

Linear filtering and modelling based on Gram-Schmidt orthogonalization concept

B. Yagoubi

Laboratory of signals and systems
 Department of Electrical Engineering
 Faculty of Sciences and technology, Mostaganem University,
 ALGERIA. E-mail: sbmsvy@gmail.com

Abstract- In this work, the approach we suggest for the linear filtering consists in considering any linear filter as a geometric hyper plane space to which the output signal vector belongs. Any signal orthogonal component to this space vanishes. So removing a non desired component from a signal is to look for a flat space to which this component is orthogonal; in other words, this non desired component will not be observed by orthogonal projection in this geometric space or it does not belong to it and hence, it is eliminated according to Gram-Schmidt orthogonalization concept. To clarify this point view, we compare this geometric filtering procedure to that of an ideal low pass filter in Fourier space and show that it is simple, more efficient and general than the traditional filtering. As an application, we extend this geometric filtering to the linear modelling by eliminating the modelling error, considered as a non-desired output signal component, in order to determine the model coefficients in the case of a linear modelling, linear model identification, and auto-regressive modelling. In addition, using Pythagoras theorem, we calculate the modelling error variance which can be used for testing the linear model approximation quality.

Keywords:- Geometric linear filtering; Gram-Schmidt orthogonalization; orthogonal component; geometric hyper plane; linear model; auto-regressive model.

1 Introduction

A filter is, usually, defined as a procedure that transforms a given signal into a second one that has some more desirable properties or information such as those that are less noisy or distorted [1, ..., 3]. The desired features in the filter output depend, mostly, on the practical applications. If the input is generated by a sensing device, such as a microphone, for example, the filter may attempt to produce an output signal having less background noise or interference. Filter specifications are commonly expressed in the frequency domain, known as the Fourier space [12], characterized by the fundamental stationary sinusoidal oscillation as its base vector. This is the space in which the input signal is projected by Fourier transform. The base vectors of the Fourier space represent stationary oscillations. Their localisation in time is, however, not taken into consideration or is not, in other words, observed and hence orthogonal to the Fourier space. This very useful information can be brought by breaking the signal in a space having both stationary and non stationary oscillations as its base vectors, giving more chance to any information to be, well, generated by this

base. We need to break the signal vector \mathbf{y} at each instant hence we obtain localised co and contra variant components in time. The latter are what we need to see in this oscillations space. The wavelet space is an example of such spaces. In the latter, instead of the contra variant, we look usually, for the co variant components of a signal given by projection of this signal \mathbf{y} in the wavelet space;

$$C_{a,b} = \int \mathbf{y} \cdot \psi_{a,b} dt$$

Where

$$\psi_{a,b} = a^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right)$$

are the base vectors of the wavelet space for the oscillation a (scale) at the instant b . This allows the observation of all different distortions a for a given wavelet shape at this instant. An alternate and more general technique is proposed in this work. This method is geometric and consists in considering any linear filter

as a flat geometric space (hyper plane) [14] with any desired fundamental information as its base vector. We will show how to eliminate any non desired signal component by orthogonal projection in this geometric hyper plane space according to Gram-Schmidt orthogonalization. We further show that our proposed method of filtering can be easily extended to the linear modeling by assuming the modeling error as the non-desired output signal component to be eliminated. This indicates, therefore, the generalized applicability which is a good advantage of our suggested way of filtering over the traditional filtering

2. Discussion

Using relativity terms, a geometric space in which a non desired component projection of a signal vanishes, could be interpreted as that if an observer is in this space (using only this space tools), cannot see this component and, thus, it is exactly eliminated, but only relatively in this space. A relative geometric space behaves, therefore, as a filter. So a linear filtering procedure corresponds geometrically to searching a geometric hyper-plane to which a non-desired component of a signal is orthogonal or equivalently to which a desired component belongs. The choice of the base vectors depends on the desired information; it should represent fundamental information that generates linearly the desired one. For example if we are looking for stationary sinusoidal oscillations, then the Fourier factor $\boldsymbol{\psi} = e^{2j\pi f_0^n}$, as the fundamental base, is known to be the more appropriate. To construct a geometric filter from this space, we reduce its dimension by eliminating the base vector that generates the non desired component. For example if the non-desired component is at f_0 , then the base vector $\boldsymbol{\psi} = e^{2j\pi f_0^n}$ must be omitted to obtain a hyper plane representing the corresponding filter [8, ..., 11].

According to Charles' vectors relation we can break any signal vector \mathbf{y} into as many components as we wish. In our case, we are interested to decompose the output signal into two main components only; the desired component \mathbf{y}_d (the linear output) and the non desired component \mathbf{y}_{nd} (to be eliminated)

$$\mathbf{y} = \mathbf{y}_d + \mathbf{y}_{nd} \quad (1)$$

The aim of this filtering is to eliminate \mathbf{y}_{nd} by orthogonal projection either entirely or partially depending on how both the desired and the non-desired components are related. When there is any interaction between them,

some of the desired information is, unfortunately, eliminated if the desired hyper plane $\boldsymbol{\psi}$ representing the filter is orthogonal to \mathbf{y}_{nd} . A compromise, depending on practical purposes, between the two components is, therefore, needed and it can be made by adjusting geometrically the hyper plane. We will make this point clearer in the following by discussing step by step the possible cases of the geometric linear filtering that may take place and illustrating them by their corresponding to the traditional low pass filtering in Fourier space.

