the following special choices:

$$\begin{aligned} p(z) &= p_{\gamma_0}(z) = (z+1)^n \gamma^n = (z-1)^n, \\ q(z) &= 0 \\ F_\gamma(z) &= \begin{pmatrix} F_{\gamma_1}(z) & 0 \\ 0 & F_{\gamma_2}(z) \end{pmatrix} \end{aligned} \tag{32}$$

$$\begin{aligned} F_{\gamma_1}(z) &\triangleq z^{-n} \left(p_{\gamma_1}(z) \ \cdots \ p_{\gamma_n}(z) \right)^T \triangleq z^{-n} T_{\gamma_1} \nu_{n+1}(z) \\ F_{\gamma_2}(z) &\triangleq z^{-n} \left(p_{\gamma_0}(z) \ \cdots \ p_{\gamma_n}(z) \right)^T \triangleq z^{-n} T_\gamma \nu_{n+1}(z) \end{aligned} \tag{33}$$

where T_γ was defined in (30) and where $T_{\gamma_1} \in \mathbb{R}^{n\times(n+1)}$ is made up of the first n rows of T_γ. The LS regression equation is then written as

$$y^*(t) = \theta_\gamma^T \Psi_\gamma(t) + v(t), \tag{34}$$

where

$$\theta_\gamma \triangleq \left(\alpha_1 \ \cdots \ \alpha_n \ \beta_0 \ \cdots \ \beta_n \right)^T .$$

Here the new process $y^*(t)$ and the new regression vector $\Psi_\gamma(t)$ are computed as (see (11) and (12)):

$$\begin{aligned} y^*(t) &\triangleq p(z)z^{-n}y(t) = (1-z^{-1})^n y(t) \\ \Psi_\gamma(t) &\triangleq F_\gamma(z) \begin{pmatrix} -y(t) \\ u(t) \end{pmatrix} = \begin{pmatrix} T_{\gamma 1} & 0 \\ 0 & T_\gamma \end{pmatrix} \begin{pmatrix} -y(t) \\ \Psi_z(t) \end{pmatrix} . \end{aligned} \tag{35}$$

It will prove useful to rewrite $\Psi_\gamma(t)$ in the following alternative form, which makes the dependence on $y(t)$ very clear:

$$\Psi_\gamma(t) = (1+z^{-1})^n \begin{pmatrix} -\gamma^{n-1}y(t) \\ \cdot \\ \cdot \\ -\gamma^0 y(t) \\ \gamma^n u(t) \\ \gamma^{n-1}u(t) \\ \cdot \\ \cdot \\ \gamma^0 u(t) \end{pmatrix} . \tag{36}$$

We have shown how to compute the regression vector $\Psi_\gamma(t)$ in this γ-operator parametrization. We now explore the properties of its information matrix. To do so, we first prove a key orthogonality property of the operator γ, which can be seen as a filter, $\gamma(z)$.

Lemma 2 : *Let $y(t) \triangleq \gamma^i(z)u(t)$ and $w(t) \triangleq \gamma^k(z)u(t)$. Then*

$$R_{yw}(0) \triangleq \int_{-\pi}^{\pi} \gamma^i(e^{j\omega})\Phi_u(e^{j\omega})\gamma^k(e^{-j\omega})d\omega = 0 \tag{37}$$

for all Φ_u if $|i-k|$ is an odd number. In particular,

$$\int_{-\pi}^{\pi} \gamma(e^{j\omega})\Phi_u(e^{j\omega})d\omega = 0 \quad \textit{for all} \quad \Phi_u . \tag{38}$$

Proof: It follows from elementary high school level calculus that $\gamma(e^{j\omega}) = jctg(\omega/2)$ and hence

$$\arg\{\gamma(e^{j\omega})\} = \frac{\pi}{2}, \quad \forall \omega \in [0,\pi]$$

and

$$\arg\{\gamma^k(e^{j\omega})\} = \frac{\pi}{2}k, \quad \forall \omega \in [0,\pi].$$

Therefore,

$$arg\{\gamma^i(e^{j\omega})\} - arg\{\gamma^k(e^{j\omega})\} = \frac{\pi}{2}(i-k) \quad \forall \omega \in [0,\pi].$$

The result then follows from Lemma 1. ∎

The structure of the information matrix corresponding to $\Psi_\gamma(t)$ follows immediately from this lemma.

