Chapter 6

Spectral Analysis by Stationary Time Series Modeling

Choosing a parametric model among all the existing models is by itself a difficult problem. Generally, this is *a priori* information about the signal and it makes it possible to select a given model. For that, it is necessary to know the various existing models, their properties, the estimation methods for their parameters and the various possible implementations.

In this chapter, we present the various parametric models of stationary processes. Chapter 4 brought an introduction to this chapter by presenting the models of the time series: the ARMA and Prony models have thus already been introduced. The subject of this chapter is to go through these models by tackling their parameter estimation problem as well as their implementation form.

6.1. Parametric models

The stationary parametric models discussed in this chapter can be regrouped into two categories:

- ARMA models,
- Prony models.

Chapter written by Corinne MAILHES and Francis CASTANIÉ.

The "geometrical" methods, based on the decomposition of a signal sub-space and of a noise sub-space (MUSIC, ESPRIT, etc.), specific of sinusoidal signals, are presented in Chapter 8.

These two types of models were presented in Chapter 4, but in order to make the reading of this chapter easier, we will recall the fundamental equations of various models.

As most of the processes can be well approximated by a linear rational model, the ARMA model (p, q) is an interesting general model. It is described by a finite order linear recursion with constant coefficients:

$$x(k) = -\sum_{n=1}^{p} a_n x(k-n) + \sum_{n=0}^{q} b_n u(k-n) \qquad [6.1]$$

in which $u(k)$ represents a stationary white noise with null mean.

The coefficients a_n correspond to the AR part of the model, while the coefficients b_n determine the MA part.

A particular case of the ARMA models is the AR model of order p, which is written:

$$x(k) = -\sum_{n=1}^{p} a_n x(k-n) + u(k) \qquad [6.2]$$

Autoregressive modeling (AR) is certainly the most popular analysis in parametric spectral analysis. The interest of this method lies in the possibility of obtaining very precise estimations of the AR parameters even for a small number of samples.

In order to have a more physical description of the signal, in terms of frequency, amplitude, phase and damping, the Prony model can also be envisaged. This model consists of describing the signal as a sum of p complex exponentials. This model is thus *a priori* deterministic, contrary to the ARMA and AR models. Thus, it makes it possible to model deterministic signals or moments (autocorrelation functions, for example) by:

$$x(k) = \sum_{m=1}^{p} B_m z_m^k$$

6.2. Estimation of model parameters

In the problem of the estimation of various model parameters, we first take interest in the estimation of AR parameters because this problem is also found again in the ARMA and Prony modelings. We will then tackle the problem of the estimation of the ARMA and Prony model specific parameters.

6.2.1. *Estimation of AR parameters*

Let us suppose that we wish to approach an ordinary signal $x(k)$, for the interval $k = 0, \ldots, N-1$, by the model [6.2]. The method, which is most currently used to calculate the autoregressive coefficients, consists of minimizing the Linear Prediction Error (LPE) in the mean square sense. By considering the recursion equation [6.2], we can realize that we can build a linear predictor of the signal both in the forward and backward direction. We define the forward predictor $\hat{x}(k)$ of $x(k)$ by:

$$\hat{x}(k) = -\sum_{n=1}^{p} a_n x(k-n) \qquad [6.3]$$

We then define the error of forward linear prediction by:

$$e(k) = x(k) - \hat{x}(k) = x(k) + \sum_{n=1}^{p} a_n x(k-n) \qquad [6.4]$$

and the error of backward linear prediction by:

$$b(k) = x(k-p-1) + \sum_{n=1}^{p} a_{p+1-n} x(k-n) \qquad [6.5]$$

The parameters $\mathrm{AR}\{a_n\}_{n=1,p}$ are estimated so that they minimize the following quadratic criterion:

$$\min_{a_k} \left\{ \sigma_e^2 = \sum_{k=p}^{N-1} |e(k)|^2 \right\} \qquad [6.6]$$

and/or the equivalent backward criterion $\min_{ak}\left\{\sigma_b^2\right\}$.

If the signal is efficiently modeled an AR model of a given order p_1, we can show [KAY 88] that the coefficients calculated this way are good estimators of the coefficients $\{a_n\}$ of equation [6.2] if the order of the chosen model p is equal to p_1.

The solution of this least-squares problem is expressed in the following way:

$$\hat{\mathbf{a}} = -\left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H x \text{ with } \hat{\mathbf{a}} = \begin{bmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{bmatrix} \quad [6.7]$$

Depending on the chosen minimization window the matrices and the vectors **X** and **x** are defined in different ways. In the case where:

$$\mathbf{X} = \mathbf{X}_1 = \begin{bmatrix} x(p-1) & \cdots & x(0) \\ \vdots & \ddots & \vdots \\ x(N-2) & \cdots & x(N-p-1) \end{bmatrix} \quad \mathbf{x} = \mathbf{x}_1 = \begin{bmatrix} x(p) \\ \vdots \\ x(N-1) \end{bmatrix} \quad [6.8]$$

this estimation method is improperly called the *covariance method* because the matrix $\mathbf{X}_1^H\mathbf{X}_1$ and the vector $\mathbf{X}_1^H\mathbf{x}_1$ are estimations of covariances with one normalizing coefficient. Morf has given an order recursive algorithm making it possible to calculate this solution without explicit inversion of the matrix [KAY 88, MOR 77].

