Institut Mines-Télécom

Applications et ouvertures

Roland Badeau
roland.badeau@telecom-paris.fr

Contents

Acronyms **4**

Mathematical notation **6**

1 High resolution methods **7**
1 Introduction . . . 8
2 Signal model . . . 8
3 Maximum likelihood method . . . 9
3.1 Application of the maximum likelihood principle to the ESM model . . . 9
3.2 Maximum likelihood and Fourier resolution . . . 11
4 High resolution methods . . . 12
4.1 Linear prediction techniques . . . 12
4.1.1 Linear recurrence equations . . . 13
4.1.2 Prony method . . . 13
4.1.3 Pisarenko method . . . 15
4.2 Subspace methods . . . 15
4.2.1 Singular structure of the data matrix . . . 15
4.2.2 Singular structure of the correlation matrix . . . 16
4.2.3 Complement: analogy between the spectrum in the matrix sense and in the Fourier sense . . . 17
4.2.4 MUltiple SIgnal Characterization (MUSIC) . . . 18
4.2.5 Estimation of Signal Parameters via Rotational Invariance Techniques . . . 18
5 Estimation of the other parameters . . . 20
5.1 Estimation of the modeling order . . . 20
5.2 Estimation of amplitudes, phases and standard deviation of noise . . . 21
6 Performance of the estimators . . . 21
6.1 Cramer-Rao bound . . . 21
6.2 Performance of HR methods . . . 22
7 Conclusion . . . 23
8 Appendices . . . 23
8.1 Constrained optimization . . . 23
8.2 Vandermonde matrices . . . 23

2 Nonnegative matrix factorization **25**
1 Introduction . . . 25
2 NMF theory and algorithms . . . 26
2.1 Criteria for computing the NMF model parameters . . . 26
2.2 Probabilistic frameworks for NMF . . . 27
2.2.1 Gaussian noise model . . . 27
2.2.2 Probabilistic latent component analysis . . . 28

2.2.3 Poisson NMF model . . . 28
2.2.4 Gaussian composite model . . . 28
2.2.5 α-stable NMF models . . . 29
2.2.6 Choosing a particular NMF model . . . 30
2.3 Algorithms for NMF . . . 30
2.3.1 Multiplicative update rules . . . 30
2.3.2 The EM algorithm and its variants . . . 31
2.3.3 Application of the EM algorithm to PLCA . . . 32
2.3.4 Application of the space-alternating generalized EM algorithm to the Gaussian composite model . . . 32
3 Advanced NMF models . . . 33
3.1 Regularizations . . . 33
3.1.1 Sparsity . . . 33
3.1.2 Group sparsity . . . 34
3.1.3 Harmonicity and spectral smoothness . . . 34
3.1.4 Inharmonicity . . . 35
3.2 Nonstationarity . . . 35
3.2.1 Time-varying fundamental frequencies . . . 35
3.2.2 Time-varying spectral envelopes . . . 36
3.2.3 Both types of variations . . . 37
4 Summary . . . 38

Licence de droits d'usage **43**

Tutorials **44**
Exercices sur les méthodes à haute résolution . . . 44

Practical works **49**
TP Analyse et synthèse de sons de cloche . . . 49
TP Factorisations en matrices positives . . . 55

Past examination papers **59**
Written examination 2023-2024 . . . 59

List of Figures

1.1 Jean Baptiste Joseph FOURIER (1768-1830) 12
1.2 Gaspard-Marie RICHE de PRONY (1755-1839) 14
1.3 Waveform ($t \mapsto x(t)$), periodogram ($f \mapsto 20\log_{10}|S(e^{i2\pi f})|$) and pseudo-spectrum ($f \mapsto 20\log_{10}|\widehat{S}(e^{i2\pi f})|$ with $K = 20$ and $n = 256$) of a piano note 19
1.4 Joseph-Louis LAGRANGE (1736-1813) 24

2.1 Decomposition of "Au clair de la Lune" spectrogram 26
2.2 Gaussian composite model (IS-NMF) by Févotte et al. [2009] 29
2.3 Harmonic NMF model by Vincent et al. [2010] and Bertin et al. [2010] 34
2.4 Decomposition of an excerpt from the first Prelude by Johann Sebastian Bach 36
2.5 Jew's harp sound decomposed with a time-frequency activation 37