In the following we will be using the contra-variant and the covariant components of a vector which are, basically, different. These are defined as follows:

The contra-variant h^j (or the filter coefficients) of a vector \mathbf{y}_L , in a linear space $\boldsymbol{\Psi}_i$, is obtained by decomposing the vector in this space such that

$$\mathbf{y}_L = h^j \boldsymbol{\Psi}_j$$

Whereas its covariant h_i is obtained by projecting the vector in this space using the following dot product

$$h_i = \langle \mathbf{y}_L, \boldsymbol{\Psi}_i \rangle = \langle h^j \boldsymbol{\Psi}_j, \boldsymbol{\Psi}_i \rangle = h^j \langle \boldsymbol{\Psi}_j, \boldsymbol{\Psi}_i \rangle \quad (2)$$

From this equation we can deduce that the contra-variant and the covariant components are equal only if the base is ortho-normal that is if we have

$$\langle \boldsymbol{\Psi}_j, \boldsymbol{\Psi}_i \rangle = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

2.1 The desired and the non desired components are independent

In the simplest case, we can suppose that the desired and the non-desired components are independent. In this case it is well known that any set of linearly independent vectors can be converted into a set of orthogonal vectors by the Gram-Schmidt process [15]. These two components are, therefore, geometrically orthogonal and hence their dot product vanishes:

$$\langle \mathbf{y}_{nd}, \mathbf{y}_d \rangle = 0 \quad (3)$$

Where $\langle \cdot, \cdot \rangle$ represents the dot product. In this case we can, easily, eliminate, entirely, the non desired component \mathbf{y}_{nd} without affecting the desired one by

choosing the hyper plane, representing the filter, to be orthogonal to this component. This can be performed by adjusting the hyper plane ψ_i such that all the signal desired components belong to it and hence the non desired component \mathbf{y}_{nd} will be orthogonal as indicated in fig. (1), for the hyper plane space and in fig. (2), in the case of Fourier space.

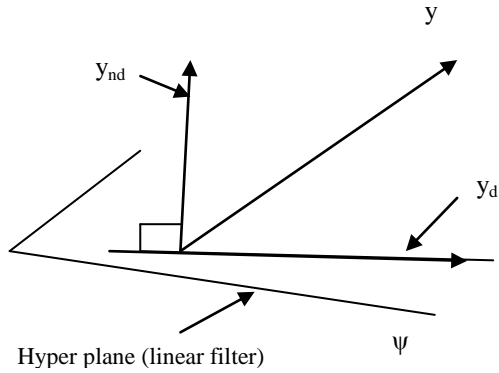


Fig.1. The non-desired \mathbf{y}_{nd} is orthogonal and the desired \mathbf{y}_d is in the space ψ

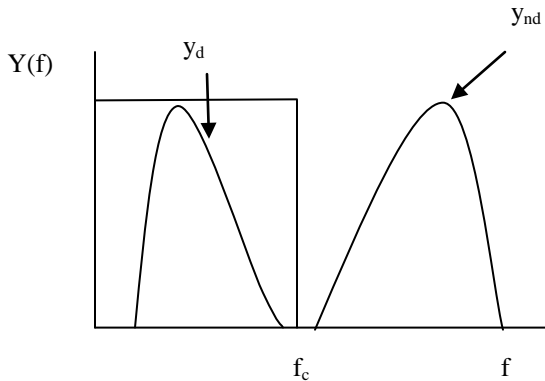


Fig.2. The non-desired \mathbf{y}_{nd} and the desired \mathbf{y}_d are independent in Fourier space

In this case the desired component is given by

$$\mathbf{y}_d = \sum_{j=1}^p h^j \boldsymbol{\psi}_j = h^j \boldsymbol{\psi}_j \quad (4)$$

We have used Einstein convention in (4) for simplification, and by substituting this expression in equation (1), we obtain

$$\mathbf{y} = h^j \boldsymbol{\psi}_j + \mathbf{y}_{nd} \quad (5)$$

h^j represents, geometrically, the j^{th} contra variant component or the filter coefficients in this hyper plane space. So, as we mentioned previously, filtering corresponds to eliminating \mathbf{y}_{nd} by projecting orthogonally equation (5) on the space $\boldsymbol{\psi}_i$ representing the geometric filter, we obtain, thus, the fundamental equation [5, 6] to determine the filter coefficients h^j (the output y contra-variant) with respect to the output y covariant component h_i

$$h_i = \langle \mathbf{y}, \boldsymbol{\psi}_i \rangle = \langle \boldsymbol{\psi}_j, \boldsymbol{\psi}_i \rangle h^j \quad (6),$$

We represent in figure (2) the role of an ideal low pass filter in Fourier space similar to the geometric filter in figure (1). It can be noticed that orthogonal components in geometry, are independent algebraically in Fourier space since there is no band overlapping. They can be, therefore, well separated from each other in the frequency space. As a consequence, the non-desired component can be entirely removed by a low pass without affecting the desired component fig (2). In this example, the aim of using a low pass filter is to reduce the Fourier dimensions in order to obtain a reduced observation space corresponding to the frequency range $[0 \text{ to } f_c]$. Any component that is not within this space is, obviously, not observed (orthogonal) or in other word eliminated.

The modeling error in Wiener filter and the prediction error in autoregressive model are a good example of the non-desired components to be eliminated by projection in the linear input space [4, ..., 7].

2.2 The desired and the non desired components are dependent

We have, here, three possible cases of filtering depending on practical reasons; 1) filtering entirely the desired component, 2) eliminating entirely the non-desired component and finally 3) a compromise between these two cases.