By adopting a least-squares solution, which minimizes the forward [6.6] and backward sum of the prediction errors, we choose:

$$\mathbf{X} = \mathbf{X}_2 = \begin{bmatrix} x^*(N-p) & \cdots & x^*(N-1) \\ \vdots & \ddots & \vdots \\ x^*(p) & \cdots & x^*(1) \\ x(p-1) & \cdots & x(0) \\ \vdots & \ddots & \vdots \\ x(N-2) & \cdots & x(N-p-1) \end{bmatrix} \quad [6.9]$$

$$\mathbf{x} = \mathbf{x}_2 = \begin{bmatrix} x^*(N-p-1) \\ \vdots \\ x^*(0) \\ x(p) \\ \vdots \\ x(N-1) \end{bmatrix}$$

This is the *modified covariance method* and it is generally more efficient than the covariance method. Sometimes it is also called the maximum entropy method

(*MEM*) [LAN 80] because it is a particular case when the noise is Gaussian. Burg developed an order recursive algorithm making it possible to obtain the reflection coefficients that minimize the sum of the forward and backward prediction errors, which makes it possible to deduce the *AR* parameters [BUR 75] *via* the Levinson-Durbin recursion. The advantage of this algorithm is that the estimated poles are always inside or on the unit circle. Its main disadvantages are a spectral line splitting in the case of a signal made up of a noisy sinusoid with a strong signal-to-noise ratio (*SNR*) and sensitivity at the initial phase [SWI 80, WIL 93]. This sensitivity at the phase can be simply highlighted by considering the following case:

$$x(k) = \cos(2\pi fk + \phi) + u(k)$$

$$\hat{\mathbf{a}} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = -\begin{bmatrix} x(1) & x(0) \\ x(2) & x(1) \end{bmatrix}^{-1} \begin{bmatrix} x(2) \\ x(3) \end{bmatrix} \qquad [6.10]$$

The sensitivity of the solution $\hat{\mathbf{a}}$ is evaluated through the conditioning κ of the system [6.10] [GOL 89]:

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} \qquad [6.11]$$

or $\lambda_{\max}$ and $\lambda_{\min}$ are the eigenvalues of the following matrix:

$$\begin{bmatrix} \cos(2\pi f + \phi) & \cos(\phi) \\ \cos(4\pi fn + \phi) & \cos(2\pi f + \phi) \end{bmatrix} \qquad [6.12]$$

$$\Rightarrow \begin{cases} \lambda_{\max} = \cos(2\pi f + \phi) + \sqrt{\cos(\phi)\cos(4\pi f + \phi)} \\ \lambda_{\min} = \cos(2\pi f + \phi) - \sqrt{\cos(\phi)\cos(4\pi f + \phi)} \end{cases}$$

The difference between these two eigenvalues impacts on the extent of the errors in $\hat{\mathbf{a}}$. It is evident that when $\phi = \frac{\pi}{2}$, the eigenvalues are identical and the errors in $\mathbf{a}$ are minimal. When the estimation of the AR parameters is realized in the least-squares sense, the sensitivity at the initial phases of the sinusoids diminishes but it can remain important if the number of samples is small.

But the estimation of the *AR* parameters of an autoregressive model of order *p* can also be done starting from:

$$\gamma_{xx}(m) = E\left[x(k)x^*(k-m)\right] = -\sum_{n=1}^{p} a_x \gamma_{xx}(m-n) + \sigma_n^2 \delta(m) \qquad [6.13]$$

which we call *Yule-Walker* equations, which were already mentioned in Chapter 4. This result shows that the autocorrelation of $x(k)$ satisfies the same recursion as the signal. A large number of estimation methods solve the Yule-Walker equations by replacing the theoretic autocorrelation $\gamma_{xx}(m)$ by an estimation $\hat{\gamma}_{xx}(m)$:

$$\hat{\mathbf{a}} = -\underbrace{\begin{bmatrix} \hat{\gamma}_{xx}(0) & \hat{\gamma}_{xx}^*(1) & \cdots & \hat{\gamma}_{xx}^*(p-1) \\ \hat{\gamma}_{xx}(1) & \hat{\gamma}_{xx}(0) & \cdots & \hat{\gamma}_{xx}^*(p) \\ \vdots & \ddots & \ddots & \vdots \\ \hat{\gamma}_{xx}(p-1) & \cdots & \cdots & \hat{\gamma}_{xx}(0) \end{bmatrix}^{-1}}_{\hat{\mathbf{R}}_x^{-1}} \underbrace{\begin{bmatrix} \hat{\gamma}_{xx}(1) \\ \hat{\gamma}_{xx}(2) \\ \vdots \\ \hat{\gamma}_{xx}(p) \end{bmatrix}}_{\hat{\mathbf{r}}_x} \qquad [6.14]$$