Acronyms

ACF *Auto-Covariance Function*

AIC *Akaike Information Criterion*

AR *Autoregressive*

ARMA *Autoregressive Moving Average*

DFT *Discrete Fourier Transform*

DNN *Deep Neural Networks*

EDC *Efficient Detection Criteria*

EDS *Exponentially Damped Sinusoids*

EM *Expectation-Maximization*

ERB *Equivalent Rectangular Bandwidth*

ESM *Exponential Sinusoidal Model*

ESPRIT *Estimation of Signal Parameters via Rotational Invariance Techniques*

EUC *Euclidean*

EVD *EigenValue Decomposition*

FFT *Fast Fourier Transform*

HR *High Resolution*

IS *Itakura-Saito*

ITC *Information Theoretic Criteria*

KL *Kullback-Leibler*

MA *Moving Average*

MAP *Maximum a Posteriori*

MDL *Minimum Description Length*

ML *Maxmimum Likelihood*

MM *Majorization-Minimization*

MMSE *Minimum Mean Square Error*

NMF *Non-negative Matrix Factorization*

PLCA *Probabilistic Latent Component Analysis*

SAGE *Space Alternating Generalized EM*

SNR *Signal to Noise Ratio*

STFT *Short Time Fourier Transform*

SVD *Singular Value Decomposition*

Mathematical notation

$\mathbb{N}$ set of natural numbers

$\mathbb{Z}$ set of integers

$\mathbb{R}$ set of real numbers

$\mathbb{C}$ set of complex numbers

$\mathcal{R}e(.)$ real part

$\mathcal{I}m(.)$ imaginary part

x (normal font, lower case) scalar

$\boldsymbol{x}$ (bold font, lower case) vector

$\boldsymbol{A}$ (bold font, upper case) matrix

$\|.\|_2$ Euclidean norm of a real vector, or Hermitian norm of a complex vector

$\|.\|_F$ Frobenius norm of a matrix

$\overline{(.)}$ conjugate of a matrix / vector / number

$.^\top$ transpose of a matrix

$.^H$ conjugate transpose of a matrix

Span(.) range space of a matrix

Ker(.) kernel of a matrix

dim(.) dimension of a vector space

rank(.) rank of a matrix

trace(.) trace of a square matrix

det(.) determinant of a square matrix

$.^\dagger$ pseudo-inverse of a matrix (if $\boldsymbol{A} \in \mathbb{R}^{M \times K}$ with $M \geq K$, $\boldsymbol{A}^\dagger = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top$)

diag(.) diagonal matrix formed from a vector of diagonal coefficients, or from a matrix with same diagonal entries

$\boldsymbol{I}_K$ $K \times K$ identity matrix

$\mathbf{1}_A$ is 1 if A is true, or 0 if A is false

$\mathbb{E}[.]$ expected value of a random variable or vector

$\mathbb{H}[.]$ entropy of a random variable or vector

$\widehat{(.)}$ estimator of a parameter

CRB{.} Cramér-Rao bound

Chapter 1

High resolution methods

In the context of speech and music signal processing, the tonal part of a wide variety of sounds is accurately modeled as a sum of sinusoids with slowly varying parameters. For example, the sounds that produce a well-defined perception of pitch have a quasi-periodic waveform (over a duration greater than a few tens of miliseconds). Fourier analysis shows that these signals are composed of sinusoids satisfying a relation of *harmonicity*, which means that their frequencies are multiples of the fundamental frequency, defined as the inverse of the period. This is the case of voiced speech signals produced by the quasi-periodic vibration of the vocal cords, such as vowels. Many wind or string instruments also produce harmonic or quasi-harmonic sounds. However, in a polyphonic music signal, the sounds emitted simultaneously by one or more instruments overlap; thus the harmonic relationship is no longer verified, but the signal remains essentially made up of sinusoids.

The estimation of sinusoids is a classic problem, more than two hundred years old. In this area, the Fourier transform is a privileged tool because of its robustness, the simplicity of its implementation, and the existence of fast algorithms (*Fast Fourier Transform* (FFT)). However, it has a number of drawbacks. First, its frequency *precision*, that is, the precision with which the frequency of a sinusoid can be estimated, is limited by the number of samples used to calculate it. This first limitation can be circumvented by extending the signal by a series of zeros (this operation is called *zero-padding*). However, its frequency *resolution*, that is to say its ability to distinguish two close sinusoids, remains limited by the duration of the observed signal. Despite these drawbacks, the Fourier transform remains the most used tool in spectral analysis. It has given rise to numerous frequency estimation methods Keiler and Marchand [2002].