2.2.1 Filtering entirely the desired component.

To filter entirely the desired component we need to adjust the hyper plane ψ representing the filter in order to have \mathbf{y}_d completely within this space. This corresponds to the location of the entire \mathbf{y}_d band in the filter pass band in Fourier space as it is shown in fig.(4). Since there is, however, an interaction between the desired and the non-desired components, some of the non-desired information will be, unfortunately, not

eliminated as it is shown, geometrically, in fig.(3) and correspondingly in Fourier space in fig.(4). Figure (3) shows that the projection of \mathbf{y}_{nd} does not vanish in the hyper plane Ψ and hence some of it \mathbf{y}'_{nd} , is filtered. Similarly, figure (4) shows an overlapping between the two bands and thus it is not possible to filter \mathbf{y}_d entirely without filtering some of the non-desired information. This filtering can be expressed geometrically by the two following relations

$$\mathbf{y}_d = h^j \Psi_j \tag{7}$$

And

$$\langle \mathbf{y}_{nd}, \mathbf{y}_d \rangle \neq 0 \tag{8}$$

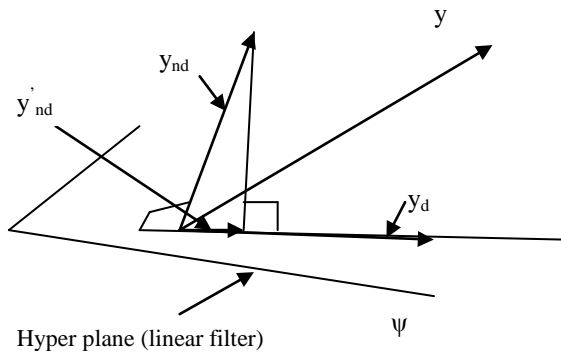


Fig.3. Some of the non-desired information \mathbf{y}'_{nd} is filtered

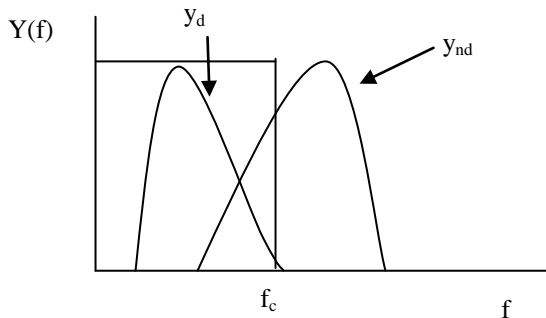


Fig.4. Some of the non-desired information \mathbf{y}'_{nd} is in the pass band (filtered).

2.2.2 Eliminating entirely the non-desired component

In this case we adjust the hyper plane to be exactly orthogonal to the non-desired component \mathbf{y}_{nd} . But as previously, some of the desired information \mathbf{y}'_d will be,

also, orthogonal and thus eliminated since the components are dependent (fig. 5). Similarly in Fourier space the non-desired component is entirely eliminated but, however, due to the band overlapping, some of the desired band still within the cutting band and hence eliminated (fig. 6). This geometric filtering can be described by

$$\langle \mathbf{y}_{nd}, \Psi_i \rangle = 0 \tag{9}$$

And

$$\langle \mathbf{y}_{nd}, \mathbf{y}_d \rangle \neq 0 \tag{10}$$

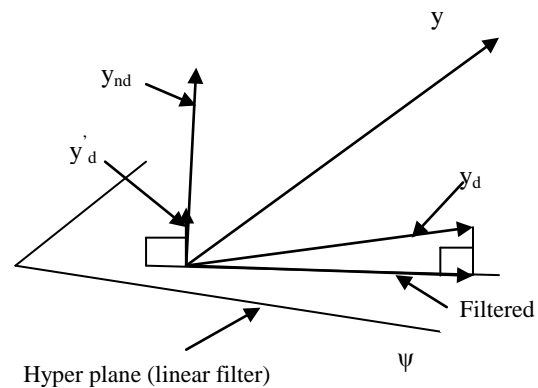


Fig.5. The non-desired and Some of the desired information \mathbf{y}'_d are eliminated.

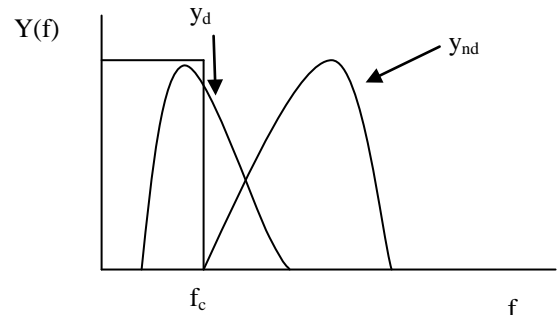


Fig.6. Some of the desired information \mathbf{y}'_d is in the cutting band (eliminated).

2.2.3 A compromise between desired and non-desired components

Since the two components are dependent, a compromise depending on practical purposes corresponds to a situation in which neither of the components is

orthogonal to the hyper plane Ψ . So, in addition to equation (8) we have

$$\langle \mathbf{y}_d, \Psi_i \rangle \neq 0 \tag{11}$$

And

$$\langle \mathbf{y}_{nd}, \Psi_i \rangle \neq 0 \tag{12}$$

By moving this plane either towards the desired or non-desired component, we obtain either more or less desired information respectively. We can use, therefore, the dot product $\langle \mathbf{y}_{nd}, \Psi_i \rangle$ to select the right geometric filter for the desired compromise according to the practical reasons. As this dot product goes to zero, we obtain more desired information. This is illustrated, geometrically, in fig.(7) and in Fourier space in fig.(8).