It is the *autocorrelation method* and $\hat{\mathbf{R}}_x$ is the (estimated) autocorrelation matrix. When $\hat{\gamma}_{xx}(m)$ is the biased estimator of the correlation, the poles are always inside the unit circle, which is not the case with the non-biased estimator which gives, however, a better estimation. The Levinson-Durbin algorithm [LEV 47] provides an order recursive solution of the system [6.14] in $O(p^2)$ computational burden. In order to reduce the noise influence and to obtain a better estimation, the system [6.14] can use a larger number of equations ($>> p$) and be solved in the least-squares sense (LS) or in the total least-squares sense (TLS) [VAN 91]. These methods are called LSYW and TLSYW (see [DUC 98]):

$$\begin{aligned} &\text{LSYW} \quad \hat{\mathbf{a}} = -\left(\hat{\mathbf{R}}_x^H \hat{\mathbf{R}}_x\right)^{-1} \hat{\mathbf{R}}_x^H \hat{\mathbf{r}}_x \\ &\text{TLSYW} \quad \hat{\mathbf{a}} = -\left(\hat{\mathbf{R}}_x^H \hat{\mathbf{R}}_x - \sigma_{\min}^2 I\right)^{-1} \hat{\mathbf{R}}_x^H \hat{\mathbf{r}}_x \\ &\hat{\mathbf{R}}_x \begin{bmatrix} \hat{\gamma}_{xx}(0) & \cdots & \hat{\gamma}_{xx}^*(p-1) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}_{xx}(N_0-1) & \cdots & \hat{\gamma}_{xx}(N_0-p) \end{bmatrix} \quad \hat{\mathbf{r}}_x = \begin{bmatrix} \hat{\gamma}_{xx}(1) \\ \vdots \\ \hat{\gamma}_{xx}(N_0) \end{bmatrix} \end{aligned} \qquad [6.15]$$

where $\sigma_{\min}$ is the smaller singular value of the matrix $\left[\hat{\mathbf{R}}_x \; \hat{\mathbf{r}}_x\right]$, $\boldsymbol{I}$ the identity matrix and N_0 the number of equations. The value of N_0 should be at most of the order of $\frac{N}{2}$ when the non-biased estimator of the correlation is used so as not to make correlations with a strong variance intervene.

The TLS solution is more efficient than the LS method because it minimizes the errors in $\hat{\mathbf{r}}_x$ and in $\hat{\mathbf{R}}_x$ at the same time, but it presents a disadvantage which we

will encounter later. It is generally better to envisage a calculation of the LS and TLS solutions via the singular value decomposition (SVD) of the matrix $\hat{\mathbf{R}}_x$, particularly when the system is not well conditioned, because the SVD proves a greater numeric stability. The LS and TLS solutions calculated by the SVD are:

$$\hat{\mathbf{R}}_x = \hat{\mathbf{U}}_p \hat{\mathbf{\Sigma}}_p \hat{\mathbf{V}}_p^H \quad \hat{\mathbf{\Sigma}}_p = diag\left(\hat{\sigma}_1,\cdots,\hat{\sigma}_p\right) \quad \hat{\sigma}_k \geq \hat{\sigma}_{k+1}$$

$$\begin{aligned} &\text{LSYW} \quad \hat{\mathbf{a}} = -\hat{\mathbf{V}}_p \left(\hat{\mathbf{\Sigma}}_p\right)^{-1} \hat{\mathbf{U}}_p^H \hat{\mathbf{r}}_x \quad \left(\hat{\mathbf{\Sigma}}_p\right)^{-1} = diag\left(\tfrac{1}{\hat{\sigma}_1},\ldots,\tfrac{1}{\hat{\sigma}_p}\right) \\ &\text{TLSYW} \quad \begin{bmatrix}\hat{\mathbf{a}} \\ 1\end{bmatrix} = \frac{\mathbf{v}_{p+1}}{\mathbf{v}_{p+1}(p+1)} \end{aligned} \qquad [6.16]$$

where $\mathbf{V}_{p+1}$ is the eigenvector associated to the smallest eigenvalue of the matrix $\left[\hat{\mathbf{R}}_x\ \hat{\mathbf{r}}_x\right]^H \left[\hat{\mathbf{R}}_x\ \hat{\mathbf{r}}_x\right]$, i.e. the $(p + 1)^{\text{th}}$ column vector of the matrix $\hat{\mathbf{V}}_{p+1}$ so that $\left[\hat{\mathbf{R}}_x\ \hat{\mathbf{r}}_x\right] = \hat{\mathbf{U}}_{p+1} \hat{\mathbf{\Sigma}}_{p+1} \hat{\mathbf{V}}_{p+1}^H$.

There exists an extension of the Levinson-Durbin algorithm for solving LSYW, it is what we call the least-squares *lattice* [MAK 77, PRO 92]. The interest of this algorithm is that it is time and order recursive, which makes it possible to implement parameter estimation adaptive procedures.

We can very easily notice that these methods are biased because the correlation at 0^{th} lag makes the power of the white noise σ_u^2 intervene. The approached value of this bias for (6.14) is:

$$\hat{\mathbf{a}} = -\left(\hat{\mathbf{R}}_{x-u} + \sigma_u^2 I\right)^{-1} \hat{\mathbf{r}}_{x-u}$$

$$\hat{\mathbf{a}} \approx \mathbf{a} + \sigma_u^2 \hat{\mathbf{R}}_{x-u}^{-2} \hat{\mathbf{r}}_{x-u} = \mathbf{a} - \sigma_u^2 \hat{\mathbf{R}}_{x-u}^{-1} \mathbf{a} \qquad [6.17]$$

where $x - u$ is the noiseless signal. Certain methods [KAY 80, SAK 79] exploit this relation in order to try to eliminate this bias, but these require an estimation of σ_u^2. The simplest solution for obtaining a non-biased estimation of the *AR* parameters consists of not making the zero lag correlation intervene in the estimation:

$$\hat{\mathbf{a}} = -\begin{bmatrix} \hat{\gamma}_{xx}(p) & \cdots & \hat{\gamma}_{xx}(1) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}_{xx}(2p-1) & \cdots & \hat{\gamma}_{xx}(p) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\gamma}_{xx}(p+1) \\ \vdots \\ \hat{\gamma}_{xx}(2p) \end{bmatrix} \qquad [6.18]$$

That is called the modified Yule-Walker method (MYW). When we solve this system in the classic or total least-squares sense, we call these methods LSMYW and TLSMYW [STO 92]. When the signal $x(k)$ is really an AR(p), these estimators are asymptotically unbiased ($N \to \infty$). However, if $x(k)$ is made up of a sum of sinusoids and white noise, the MYW estimators are also asymptotically unbiased [GIN 85].