The *High Resolution* (HR) methods, which find their applications both in antenna processing and in spectral analysis Marcos et al. [1998], have the advantage of overcoming the natural limitations of Fourier analysis. Indeed, in the absence of noise, their frequency precision and resolution are virtually infinite (although in practice limited by the finite precision of computers). This is made possible by exploiting a parametric signal model. Thus, unlike the Fourier analysis which consists in representing the signal in a transformed domain, the HR methods are parametric estimation methods. In the context of audio signal processing, despite their superiority in terms of spectral resolution (in particular over short time windows), they remain little used because of their high computational complexity. Nevertheless, the HR methods are well suited for estimating the parameters of a sum of sinusoids whose amplitudes vary exponentially (*Exponential Sinusoidal Model* (ESM) model). This type of modulation makes it possible to describe the natural damping of free vibratory systems, such as the vibration of a plucked string Jensen et al. [2004]. On the other hand, it has been shown in Laroche [1993] that the HR methods prove to be particularly effective in the case of strongly attenuated signals. More generally, the ESM model makes it possible to describe signals with large amplitude variations Hermus et al. [2002]. In addition, the music signals often contain pairs or triplets of very close frequencies which generate a beat phenomenon. These beats strongly contribute to the natural appearance of the sound. They often result from the special properties of vibration systems. For example, a minor asymmetry in the geometry of a bell leads to pairs of vibration modes. In the case of a guitar, the coupling between the strings and the bridge can be represented by a so-called mobility matrix, from which it is possible to deduce frequency pairs Lambourg and Chaigne [1993]. In the case of the piano, the coupling of the horizontal and vertical vibration modes of each string and the presence of pairs or triplets of strings for most notes explain

the presence of four or six neighboring frequencies at the level of each harmonic Weinreich [1977]. The Fourier analysis generally does not make it possible to distinguish all these frequencies. Studies in Laroche [1993] on piano and guitar sounds have shown the superiority of HR methods in this area. The same technique was used to estimate physical parameters, such as the radiation factor of a guitar David [1999], and to study the propagation of mechanical waves in solid materials Jeanneau et al. [1998].

1 Introduction

This chapter is devoted to the parametric estimation of a signal composed of a sum of exponentially modulated sinusoids, perturbed by an additive noise. The maximum likelihood principle then reduces the estimation of amplitudes and phases to a simple least squares problem, while the estimation of frequencies and damping factors requires more sophisticated methods, called *high resolution methods*, because they overcome the limits of Fourier analysis in terms of spectral resolution.

The origin of the HR methods dates back to Prony's work published in 1795, which aims to estimate a sum of exponentials by linear prediction techniques Riche de Prony [1795]. More recently, this approach was further developped by Pisarenko to estimate sinusoids of constant amplitude Pisarenko [1973]. In comparison, modern HR methods are based on the particular properties of the signal covariance matrix. Thus, the study of its rank makes it possible to separate the data space into two subspaces, the signal subspace spanned by the sinusoids, and the noise subspace which is its orthogonal complement. The HR methods resulting from this decomposition into subspaces are known to be more robust than linear prediction techniques. This is the case of the MUSIC Schmidt [1986] and root-MUSIC Barabell [1983] methods (which are based on the noise subspace), of the *Toeplitz Approximation Method* (TAM) algorithm Kung et al. [1983], as well as the *Estimation of Signal Parameters via Rotational Invariance Techniques* (ESPRIT) algorithm Roy et al. [1986] and its variants TLS-ESPRIT Roy and Kailath [1987] and PRO-ESPRIT Zoltawski and Stavrinides [1989] (which are based on the signal subspace). All of these estimation methods can be applied to the ESM (Exponential Sinusoidal Model), which represents the signal as a sum of exponentially modulated sinusoids. This model is also called *Exponentially Damped Sinusoids* (EDS) when the modulation is decreasing Nieuwenhuijse et al. [1998]. Other estimation techniques have been specifically developed for the ESM model, such as the *Kumaresan and Tufts* (KT) algorithm, also called Min-Norm method Kumaresan and Tufts [1982], its modified version Li et al. [1997] (based on linear prediction), and the *Matrix Pencil* method Hua and Sarkar [1990] (based on subspaces). A more complete list of methods can be found in Van der Veen et al. [1993].