Theorem 1 : *Let $\Psi_\gamma(t)$ be the regression vector (36) obtained from the γ-operator parametrization. Then the corresponding information matrix has the form:*

$$R_\gamma = \begin{pmatrix} R_{\gamma_1} & X \\ X^T & R_{\gamma_2} \end{pmatrix} \tag{39}$$

where X has no particular structure while $R_{\gamma_1} \in \mathbb{R}^{n \times n}$ and $R_{\gamma_2} \in \mathbb{R}^{(n+1)\times(n+1)}$ are of the following form: if n is even, then

$$R_{\gamma_1} = \begin{pmatrix} x & 0 & x & 0 & \cdot & \cdot & x & 0 \\ 0 & x & 0 & x & \cdot & \cdot & 0 & x \\ x & 0 & x & 0 & \cdot & \cdot & x & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x & 0 & x & 0 & \cdot & \cdot & x & 0 \\ 0 & x & 0 & x & \cdot & \cdot & 0 & x \end{pmatrix} \tag{40}$$

while $R_{\gamma 2}$ has the same structure with one more row and column.

Proof: Follows directly from (36) and Lemma 2. ∎

Thus, using a γ-operator parametrization yields a regression vector whose information matrix has the nice structure (39)-(40), independently of the data processes $\{u(t)\}$ and $\{y(t)\}$.

We have seen that the presence of off-diagonal coupling terms in the information matrix has the effect of slowing down the convergence of parameter estimation schemes. We shall contrariwise illustrate the beneficial effects of the zero off-diagonal elements in R_γ through simulations. But first we consider an alternative polynomial parametrization based on the δ-operator.

5.2 Comparison with δ-operator parametrizations

In [Goo88] the δ-operator parametrization was studied in the context of parameter estimation, and it was claimed that estimating parameters obtained from a δ-operator transfer function leads to regressors whose information matrices have approximately zero elements in the odd off-diagonals. We now examine this claim. Consider a system represented in a δ-operator transfer function:

$$H(z) = \frac{\beta_0 \delta^n + \beta_1 \delta^{n-1} + \ldots + \beta_n \delta^0}{\delta^n + \alpha_1 \delta^{n-1} + \ldots + \alpha_n \delta^0} \triangleq H_\delta(\delta). \tag{41}$$

The input-output relation can be written as

$$\delta^n y(t) = -\sum_{i=1}^{n} \alpha_i \delta^{n-i} y(t) + \sum_{i=0}^{n} \beta_i \delta^{n-i} u(t).$$

It is easy to see that the corresponding regression vector has the form:

$$\Psi_\delta^T(t) = z^{-n}\left(-\delta^{n-1}y(t), \ldots, -\delta^0 y(t), \delta^n u(t), \ldots, \delta^0 u(t)\right). \tag{42}$$

The properties of the information matrix corresponding to $\Psi_\delta(t)$ are an immediate consequence of the following lemma.

Lemma 3 : *Let $y(t) \triangleq \delta^i(z)u(t)$ and $w(t) \triangleq \delta^k(z)u(t)$. Then*

$$R_{yw}(0) \triangleq \int_{-\pi}^{\pi} \delta^i(e^{j\omega})\Phi_u(e^{j\omega})\delta^k(e^{-j\omega})d\omega \approx 0 \tag{43}$$

if $|i-k|$ is an odd number and if u has a narrow band low pass spectrum.