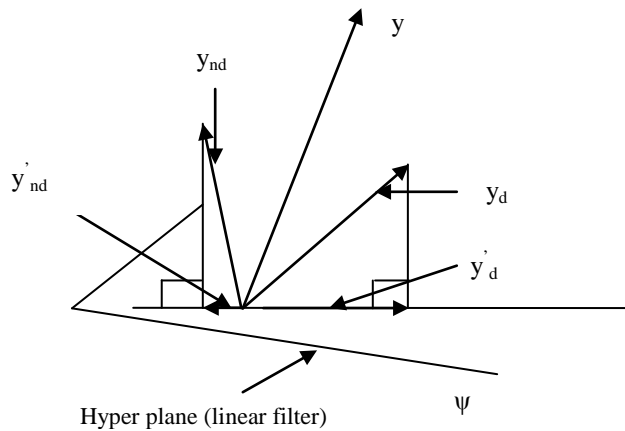


Fig.7. neither of the components are orthogonal to the hyper plane Ψ .

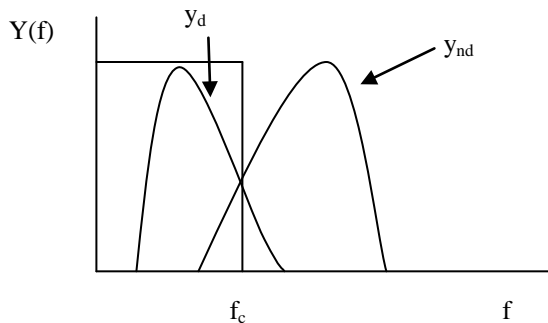


Fig.8. neither of the components is entirely filtered in Fourier space

In all the above cases of filtering, we can notice that

adjusting, geometrically, the hyper plane corresponds to varying the cutting frequency or the pass band of the low pass filter in Fourier space.

2.3. Applications

2.3.1 Linear modelling

The aim of our suggested method of filtering is to extend it, particularly, to the linear modelling by considering the linear modelling error as the non-desired component to be removed from the real model output signal. The real model output \mathbf{y} is, therefore, decomposed according to equation (1) into a linear component as a desired component, and a modelling error \mathbf{e}_{mod} as a non desired component, as follows

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^{i=p} h^i \mathbf{x}_i + \mathbf{e}_{mod} \\ &= \mathbf{y}_d + \mathbf{y}_{nd} \end{aligned} \tag{13}$$

Where

$$\mathbf{y}_d = \sum_{i=1}^{i=p} h^i \mathbf{x}_i \tag{14}$$

and

$$\mathbf{y}_{nd} = \mathbf{e}_{mod} \tag{15}$$

Note that any error is made because it is not observed in our desired space of observation (linear model) or not generated by its base vectors. So, since our desired component is generated by this space, then this component and the modelling error (non-desired) are always independent. This can be also interpreted, geometrically, according to Gram-Schmidt orthogonalization, as that the error is, always, orthogonal to our space of observation to which the desired component (linear) belongs, hence it vanishes by orthogonal projection. The aim, in a linear modelling, is, therefore, simply to choose a geometric hyper plane space to represent a linear filter and to which the desired component (linear) belongs. The modelling error is, automatically, orthogonal to this hyper plane as indicated in figure (9). Since the modelling error (non-desired) and the linear component (desired) are, geometrically, orthogonal, equation (6), in which we have substituted $\Psi_i = \mathbf{x}_i$, allows the determination of the linear model coefficients, hence;

$$\langle \mathbf{y}, \mathbf{x}_i \rangle = \langle \mathbf{x}_j, \mathbf{x}_i \rangle h^j \tag{16}$$

Using the ordinary multiplication and applying the mathematical expectation to both sides of (16), we obtain the linear model coefficients in Hilbert space.

$$E[\mathbf{y} \cdot \mathbf{x}_i] = E[\mathbf{x}_j \mathbf{x}_i] h^j$$

$$= R_x(i, j) h^j \quad (17)$$

This relation is known as Wiener filter[5, 6]. The quantity $R_x(i, j)$ is the auto-covariance matrix element of the inputs. Note that the relation (17) expressing Wiener filter in Hilbert space is, just, a simple covariant and contra-variant components relation obtained, geometrically, by projecting and decomposing the output vector in the geometric input space represented by the inputs \mathbf{x}_i .

So, since the desired (linear) \mathbf{y}_d and the non-desired (the modelling error) ($\mathbf{y}_{nd} = \mathbf{e}_{mod}$) are orthogonal, then according to equation (1) these two components and the output \mathbf{y} form a rectangular triangle as shown in fig.(9). So, to test the linear model approximation, we can determine the modelling error variance from fig.(9) using Pythagoras theorem and the mathematical expectation E, equivalent to Gram-Schmidt process, but simpler, as follows

$$\mathbf{y}^2 = \mathbf{y}_d^2 + \mathbf{y}_{nd}^2$$

$$\Rightarrow E[\mathbf{y}^2] = E[\mathbf{y}_d^2] + E[\mathbf{y}_{nd}^2] \quad (18)$$

Substituting the linear desired component by its expression (14) in this equation (18) we obtain the modelling error variance

$$E[(\mathbf{e}_{mod})^2] = E[\mathbf{y}^2] - \sum_{i=1}^{i=P} (h^i)^2 R_x(i, i) - \sum_{i \neq j} \sum_{j \neq i} h^i h^j R_x(i, j)$$

To obtain a final expression for the error variance, we can simplify this equation using the vector transposition and the Einstein convention mentioned earlier, hence;

$$E[(\mathbf{e}_{mod})^2] = E[\mathbf{y}^2] - \mathbf{h} \cdot \mathbf{h}^T R_x(i, i) - h^i h^j R_x(i, j) \quad (19)$$

$$i \neq j$$

where \mathbf{h}^T is vector transposition representing the linear model coefficients.