This chapter is not intended to present the HR methods exhaustively, but rather to familiarize the reader with the concepts on which they are based. This is why only some of them are presented here: the Prony, Pisarenko, MUSIC and ESPRIT methods. This presentation will start with the definition of the signal model (section 2). Then the maximum likelihood method, which makes it possible to establish a link with the Fourier transform, will be presented in section 3. Then the high resolution methods that estimate the complex poles will be introduced in section 4, and techniques for estimating the other parameters of the model will be presented in section 5. Section 6 will be devoted to the analysis of the performance of HR methods. Finally, the results of this chapter will be summarized in section 7.

2 Signal model

Consider the discrete signal model (defined for all $t \in \mathbb{Z}$)

$$\boxed{s(t) = \sum_{k=0}^{K-1} \alpha_k {z_k}^t} \tag{1.1}$$

where $K \in \mathbb{N}^*$, $\forall k \in \{0 \dots K-1\}$, $\alpha_k \in \mathbb{C}^*$, and the poles $z_k \in \mathbb{C}^*$ are pairwise distinct. In the particular case where all the poles belong to the unit circle, the signal is represented as a sum of complex sinusoids. Thus, each pole z_k is written in the form $z_k = e^{i2\pi f_k}$ where $f_k \in \mathbb{R}$ is the frequency of the sinusoid. More generally, if the poles are not on the unit circle, the sinusoids are exponentially modulated (ESM). In this case, each pole z_k is written in polar

form $z_k = e^{\delta_k}\, e^{i2\pi f_k}$, where $\delta_k \in \mathbb{R}$ is the damping factor (or attenuation rate) of the sinusoid. In particular, poles with the same polar angle and different modules are associated with the same frequency. The complex amplitudes α_k are also written in polar form $\alpha_k = a_k\, e^{i\phi_k}$, where $a_k \in \mathbb{R}_+^*$ and $\phi \in \mathbb{R}$.

In addition, the observed signal $x(t)$ can be modeled as the sum of the deterministic signal $s(t)$ defined above and of a complex centered white Gaussian noise $b(t)$ of variance σ^2. Remember that a complex centered white Gaussian noise is a sequence of *i.i.d* random variables with complex values, of probability density $p(b) = \frac{1}{\pi\sigma^2}\, e^{-\frac{|b|^2}{\sigma^2}}$. We thus obtain the relation

$$x(t) = s(t) + b(t). \tag{1.2}$$

The signal is observed over time windows of length $N \geq K$. Thus, for all $t \in \mathbb{Z}$, we consider the time window $\{t-l+1 \ldots t+n-1\}$, where the integers n and l are such that $N = n + l - 1$, and we define the vector $\boldsymbol{s}(t) = [s(t-l+1), \ldots,\ s(t+n-1)]^\top$, of dimension N. For all $z \in \mathbb{C}$, let us define $\boldsymbol{v}(z) = [1,\ z, \ldots,\ z^{N-1}]^\top$. However $\boldsymbol{s}(t) = \sum_{k=0}^{K-1} \alpha_k\, z_k^{t-l+1}\, \boldsymbol{v}(z_k)$. This equality can be rewritten in the form of a product: $\boldsymbol{s}(t) = \boldsymbol{V}^N\, \boldsymbol{D}^{t-l+1}\, \boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = [\alpha_0, \ldots,\ \alpha_{K-1}]^\top$ is a vector of dimension K, $\boldsymbol{D} = \mathrm{diag}(z_0, \ldots,\ z_{(K-1)})$ is a diagonal matrix of dimension $K \times K$, and $\boldsymbol{V}^N = [\boldsymbol{v}(z_0), \ldots,\ \boldsymbol{v}(z_{(K-1)})]$ is a Vandermonde matrix of dimensions $N \times K$: (*cf.* definition 2 in appendix 8.2 page 23)

$$\boldsymbol{V}^N = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{N-1} & {z_1}^{N-1} & \ldots & {z_{K-1}}^{N-1} \end{bmatrix}.$$

Then define the vector of amplitudes at time t, $\boldsymbol{\alpha}(t) = \boldsymbol{D}^{t-l+1}\boldsymbol{\alpha}$, so that $\boldsymbol{s}(t) = \boldsymbol{V}^N\boldsymbol{\alpha}(t)$. It is known that the determinant of the square Vandermonde matrix $\boldsymbol{V}^K$ extracted from the first K rows of $\boldsymbol{V}^N$ (remember that $N \geq K$) is (*cf.* proposition 9 in appendix 8.2 page 23)

$$\boxed{\det(\boldsymbol{V}^K) = \prod_{0 \leq k_1 < k_2 \leq K-1} (z_{k_2} - z_{k_1}).} \tag{1.3}$$

Thus, matrix $\boldsymbol{V}^N$ has full rank if and only if all the poles are distinct. The relation $\boldsymbol{s}(t) = \boldsymbol{V}^N\boldsymbol{\alpha}(t)$ therefore shows that for each time t the vector $\boldsymbol{s}(t)$ lies in the range space of matrix $\boldsymbol{V}^N$, of dimension less than or equal to K in the general case, and equal to K if all poles are distinct.