The case of a colored noise slightly modifies the MYW estimators. The hypothesis that the noise $u(k)$ is of the form of a moving average process (MA) of order q is the most common:

$$u(k) = \sum_{n=0}^{q} b_n \varepsilon(k-n), \qquad [6.19]$$

where $\varepsilon(n)$ is a null mean Gaussian white noise. In this case, the correlation $\gamma_{xx}(m)$ of $u(k)$ is null when $m > q$. Thus, an estimator of the *AR* parameters of an ARMA (p, q) or of a sum of noisy sinusoids by a MA(q) is:

$$\hat{\mathbf{a}} = -\begin{bmatrix} \hat{\gamma}_{xx}(p+q) & \cdots & \hat{\gamma}_{xx}(q+1) \\ \vdots & \ddots & \vdots \\ \hat{\gamma}_{xx}(2p+q-1) & \cdots & \hat{\gamma}_{xx}(p+q) \end{bmatrix} -1 \begin{bmatrix} \hat{\gamma}_{xx}(p+q+1) \\ \vdots \\ \hat{\gamma}_{xx}(2p+q) \end{bmatrix} \qquad [6.20]$$

Of course, identifying the noise structure in practice is far from being evident and very often we satisfy ourselves with the whiteness hypothesis.

These methods suppose that the order p is known. In practice it is not necessarily simple to determine it (we will see the methods further) but it should not be underestimated in order not to forget to estimate all the signal poles. When it is overestimated, besides the signal poles, poles linked to the noise also appear; they are more or less damped but in any case they are generally more damped than the signal poles (if the SNR is not too weak). This makes it possible to distinguish between the signal poles and the noise poles. The LSMYW method of an overestimated order is called high-order Yule-Walker method (HOYW). An order overestimation presents the advantage of improving the signal poles estimation. The LSMYW method is relatively efficient in order to distinguish between the signal and the noise poles, on the contrary the TLSMYW method has the disadvantage of attracting the poles (including those of the noise) on the unit circle making the distinction very difficult. We will show this phenomenon on one example; we consider N = 100 samples of a signal made up of non-damped sinusoids of frequencies 0.2 and 0.3 and of a white noise such as SNR = 10 dB. The poles are estimated by the LSMYW and TLSMYW methods with a number of equations $\frac{N}{2}$

and an order p = 10. Figure 6.1 represents the plot of the poles estimated in the complex plane for a signal realization.

We notice that the poles linked to the noise of the LSMYW method are clearly damped in relation to those obtained with TLSMYW which sometimes fall outside the unit circle. When the order is overestimated, the TLSMYW method should be avoided. However, for a correct order, the estimation by TLSMYW is better than that of LSMYW.

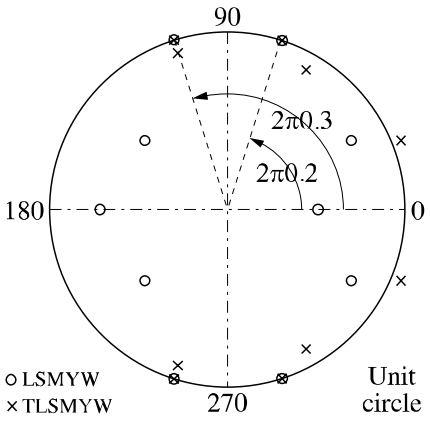

Figure 6.1. *Influence of the order overestimation for the LSMYW and TLSMYW methods*

Up to now, the LSMYW method is one of the most efficient of those that we have described [CHA 82, STO 89]. Stoica and Söderström [SÖD 93] have given the asymptotic performances of this method in the case of a noisy exponential (white noise) of circular frequency $\omega = 2\pi f$ and amplitude A:

$$Var(\hat{\omega}) = \lim_{N\to\infty} E\left[(\hat{\omega}-\omega)^2\right] = \frac{1}{N}\frac{\sigma_u^4}{A^4}\frac{2(2p+1)}{3p(p+1)N_0^2} \qquad [6.21]$$

where p represents the order and $N_0 \geq p$ the number of equations. This expression can be compared to the asymptotic Cramer-Rao bound of the circular frequency ω (see section 3.4.1):

$$CRB(\omega) = \frac{6\sigma_u^2}{A^2N^3} \qquad [6.22]$$

Or we could think that the variance of HOYW [6.21] can become smaller than the Cramer-Rao bound. Actually, this is not the case at all if it is only asymptotically that it can become very close to Cramer-Rao. [SÖD 91] compares this method to the Root-MUSIC and ESPRIT methods (see Chapter 8) in terms of precision and

computational burden. HOYW makes a good compromise between these two aspects.