Proof: It follows from even more elementary high school level calculus that $\delta(e^{j\omega}) = \frac{e^{j\omega}-1}{\Delta} \approx (j\frac{\omega}{\Delta})$ for $|\omega| \ll 1$ which leads to $\arg\{\delta(e^{j\omega})\} \approx \frac{\pi}{2}$ for small ω. Therefore $\arg\{\delta^i(e^{j\omega})\} - \arg\{\delta^k(e^{j\omega})\} \approx \frac{\pi}{2}(i-k)$ for $|\omega| \ll 1$ and for $i, k = 1, \ldots, n$. It then follows from Lemma 1 that the information matrix R_δ formed from the corresponding regression vector $\Psi_\delta(t)$ is approximately of the form (39)-(40), but only for small ω. ■

Comments:

1. We notice that with a δ-operator parametrization the information matrix has the same interesting properties as those achieved with a γ-operator parametrization, but only for narrow band input signals.
2. On the other hand, one advantage of the δ-operator parametrization is that the transformation matrices T_{δ_1} and T_δ, corresponding to T_{γ_1} and T_γ in (33) for the γ-operator parametrization, are upper triangular while T_{γ_1} and T_{γ_2} are fully parametrized. Thus a δ-operator parametrization requires less computational work.

We now illustrate some of our claims with simulations.

6 Applications in estimation and adaptive filtering

We consider the identification of a finite impulse response (FIR) model:

$$H(z) = 1 - 2.1000z^{-1} + 1.4600z^{-2} - 0.3360z^{-3},$$

corresponding to an input-output relation

$$y(t) = \sum_{i=1}^{4} b_i u(t-i) = \theta_z^T \Psi_z(t), \tag{44}$$

with $\Psi_z(t) = (u(t) \;\; u(t-1) \;\; u(t-2) \;\; u(t-3))^T$. We thus consider the estimation of the parameters b_i from input-output data. The input $\{u(t)\}$ is obtained by feeding white Gaussian noise through a low pass filter with a bandwidth of about $\pi/4$. With 950 data points, the regression vector $\Psi_z(t)$ produces an information matrix R_z that has the following properties: $\lambda_{min}(R_z) = 2.1221 \times 10^{-4}, \lambda_{max}(R_z) = 0.7958$. Its condition number is $\kappa \triangleq \frac{\lambda_{max}}{\lambda_{min}} = 3.7500 \times 10^3$.

In the same way, we compute the information matrix corresponding to the regressor $\Psi_\gamma(t)$. The scaling constant ϵ in the γ-operator (see (26)) is selected such as to produce diagonal elements of the same order of magnitude. This can be done on the basis of knowledge of the bandpass of the input data only, and we refer the reader to [GL92] for details. The corresponding information matrix, R_γ, has the following properties: $\lambda_{min}(R_\gamma) = 0.0466$, $\lambda_{max}(R_\gamma) = 0.4727$, $\kappa = 10.1410$.

Comments:

The condition number of R_γ is much smaller than that of R_z. In addition, the computations confirm that the γ-operator parametrization yields an information matrix of zero odd-off-diagonal elements, which guarantees the nice decoupling property of this parametrization.

To confirm our theoretical results, some simulations have been performed in which the parameters have been estimated using the following Least Mean Square (LMS) adaptive algorithm:

$$\begin{aligned} \hat{\theta}_\rho(t+1) &= \hat{\theta}_\rho(t) + \mu \Psi_\rho(t) e_\rho(t) \\ e_\rho(t) &= y(t) - \hat{\theta}_\rho^T(t) \Psi_\rho(t) \end{aligned} \tag{45}$$

where ρ stands for either z or γ, $\hat{\theta}_\rho(t)$ and $\Psi_\rho(t)$ are the parameter vector estimate and the regressor vector at time t in their particular parametrization, respectively. $e_\rho(t)$ is the prediction error, while μ is a constant adaptation step size. The results of our simulations are shown in Figures 1 and 2. They present, respectively, the evolution over time of the prediction error processes $e_z(t)$ and $e_\gamma(t)$ for the shift and γ-operator parametrizations. For the sake of comparison, the step size μ has been set at 0.1 and the initial condition for the parameter vector has been set at zero for both cases.