Note: that equation (18) obtained geometrically can be, also, obtained algebraically in Hilbert space using the fact that for two independent processes the average of their product is zero, as follows;

$$\mathbf{y}^2 = (\mathbf{y}_d + \mathbf{y}_{nd})^2 \Rightarrow E[\mathbf{y}^2] = E[\mathbf{y}_d^2 + \mathbf{y}_{nd}^2 + 2\mathbf{y}_d \cdot \mathbf{y}_{nd}]$$

$$= E[\mathbf{y}_d^2] + E[\mathbf{y}_{nd}^2] + 2E[\mathbf{y}_d \cdot \mathbf{y}_{nd}]$$

$$= E[\mathbf{y}_d^2] + E[\mathbf{y}_{nd}^2]$$

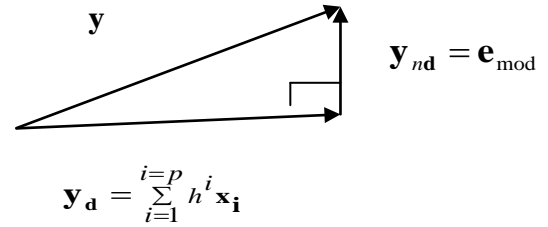


Fig.9 $\mathbf{y}_d = \sum_{i=1}^{i=P} h^i \mathbf{x}_i$ is the linear model output in the hyper-plan space (linear filter) $\Psi_i = \mathbf{x}_i$

2.3.1.1. Consequences of this proposed geometric method.

Definition of the LMR and the LRR

In the linear modelling, the aim is to represent the real model by an adequate linear model. According to figure (9), the real model is more linear if the linear component \mathbf{y}_d (desired) is more important than the modelling error \mathbf{e}_{mod} (non-desired). The latter can be, therefore, considered as an additive noise to the linear component. It is, therefore, convenient to define a ratio (LMR) of the linear component to that of the modelling error, similar to the signal to noise ratio to see the linearity of the real model. This can also be done by calculating another ratio (LRR) of the linear component to that of the real model. These two important ratios, to study the linearity of the real model, are discussed in the following with an example illustrating their role in linear modelling.

a)-The linear output to the modelling error ratio (LMR)

We can calculate the ratio of the desired linear output power P_l to that of the modeling error P_e (non desired) using relation (18) and (19), or from its illustrated version in fig.(9) using the squared cotangent.

$$LMR = \frac{P_l}{P_e} = \frac{E[y_d^2]}{E[(e_{mod})^2]} = \cot^2 g^2(y_d, y) \quad (20)$$

b)-The linear output to the real output ratio (LRR)

An important consequence of this geometric method of filtering, is the cosine square calculated from fig.(9) using Hilbert space norm. It represents the ratio

$$LRR = \frac{P_l}{P_M}$$

of the desired linear output power to that of the real output.

$$LRR = \frac{P_l}{P_M} = \cos^2(y_d, y) = \frac{E[y_d^2]}{E[y^2]} \leq 1 \Rightarrow \quad (21)$$

$$LRR = \frac{E[y_d^2]}{E[y_d^2] + E[(e_{mod})^2]} = \frac{1}{1 + 1/LMR} \leq 1 \quad (22)$$

Notice that the LRR given by equation (21) or (22) indicates, also, the gain or the fraction of the measured output that can be generated linearly with respect to the inputs.

c) – Simulation

In this section, we illustrate the role of the LMR or the LRR ratios given respectively by (20) and (22) in the case of an ideal digital low pass filter.

If the LRR ratio is very low, then the non linear power which is, at the same time, the linear modeling error variance $E[(e_{mod})^2]$, is very high. As a consequence, the linear model coefficients will be smeared out by the second term in the following LMS algorithm (23) if this is used to determine the linear model coefficients.

$$\mathbf{h}_e(i) = \mathbf{h}_e(i-1) + \mu \cdot \mathbf{e}_{mod} \cdot \mathbf{x} \quad (23)$$

Where \mathbf{e}_{mod} is the linear modeling error given by equation (4). μ is the gradient step. So, if the non linear component, present in the output, is very dominant, the LMS algorithm should not be used to determine a linear model. This can be clarified by the following simulation. The linear model, in our simulation, is an ideal low pass whose impulse response (filter coefficients) is given by

$$h_i = \frac{\sin(2\pi i f_c)}{\pi i} \quad (24)$$

Its corresponding finite impulse response (FIR) is obtained, in the literature using, generally, Hamming's window. In our case, we limited its width to P=19 samples corresponding to the number of the filter coefficients. The error measurement is assumed to be a Gaussian centered white noise with a variance $E[(e_m)^2] = \sigma^2 = 10^{-4}$, and the non linear model is of an exponential form. Figure (10) shows the theoretical model coefficients \mathbf{h} eq.(24) in solid curve and the estimated \mathbf{h}_e in dotted curve. It can be seen in