6.2.2. *Estimation of ARMA parameters*

The estimation of ARMA parameters [KAY 88, MAR 87] is made in two steps: first, we estimate the AR parameters and then the MA parameters. This suboptimal solution leads to a considerable reduction of the computational complexity with respect to the optimal solution that would consist of estimating the AR and MA parameters. The estimation can be efficiently performed in two steps. Starting from equation [6.1], we show that the autocorrelation function of an ARMA process itself follows a recursion of the AR type but starting from the lag q +1 (see section 4.2.2, equation [4.17]):

$$\gamma_{xx}(m) = -\sum_{n=1}^{p} a_x \gamma_{xx}(m-n) \text{ for } m > q$$

The estimation of the AR parameters is done as before by taking these modified Yule-Walker equations into account (starting from the rank q + 1).

The estimation of the MA parameters is done first by filtering the process by the inverse AR filter, using the AR estimated parameters so that we come back to an MA of order q, according to the principle in Figure 6.2.

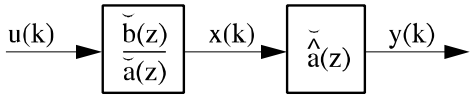

Figure 6.2. *Principle of the estimation of MA parameters in an ARMA*

If the inverse AR filter is supposed perfectly known, we obtain (by using equation [6.1]):

$$y(k) = x(k) + \sum_{n=1}^{p} a_n x(k-n) = \sum_{n=0}^{q} b_n u(k-n)$$

and the filtered process $y(k)$ is a pure MA(q). Thus, we purely come back to a problem of estimation of MA parameters. This model MA(q) can theoretically be modeled by an infinite order AR by writing:

$$\breve{b}(z) = \frac{1}{\breve{c}(z)} \text{ with } \breve{c}(z) = \sum_{n=0}^{+\infty} c_n z^{-n}$$

$$c_k = -\sum_{n=1}^{q} b_n c_{k-n} + \delta(k)$$

We obtain a relation between the MA coefficients and the coefficients of the equivalent AR model:

$$c_k = -\sum_{n=1}^{q} b_n c_{k-n} + \delta(k)$$

$\delta(k)$ standing for the Kronecker symbol ($\delta(0) = 1$ and $\delta(k) = 0$, for any $k \neq 0$).

In practice, we choose to estimate an AR model of high order M such as $M >> q$. By using the AR parameters estimated this way, we obtain a set of equations in the form:

$$\varepsilon(k) = \hat{c}_k + \sum_{n=1}^{q} b_n \hat{c}_{k-n} \qquad [6.23]$$

Ideally, $\varepsilon(k)$ should be null everywhere except for $k = 0$ where it should be equal to 1. As the order of the estimated AR is not infinite, it's nothing of the sort and the estimation of the MA parameters should be done by minimizing a quadratic criterion of the form:

$$\sum_{k} |\varepsilon(k)|^2$$

The index k varies on a domain, which differs according to the used estimation methods. Indeed, equation [6.23] is not unlike the linear prediction error of an AR model in which the parameters would be the b_n and the signal $\hat{c}_k$. From this fact, the AR estimation techniques can be envisaged with, as particular cases: $k = 0, ..., M + q$ corresponding to the method of autocorrelations and $k = q, ..., M$ corresponding to the covariance method.

6.2.3. *Estimation of Prony parameters*

The classic estimation method of the Prony model parameters is based on the recursion expression of a sum of exponentials:

$$x(k) = \sum_{n=1}^{p} B_n z_n^k = -\sum_{n=1}^{p} a_n x(k-1) \text{ for } k \geq p \qquad [6.24]$$

which leads to:

$$x(k) = \sum_{n=1}^{p} a_n x(k-n) + \sum_{n=0}^{p} a_n u(k-n) \quad p \leq k \leq N-1 \quad (a_0 = 1) \qquad [6.25]$$

We find the well-known equivalence between a sum of p noisy exponentials and the ARMA(p, p) model for which the AR and MA coefficients are identical. The poles z_n are deduced from the polynomial roots:

$$\breve{a}(z) = \sum_{n=0}^{p} a_n z^{-n} = \prod_{n=1}^{p} \left(1 - z^{-1} z_n\right) \qquad [6.26]$$

The classic estimation method of the AR parameters in the estimation procedure of the Prony model parameters consists of minimizing the quadratic error:

$$e = \min_{a_n} \sum_{k=p}^{N-1} \left| x(k) + \sum_{n=1}^{p} a_n x(k-n) \right|^2 \qquad [6.27]$$

which, in a matrix form, is written as:

$$\min_{a_n} \left\| \begin{bmatrix} x(p-1) & \cdots & x(0) \\ \vdots & \ddots & \vdots \\ x(N-1) & \cdots & x(N-p) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} + \begin{bmatrix} x(p) \\ \vdots \\ x(N) \end{bmatrix} \right\|^2 \qquad [6.28]$$

$$\min_{\mathbf{a}} \left\| \mathbf{X}\mathbf{a} + \mathbf{x} \right\|^2$$

and leads to the least-squares solution $\mathbf{a}_{LS}$:

$$\mathbf{a}_{LS} = -\left(\mathbf{X}^H \mathbf{X}\right)^{-1} \mathbf{X}^H \mathbf{x}. \qquad [6.29]$$