The superiority of the γ-operator parametrization over the classical shift-operator parametrization in terms of convergence speed is self-evident. But let us further comment that, after 900 iterations, the parameter estimate $\hat{\theta}_z(t)$ still shows no tendency to converge to its true value $[1 \quad -2.1000 \quad 1.4600 \quad -0.3360]^T$. In contrast, for the γ-operator parametrization the estimate $\hat{\theta}_\gamma(450)$ is very close to its true value $[0.1121 \quad 0.1815 \ 0.1016 \quad 0.0240]^T$, while the estimate is almost exact after 900 iterations: $\hat{\theta}_\gamma(900) = [0.1110 \quad 0.1813 \quad 0.0999 \quad 0.0238]^T$.

The comparison just made in terms of convergence speed of the parameter estimates is somewhat unfair to $\hat{\theta}_z(t)$, because the zero initial condition is closer to the exact value of θ_γ than it is to the exact value of θ_z. We have therefore run another simulation with an initial condition $[\ 1.0000 \quad -1.5000 \quad 0.7500 \quad -0.1250\ ]^T$ for both parametrizations. This choice now favors $\hat{\theta}_z$. The evolution of the estimates of the first component of $\hat{\theta}_z(t)$ and $\hat{\theta}_\gamma(t)$ are shown in Figure 3. The values of the parameter vector estimates after 900 iterations are: $\hat{\theta}_z(900) = [\ 0.8343 \quad -1.5567 \quad 0.8164 \quad -0.0346\ ]^T$ and $\hat{\theta}_\gamma(900) = [\ 0.1195 \quad 0.1817 \quad 0.1124 \quad 0.0233\ ]^T$. Note that $\hat{\theta}_{z,1}(900)$ is still far from its exact value, even though the initial value was correct.

These simulations clearly exhibit the perverse effect of the off-diagonal terms of the information matrices in the convergence behavior of the estimation algorithm, since even when the initial condition of a component of $\hat{\theta}_z$ is exact, it is perturbed by the other non-exact terms through the coupling effects.

Another application, to an adaptive noise cancelling problem, is presented in [GL92]: it shows equally dramatic improvements when a γ-operator parametrization is used for the adaptive filter.

7 Conclusions

We have extended an idea of Goodwin [3] who showed that by estimating the parameters of a linear transfer function model using a δ-operator model of this tranfer function, rather than the classical shift-operator model, advantages could be gained in both the accuracy of the estimated prameters and the convergence speed of the estimator. This superior numerical behavior of the δ-operator parametrization is essentially due to an orthogonalizing property of the δ operator, at least when applied to narrow band signals.

We have formally established the connection between reparametrization of the transfer function and data filtering or, equivalently, regression filtering. Using these formal links, one can then play some optimal data filtering (or optimal parametrization) games in order to ameliorate the properties of the ensuing information matrix. This is effectively what the properly chosen algorithms for solving LS regression problems do. However, to perform these optimal transformations, one need to know either the spectral densities of the input and output data, or the spectral density of the input process and the parameters of the unkown system that one is identifying.

In off-line identification, the spectral information can indeed be estimated from the data, but in adaptive applications this information is not available. It is then of interest to develop transformations of the data processes, or equivalently reparametrizations, that yield nicely conditioned and almost diagonal information matrices independently of the data processes. The δ-operator parametrization yields an information matrix with zero odd-off-diagonal terms, but this property evanesces if the signals are wide band. We have shown that an alternative parametrization, called γ-operator parametrization, achieves the same robust (i.e. data-independent) orthogonalization over all frequencies.

Simulations have shown the dramatic improvements that can be achieved using the γ-operator parametrization in adaptive parameter estimation. One such example has been presented here.

We conclude by saying that the formalism introduced by the dual idea of reparametrizing the transfer function or filtering the data opens up many alternative ways of representing transfer functions using basis functions other than the powers of z. The recently introduced Laguerre models can also be seen to fit into this framework [12]; they have also been shown to have interesting numerical properties in the context of the approximation of transfer functions.