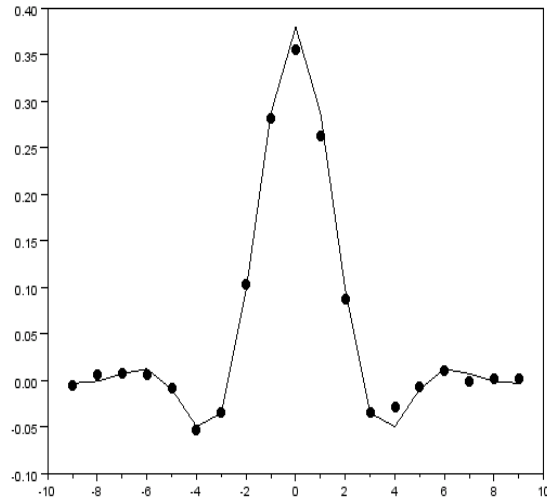


Fig.10. Theoretical model coefficients \mathbf{h} (solid curve) and the estimated \mathbf{h}_e (dotted curve). *LRR* is high

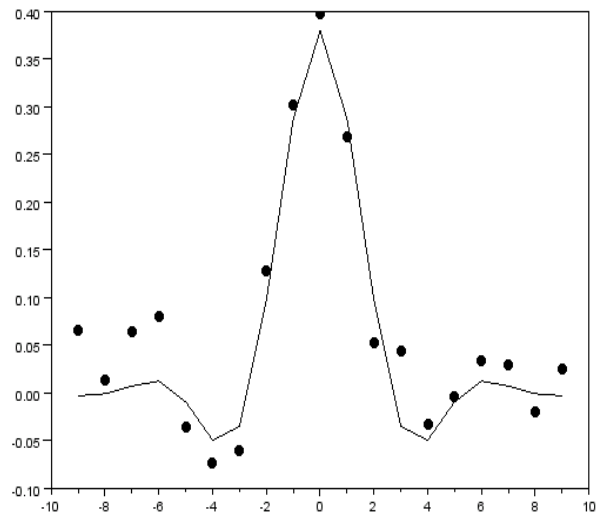


Fig.11. Theoretical model coefficients \mathbf{h} (solid curve) and the estimated \mathbf{h}_e (dotted curve). *LRR* is very low.

this figure that when the linear component is dominant, the best linear approximation, represented by the linear model, is well determined by the LMS algorithm. But as

the non linear component becomes more dominant, by increasing its amplitude and keeping the measurement error variance constant, the estimated model coefficients do not fit the theoretical ones given by equation (24), as it is seen in fig. (11).

2.3.2. Linear model identification.

A linear model output corresponds, geometrically, to a linear geometric space generated linearly by the input base vectors \mathbf{x}_i , hence;

$$\mathbf{Y}_l = \sum_{i=1}^{i=P} h^i \mathbf{x}_i$$

Or $\mathbf{Y}_l = h^i \mathbf{x}_i$ (according to Einstein convention [5])
(25)

This relation expresses a linear model output \mathbf{Y}_l or, simply, a vector in a linear geometric space [5] such that the contra variant components h^i represent the linear model coefficients. In the case of the linear model identification [13], the real model is supposed to be linear. However, the measurement errors as well as other perturbations are, unfortunately, introduced in practice during the output observation. The measured output signal \mathbf{Y} is, therefore, given by

$$\mathbf{Y} = \mathbf{Y}_l + \mathbf{e}_{id} \tag{26}$$

Where \mathbf{Y}_l is the desired linear component given by (1) and \mathbf{e}_{id} is the non-desired component representing all the possible perturbations such as the measurement error and, eventually, other noises that can be observed at the output. These perturbations are not generated by the inputs \mathbf{x}_i , hence orthogonal to any space generated by these inputs. The desired and the non-desired components are, therefore, orthogonal and, as consequence, equation (17) or (6) can be used to determine the linear model coefficients for the linear identification.

2.3.3. Application to Auto-regressive model

In the autoregressive modelling, the vector to be predicted is given, in the literature [1,..., 4], by

$$\mathbf{x}(n) = \sum_{i=1}^{i=P} a^i \mathbf{x}(n-i) + \mathbf{e}_{mod} \tag{27}$$

Where $\sum_{i=1}^{i=P} a^i \mathbf{x}(n-i)$ is the linear prediction representing the desired component, and \mathbf{e}_{mod} is the prediction error or the modelling error representing the non-desired component.

This model is a particular case of a linear modelling with the input past values $\Psi_i = \mathbf{x}_i = \mathbf{x}(n-i)$ (with $i = 1, \dots, P$) as the base vectors, and $\mathbf{y} = \mathbf{x}(n)$ is the real signal to be predicted.

a)- AR-p Coefficients

Substituting $\Psi_i = \mathbf{x}_i = \mathbf{x}(n-i)$ (with $i = 1, \dots, P$) and $\mathbf{y} = \mathbf{x}(n)$ in equation (17), we obtain the auto-regressive model coefficients,

$$E[\mathbf{x}.\mathbf{x}_i] = E[\mathbf{x}_i.\mathbf{x}_j] a^j = \sum_{j=1}^{j=P} E[\mathbf{x}_i.\mathbf{x}_j] a^j \Rightarrow$$

$$R(i) = \sum_{j=1}^{j=P} R(|j-i|) a^j \tag{28}$$

b)-The predicted error variance for AR-p

Substituting the same quantities in equation (19) we obtain the following expression for the prediction error variance

$$E[(\mathbf{e}_{mod})^2] = E[\mathbf{x}(n)^2] - \mathbf{h}.\mathbf{h}^T R_x(i,i) - \sum_{i \neq j} h^i h^j R_x(i,j) \tag{29}$$

If, in addition, the input is a wide sense second order stationary process, this equation will be

$$E[(\mathbf{e}_{mod})^2] = R_x(0) [1 - \sum_{i=1}^{i=P} (a^i)^2] - \sum_{i \neq j} \sum_{j \neq i} a^i a^j . R_x(j-i) \tag{30}$$

Where we have set $h^i = a^i$; the usual notation of the AR coefficients in the literature.