This method is sometimes called *LS-Prony*. When the system is solved in the total least-squares sense, we speak of *TLS-Prony*. We will note that the matrix $\mathbf{X}^H\mathbf{X}$ is the estimation of the covariance matrix, with one multiplying factor ($\frac{1}{N-p}$). This method is thus identical to the covariance method for the estimation of the *AR*

coefficients and of the signal poles. This method can be slightly changed in the case of $\frac{p}{2}$ non-damped real sinusoids to force the estimated poles to be of unit module:

$$\min_{a_n} \left\| \begin{bmatrix} x^*(N-p+1) & \cdots & x^*(N) \\ \vdots & \ddots & \vdots \\ x^*(1) & \cdots & x^*(p) \\ x(p-1) & \cdots & x(0) \\ \vdots & \ddots & \vdots \\ x(N-1) & \cdots & x(N-p) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_{\frac{p}{2}-1} \\ a_{\frac{p}{2}} \\ a_{\frac{p}{2}-1} \\ \vdots \\ a_1 \\ 1 \end{bmatrix} + \begin{bmatrix} x^*(N-p) \\ \vdots \\ x^*(0) \\ x(p) \\ \vdots \\ x(N) \end{bmatrix} \right\|^2 \qquad [6.30]$$

We speak then of the *harmonic Prony* method. Other estimation methods of the *AR* parameters, implementing the correlation, can be used: *LSYW* and *LSMYW*, for example.

Once the poles are estimated $(\hat{z}_n)$, the complex amplitudes are obtained by solving the Vandermonde system:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \hat{z}_1 & \hat{z}_2 & \cdots & \hat{z}_p \\ \hat{z}_1^2 & \hat{z}_2^2 & \cdots & \hat{z}_p^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{z}_1^{M-1} & \hat{z}_2^{M-1} & \cdots & \hat{z}_p^{M-1} \end{bmatrix} \begin{bmatrix} B_1 \\ \vdots \\ B_p \end{bmatrix} \approx \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(M-1) \end{bmatrix} \qquad [6.31]$$

$$\hat{\mathbf{V}}\mathbf{B} \approx \mathbf{x}$$

Solutions in the least-squares sense and in the total least-squares sense can be envisaged. It was shown in [DUC 97] that the total least-squares solution is less efficient in terms of bias than that of the classic least-squares. On the other hand, the number of equations *M* of the Vandermonde system can be chosen in an optimal manner, as detailed in [DUC 95].

6.2.4. *Order selection criteria*

Choosing a model system is a problem which is as important as the choice of the model itself. Selecting too small an order means smoothing the obtained spectrum, while choosing too large an order introduces secondary spurious peaks.

There is a large number of order selection criteria which are for most cases based on the statistic properties of the signal (maximum likelihood estimation: MLE). Others, simpler and less efficient are based on the comparison of the eigenvalues of the correlation matrix to some threshold correlation matrix [KON 88].

A large number of order selection criteria use the prediction error power decrease when the order increases. When the theoretic order is reached, this power remains constant. However, a criterion based only on the prediction error power shape does not make it possible to take the estimated spectrum variance increase into account when the order is overestimated. That is why the criteria integrate these two phenomena. One of the first criteria proposed by Akaike [AKA 70] was the *FPE* (Final Prediction Error): the estimated error corresponds to the value that minimizes:

$$FPE(k) = \frac{N+k}{N-k}\hat{\rho}_k \qquad [6.32]$$

where:

$$\hat{\rho}_k = \hat{\gamma}_{xx}(0) + \sum_{l=1}^{k} \hat{a}_l \hat{\gamma}_{xx}(l) \qquad [6.33]$$

is the power of the prediction error that decreases with k while the term $\frac{N+k}{N-k}$ increases with k (to take the estimated spectrum variance augmentation into account when k increases). The AR parameters are estimated through Yule-Walker equations with the biased estimator of the correlation. The most well known criterion proposed by Akaike is the AIC (Akaike Information Criterion) [AKA 74]:

$$AIC(k) = N\ \ln(\hat{\rho}_k) + 2k \qquad [6.34]$$

This criterion is more general than FPE and it can be applied by determining the order of an MA part of an ARMA model. Asymptotically $(N \to \infty)$ FPE and AIC are equivalent, but for a small number of samples AIC is better. It was proved that AIC is inconsistent and it tends to overestimate the order [KAS 80]. [RIS 83] proposed to modify AIC by replacing the term *2k* by a term, which increases more

rapidly (depending on N) k ln (N). This criterion is named MDL (Minimum Description Length):

$$\mathrm{MDL}(k) = N \ln(\hat{\rho}_k) + k \ln(N) \qquad [6.35]$$

This criterion is consistent and gives better results than AIC [WAX 85].

It would be tedious to present all the criteria that were developed; for more information on this, see the following references: [BRO 85, BUR 85, FUC 88, PUK 88, YIN 87, WAX 88].

[WAX 85] expressed the AIC and MDL criteria depending on the eigenvalues of the autocorrelation matrix $\hat{R}_y$:

$$AIC(k) = N(p-k)\ln\left(\frac{\left(\prod_{t=k+1}^{p}\hat{\lambda}_t\right)^{\frac{1}{p-k}}}{\frac{1}{p-k}\sum_{t=k+1}^{p}\hat{\lambda}_t}\right) + k(2p-k) \quad k = 1,\cdots,p-1 \qquad [6.36]$$

$$MDL(k) = N(p-k)\ln\left(\frac{\left(\prod_{t=k+1}^{p}\hat{\lambda}_t\right)^{\frac{1}{p-k}}}{\frac{1}{p-k}\sum_{t=k+1}^{p}\hat{\lambda}_t}\right) + \frac{1}{2}k(2p-k)\ln(N) \qquad [6.37]$$

where $\hat{\lambda}_t$ are the ordered eigenvalues $\left(\hat{\lambda}_t \geq \hat{\lambda}_{t+1}\right)$ of the matrix $\hat{R}_v(p \times p)$. It is possible to define these criteria according to the singular values $\hat{\sigma}_t$ of the matrix $\hat{R}_v(M \times p, M > p)$ by replacing $\hat{\lambda}_t$ by $\hat{\sigma}_t^2$ in the matrices [6.36] and [6.37] [HAY 89].