References

[1] G. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Academic Press, 1977.

[2] M. Gevers and G. Li, *Parametrizations in Control, Estimation and Filtering Problems: Accuracy Aspects*, Springer Verlag, Communication and Control Engineering Series, 1992.

[3] G.C. Goodwin, 'Some Observations on Robust Stochastic Estimation,' in *Proc. 8th IFAC/IFORS Symp. on Identification and System Parameter Estimation*, Beijing, pp. 22-32, 1988.

[4] S.Y. Hwang, 'Minimum Uncorrelated Unit Noise in State-Space Digital Filtering,' *IEEE Trans. on Acoust. Speech and Signal Processing*, Vol. ASSP-25, No. 4, pp. 273-281, August 1977.

[5] G. Li and M. Gevers, 'Optimal Finite Precision Implementation of a State-Estimate Feedback Controller,' *IEEE Trans. on Circuits and Systems*, Vol. CAS-38, No. 12, pp. 1487-1498, December 1990.

[6] S. Ljung and L. Ljung, 'Error Propagation Properties of Recursive Least Square Adaptive Algorithms,' *Automatica*, Vol. 21, No. 2, pp. 157-167, 1985.

[7] J. Makhoul, 'A Class of All-Zero Lattice Digital Filters: Properties and Applications,' *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-26, No. 4, pp. 304-314, August 1978.

[8] R.H. Middleton and G.C. Goodwin, *Digital Control and Estimation : A Unified Approach*, Prentice Hall, 1990.

[9] C.T. Mullis and R.A. Roberts, 'Synthesis of Minimum Roundoff Noise Fixed-Point Digital Filters,' *IEEE Trans. Circuits and Systems*, Vol. CAS-23, pp. 551-562, 1976.

[10] M. Verhaegen and P. Van Dooren, 'Numerical Aspects of Different Kalman Filter Implementations,' *IEEE Trans. on Automatic Control* Vol. AC-31, No. 10, pp. 907-917, October 1986.

[11] M. Verhaegen , 'Roundoff Noise Propagation in Four Generally-Applicable, Recursive, Least-Square Estimation Schemes,' *Automatica*, Vol. 25, No. 3, pp. 437-444, 1989.

[12] B. Wahlberg, 'System Identification using Laguerre Models' , *IEEE Trans. on Automatic Control*, Vol. 36, No. 5, pp. 551-562, May 1991.

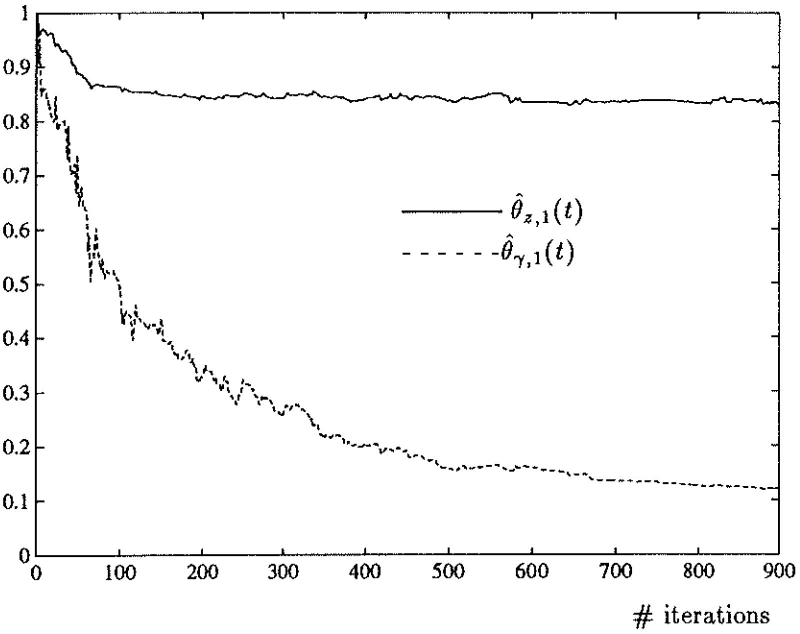

Figure 3: Evolution of $\hat{\theta}_{\rho,1}(t)$ with non-zero initial condition