$R_x(0) = E[\mathbf{x}(n-i)^2] \forall i = j$ and $R_x(j-i) = E[\mathbf{x}(n-i)\mathbf{x}(n-j)]$ depends only on the interval length $(j-i)$. Using the vector transposition and Einstein convention, the last expression (30) can be, further, simplified as follows

$$E[(\mathbf{e}_{\text{mod}})^2] = R_x(0)(1 - \mathbf{h}\mathbf{h}^T) - \sum_{i \neq j} a^i a^j R_x(j-i) \quad (31)$$

The LMR and LRR parameters can be calculated from figure (9).

c) - Determination of the LRR for AR-p

The cosine square calculated in Hilbert space norm,

represents the ratio $LRR = \frac{P_l}{P_r}$ of the linear prediction power to that of the real value to be predicted according to equation (20), hence

$$\cos^2 \left(\sum_{i=1}^P a^i \mathbf{x}(n-i), \mathbf{x}(n) \right) = \frac{E \left[\left(\sum_{i=1}^P a^i \mathbf{x}(n-i) \right)^2 \right]}{E[\mathbf{x}(n)^2]} = \frac{P_l}{P_r} \leq 1$$

$$\frac{P_l}{P_r} = \sum_{i=1}^P (h^i)^2 + \frac{1}{R(0)} \sum_{i \neq j} a^i a^j R_x(j-i) \leq 1$$

Using the vector transposition and Einstein convention as previously, the last expression can be written as follows

$$LRR = \frac{P_l}{P_r} = \mathbf{h}\mathbf{h}^T + \frac{1}{R(0)} \sum_{i \neq j} a^i a^j R_x(j-i) \leq 1 \quad (32)$$

with $i, j = 1, \dots, P$

d)-Determination of the LMR for AR-p

We can calculate the ratio of the linear prediction power to that of the prediction error using (20) as follows

$$LMR = \frac{P_l}{P_{pr}} = \frac{E \left[\left(\sum_{i=1}^P a^i \mathbf{x}(n-i) \right)^2 \right]}{E[(\mathbf{e}_{\text{mod}})^2]} \quad (33)$$

This parameter can be, also, deduced from equation (22) as follows

$$LMR = \frac{P_l}{P_{pr}} = \frac{1}{\frac{P_r}{P_l} - 1} = \frac{\frac{P_l}{P_r}}{1 - \frac{P_l}{P_r}} = \frac{LRR}{1 - LRR} \quad (34)$$

To see the application of our geometric method of a linear filtering, we review, in the following, two

particular auto-regressive models that are, usually, used in practice; the AR-1 and AR-2.

2.3.3.1-Application to AR-1

In this case we have a mono-dimensional space $\Psi_1 = \mathbf{x}(n-1)$. The non desired component is the modelling error which is also called the prediction error $\mathbf{y}_{nd} = \mathbf{e}_{\text{mod}}$ that should not be observed, hence orthogonal to $\Psi_1 = \mathbf{x}(n-1)$, to obtain the adequate auto-regressive model coefficients for the best linear prediction. The desired component is the linear prediction $\mathbf{y}_d = \mathbf{a}\mathbf{x}(n-1)$, and $\mathbf{y} = \mathbf{x}(n)$ is the vector (real signal) to be predicted. Equation (27) becomes

$$\mathbf{x}(n) = \mathbf{a}\mathbf{x}(n-1) + \mathbf{e}_{\text{mod}} \quad (35)$$

a)- AR-1 Coefficient

By applying equation (16), we obtain the AR-1 coefficient a

$$E[\mathbf{x}(n)\mathbf{x}(n-1)] = aE[\mathbf{x}^2(n-1)] \Rightarrow a = \frac{R(1)}{R(0)} \quad (36)$$

R is the auto-covariance.

b)-The predicted error variance for AR-1

This is obtained either from the figure (9) through Pythagoras theorem or the relation (31).

$$E[(\mathbf{e}_{\text{mod}})^2] = R_x(0) - R_x(0)a^2 = R_x(0)(1 - a^2) \quad (37)$$

c)- Determination of the LRR for AR-1

Using equation (31), this parameter reduces to

$$LRR = \mathbf{h}\mathbf{h}^T = a^2 \leq 1 \quad (38)$$

Where we have used $R(0) = R(|j-i|) \forall i = j$; the principal diagonal element of the wide sense second order stationary process (WSS) auto covariance matrix. This relation (38) indicates, clearly, that an AR-1 modelling with a coefficient greater than one is, in practice, not possible or it leads, at least, to a non-causal situation.