The increase in the number of lines of the matrix $\hat{R}_v$ evidently makes an improvement of the order estimation performances possible. We have compared the two expressions [6.35] and [6.37] of the MDL criterion on the following example. We consider N = 100 samples of a signal made up of two sinusoids of identical amplitudes of frequencies 0.1 and 0.2 and of one white noise. The dimension of the matrix $\hat{R}_v$ is (30 x 30). The simulations were carried out for signal-to-noise ratios of 10 dB and 0 dB and are presented in Figures 6.3 and 6.4.

SNR = 10 dB (Figure 6.3): the dimension of the subspace signal is 4 because the signal has two sinusoids. The criterion [6.37] (top curve) reaches its minimum for $k = 4$. However, the criterion [6.35] (bottom curve) is minimum when $k = 8$. These results illustrate the efficiency of the criterion defined starting from the eigenvalues of the autocorrelation matrix and the mediocre performances of the criterion defined starting from the prediction error power.

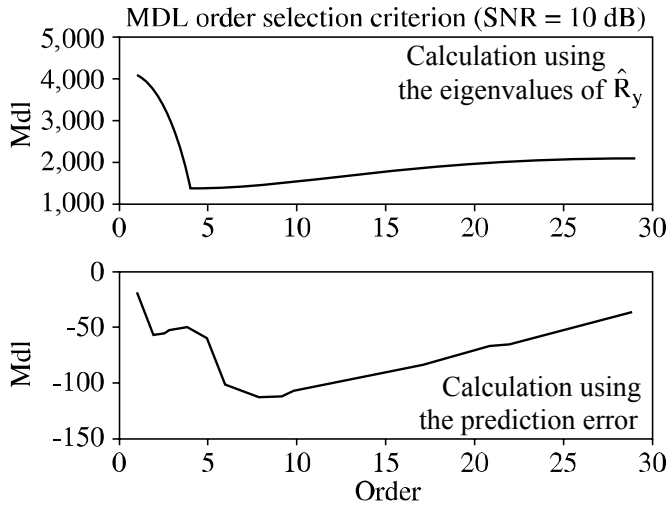

Figure 6.3. *Comparison of the MDL criteria [6.35] and [6.37] for SNR = 10 dB*

SNR = 0 dB (Figure 6.4): for a weaker SNR, the criterion [6.35] (top curve) is inefficient while the criterion [6.37] (top curve) gives a correct result.

These simulations have highlighted the efficiency of the MDL criterion [6.37] and, in a more general manner, the efficiency of the criteria built starting from the eigenvalues of the autocorrelation matrix $\hat{\mathbf{R}}_x$.

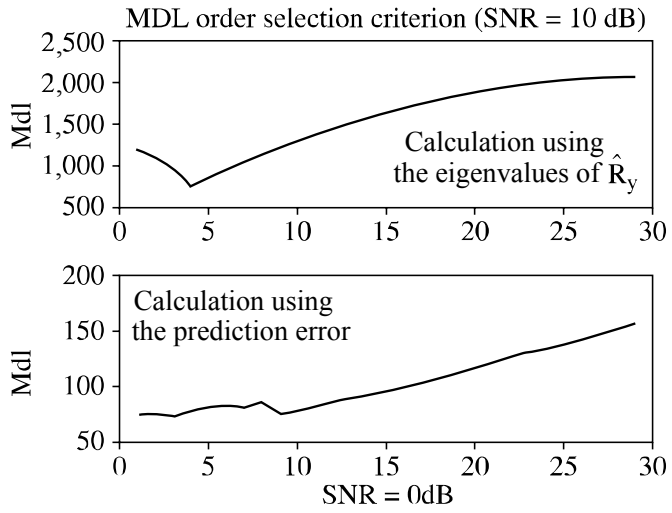

Figure 6.4. *Comparison of the MDL criteria [6.35] and [6.37] for SNR = 0 dB*

In the case of a Prony model, the model order corresponds to the number of searched exponential components, which evidently reduces to an order determination problem of AR model. All the preceding criteria are applicable.

In the case of an ARMA model, selecting the models p and q of the respective AR and MA parts is not a simple problem. Few published works tackle this problem, excepting the very simple cases. The AIC criterion is one of the most used in the form:

$$\mathrm{AIC}(\mathrm{p,q}) = N \ln\left(\hat{\rho}_{pq}\right) + 2\left(p+q\right)$$

where $\hat{\rho}_{pq}$ represents the estimated power of the entrance noise of the ARMA model. As in the case of the AR model, the minimum of this function with two variables provides the couple (p, q) of the AR and MA orders to be taken into account.

6.3. Properties of spectral estimators produced

In the case of a signal made up of a sinusoid and an additive white noise, we take interest in the AR model spectrum evolution, according to the signal to noise ratio (SNR), for a fixed AR order. Various results can be found in the literature, depending on the hypotheses taken for the modeled signal. For an example, we can

mention [LAC 84] which provides the equivalent bandwidth expression for a pure sinusoid with additive white noise, modeled by an AR model of order p:

$$\Delta f_p = \frac{6}{\pi p(p+1)\beta}$$

β being the signal to noise ratio. Figure 6.5 makes it possible to highlight that the bandwidth at -3dB of the spectrum lobe depends on the SNR: the weaker the SNR is, the more the lobe enlarges. The AR spectral resolution is affected by the noise presence.