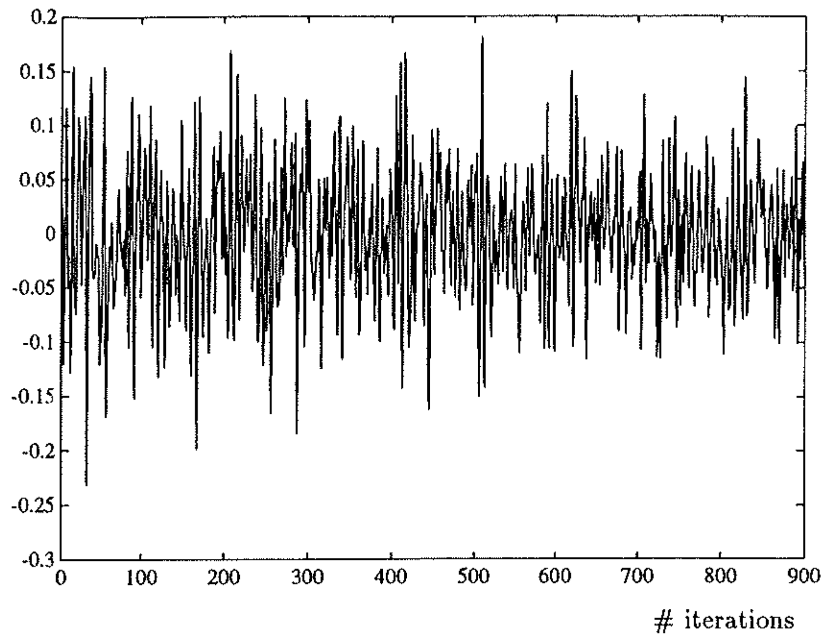

Figure 1: Prediction error sequence $e_z(t)$

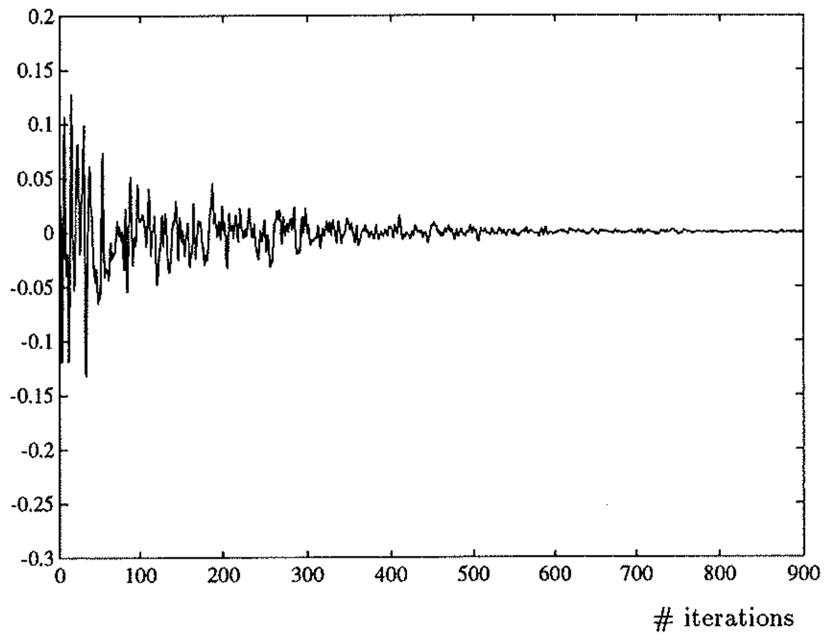

Figure 2: Prediction error sequence $e_\gamma(t)$