d)-Determination of the LMR for AR-1

This parameter can be obtained by substituting (38) into (34)

$$LMR = \frac{a^2}{1-a^2} \quad (39)$$

3.3.3.2-Application to AR-2

a)-Determination of the coefficients a^1 and a^2 for AR-2

In this case we have a two-dimensional observation space ($i=1,2$), where $\mathbf{x}_1 = \mathbf{x}(n-1)$ and $\mathbf{x}_2 = \mathbf{x}(n-2)$ are the base vectors. Equation (28) for the two model coefficients will be

For $i=1$:

$$R(1) = a^1 R(0) + a^2 R(1)$$

And for $i=2$

$$R(2) = a^1 R(1) + a^2 R(0)$$

The determinant of this equations system is

$$Det = R(0)^2 - R(1)^2$$

The model coefficients are, thus,

$$a^1 = \frac{R(0)R(1) - R(2)R(1)}{Det}$$

$$a^2 = \frac{R(0)R(2) - R(1)^2}{Det}$$

b)-Determination of the prediction error variance for AR-2

To determine the prediction error variance fig.(9), we apply the general relation (31), as previously, with the coefficients vector $\mathbf{h} = [a^1 \ a^2]$;

$$E[(\mathbf{e}_p)^2] = R(0)[1 - (a^1)^2 - (a^2)^2] - 2a^1 a^2 R(1) \quad (40)$$

c) -Determination of the LRR for AR-2

We can obtain a criterion for AR-2 by using (32);

$$LRR = (a^1)^2 + (a^2)^2 + 2a^1 a^2 \frac{R(1)}{R(0)} \leq 1 \quad (41)$$

This relation is a useful criterion that shows a good AR-2 representation of the real model as the left expression gets closer to 1.

d)-Determination of the LMR for AR-2

By substituting (41) in (34) we obtain the LMR for the AR-2.

Commentary

The application of our geometric method to determine the model coefficients, particularly, for the autoregressive model is clearly much easier and more efficient than the algebraic method, usually, applied in the literature.

4. Conclusion

We have presented a geometric way of a linear filtering in which the fundamental information representing the base vectors should generate the desired component. To find a geometric linear filter in order to eliminate a non-desired component of any signal is to look for a flat space in which it is orthogonal. Mathematically, this can be done by adjusting the hyper plane space to be orthogonal to this component according to Gram-Schmidt orthogonalization concept. In order to clarify the geometric filtering, we compared it to that of a linear filtering in Fourier space, where the fundamental information is supposed to be a stationary sinusoidal osculation. When the desired information is non-stationary, it is better to use non stationary fundamental information, such as a wavelet for the space base vectors. We have seen that when the non desired component is much correlated with the others, it is difficult to isolate it using a flat space without losing useful information. But as the correlation gets lower, a relative flat space, to eliminate a non desired component, becomes a better approximation. We believe that this geometric filtering which is another alternative to traditional filtering is very simple and it can be extended, particularly, to the linear modelling, the linear identification and the linear estimation or prediction such as in the case of the auto-regressive model. The linear modelling was performed by eliminating the modelling error according to Gram-Schmidt orthogonalization, whereas the modelling error variance was calculated using, simply, Pythagoras theorem instead of Gram-Schmidt orthogonalization process.

References

[1] W. Krabs. Mathematical Foundations of Signal Theory. Heldermann Verlag, 1995.

- [2] T.W. Parks, C.S. Burrus. Digital filter design. Wiley-Interscience, New York , 1987.
- [3] C. Herley . Boundary filters for finite-length signals and time-varying filter banks. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, Vol. 42, 1995, 102–114.
- [4] T. Liu, J. Wang, OTHR Impulsive Interference Detection based on AR Model in Phase Domain, *WSEAS transactions on signal processing*, Issue 4, Volume 5, April 2009
- [5] B. Yagoubi, ‘ A geometric Approach to the Linear Modeling’, *Proc. of the 5th WSEAS international conf. on circuits, systems and signals(CSS’11)*, p. 17, (2011).
- [6] H. Furuya, S. Eda, and T. Shimamura, Image Restoration via Wiener Filtering in the Frequency Domain , *WSEAS transactions on signal processing*, Issue 2, Volume 5, February (2009
- [7] P. Y. C. Hwang, R. G. Brown. Introduction to random signals and applied Kalman filtering: with Matlab exercises and solutions, John Wiley sons, Inc., third edition, (1997).
- [8] B. D. O. Anderson, J. B. Moore, *Optimal filtering*, Prentice Hall, (1979).
- [9] S. Haykin, *Adaptive filter theory*, Prentice Hall, (1996).
- [10] K. Suresh, T.V. Sreenivas, Linear filtering in DCT IV/DST IV and MDCT/MDST domain, *signal processing Journal, Elsevier*, V.89(2009
- [11] Chunyan Han, Huanshui Zhang, Linear optimal filtering for discrete-time systems with random jump delays, *signal processing Journal, Elsevier*, V.89(2009)
- [12] R. Bracewell. Fourier Analysis and Imaging. Kluwer Academic /Plenum Publishers, New York, 2003.
- [13] L. Ljung. System identification. Theory for the user. Second edition. Prentice-Hall PTR. 1999..
- [14] Ioannis Z. Emirs, Elias P. Tsigaridas and George M. Tzoumas ‘The Predicates for the exact Voronoi diagram of ellipses under The Euclidean metric’, *International Journal of Computational Geometry and Applications* p. 567, V.(18), (2008)
- [15] A. Ruhe. Numerical Aspects of Gram-Schmidt Orthogonalization of Vectors. *Linear Algebra and its Applications* 52/53 (1983), 591-601.