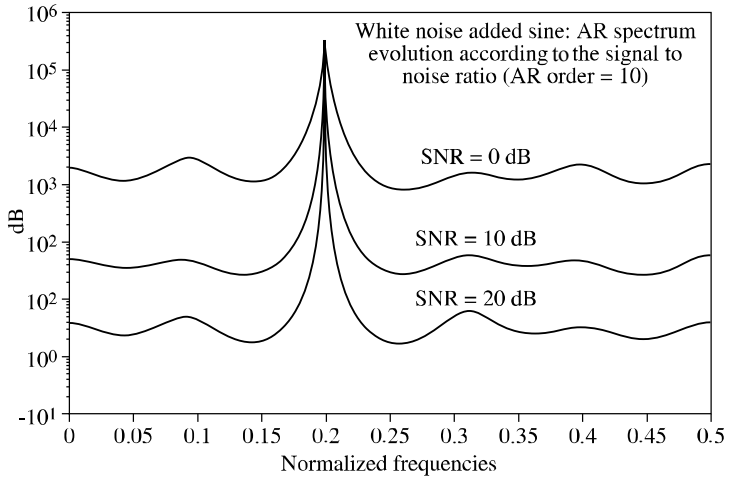

Figure 6.5. *Evolution of the AR spectrum according to the SNR in the case of a noisy sinusoid*

In the case of a noisy sinusoidal signal, Figure 6.6 plots the evolution of the AR spectrum according to the order of the chosen model, for a previously fixed SNR. The larger the order is, the better the spectral resolution is (inversely related to the width of the "peaks"). But if the chosen order is too big, false (or spurious) "peaks" appear.

The effect of the order choice is again illustrated on the case of a signal made up of three sinusoids in white noise. The minimal order necessary for the estimation of these three sinusoids is of 6. Of course, the presence of the additive white noise makes it necessary to choose a higher order. Figure 6.7 presents the AR spectrum evolution, while the chosen order is below 6: if we could see that like an animated cartoon, we would see the successive lines "push" according to the chosen order. Below, we have represented the periodogram, which makes it possible to appreciate the smoothing produced by the AR spectral analysis.

Figure 6.8 presents the AR spectrum evolution on the same signal made up of three sinusoids in white noise when the chosen order is superior to 6. The three lines are better and better estimated but at the order 50, the AR modeling makes two false "peaks" appear, which are actually noise peaks.

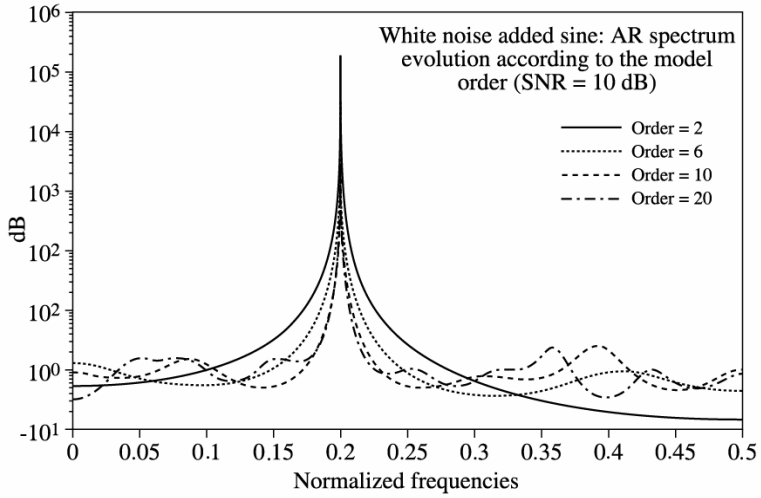

Figure 6.6. *AR spectrum evolution according to the order of the chosen model in the case of a noisy sinusoid*

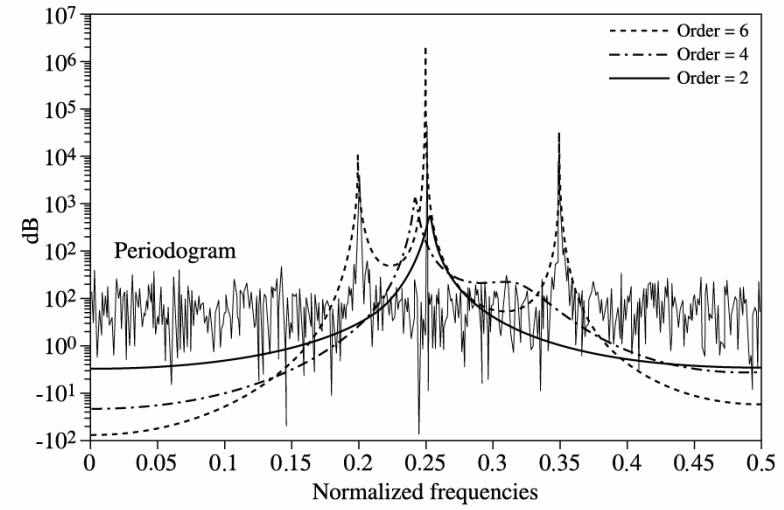

Figure 6.7. *AR spectrum evolution (compared to the periodogram) according to the order chosen for a signal made up of 3 sinusoids in white noise*