Let $\boldsymbol{b}(t) = [b(t-l+1), \ldots,\ b(t+n-1)]^\top$ be the vector containing the samples of the additive noise. It is a centered Gaussian random vector, whose covariance matrix is $\boldsymbol{R}_{bb} = \sigma^2\boldsymbol{I}_N$. Finally, let $\boldsymbol{x}(t) = [x(t-l+1), \ldots,\ x(t+n-1)]^\top$ be the vector of observed data. This vector therefore satisfies $\boldsymbol{x}(t) = \boldsymbol{s}(t) + \boldsymbol{b}(t)$. The model being posed, the analysis of the signal $s(t)$ will consist in estimating the parameters σ^2, z_0, $\ldots$, $z_{(K-1)}$ and $\boldsymbol{\alpha}(t)$. A classical parametric estimation technique, the maximum likelihood method, is applied to this model in the next section.

3 Maximum likelihood method

The maximum likelihood principle is a general parametric estimation method. It provides asymptotically efficient and unbiased estimators. This is why it is often preferred to other estimation techniques when it has a simple closed-form solution.

3.1 Application of the maximum likelihood principle to the ESM model

The maximum likelihood principle consists in maximizing the conditional probability of observing the signal x over the interval $\{t-l+1,\ \ldots,\ t+n-1\}$, given the parameters σ^2, $z_0, \ldots,\ z_{(K-1)}$ and $\boldsymbol{\alpha}(t)$ (or the natural logarithm of this probability, called *log-likelihood* of the observations). Since $\boldsymbol{x}(t) = \boldsymbol{s}(t) + \boldsymbol{b}(t)$, where $\boldsymbol{s}(t) = \boldsymbol{V}^N\boldsymbol{\alpha}(t)$ is a deterministic vector and $\boldsymbol{b}(t)$ is a centered complex Gaussian random vector of covariance matrix $\boldsymbol{R}_{bb} = \sigma^2\boldsymbol{I}_N$, $\boldsymbol{x}(t)$

is itself a complex Gaussian random vector with expected value $\boldsymbol{s}(t)$ and covariance matrix $\boldsymbol{R}_{bb}$. Remember that the probability density of such a random vector is

$$p(\boldsymbol{x}(t)) = \frac{1}{\pi^N \det(\boldsymbol{R}_{bb})} e^{-(\boldsymbol{x}(t)-\boldsymbol{s}(t))^H \boldsymbol{R}_{bb}^{-1}(\boldsymbol{x}(t)-\boldsymbol{s}(t))}.$$

So the log-likelihood of the observations is

$$L(\sigma^2, z_0 \dots z_{K-1}, \boldsymbol{\alpha}(t)) = -N \ln(\pi\sigma^2) - \frac{1}{\sigma^2} g(z_0 \dots z_{K-1}, \boldsymbol{\alpha}(t))$$

where

$$g(z_0 \dots z_{K-1}, \boldsymbol{\alpha}(t)) = \left(\boldsymbol{x}(t) - \boldsymbol{V}^N \boldsymbol{\alpha}(t)\right)^H \left(\boldsymbol{x}(t) - \boldsymbol{V}^N \boldsymbol{\alpha}(t)\right).$$

Maximizing this log-likelihood with respect to the parameters $(\sigma^2, z_0 \dots z_{K-1}, \boldsymbol{\alpha}(t))$ can be done by first minimizing g with respect to the pair $(z_0 \dots z_{K-1}, \boldsymbol{\alpha}(t))$, then by maximizing L with respect to σ. We thus obtain $\sigma^2 = \frac{1}{N} g(z_0 \dots z_{K-1}, \boldsymbol{\alpha}(t))$, i.e.

$$\boxed{\sigma^2 = \frac{1}{N} \left\|\boldsymbol{x}(t) - \boldsymbol{V}^N \boldsymbol{\alpha}(t)\right\|^2.} \tag{1.4}$$

It appears that σ^2 is estimated by calculating the power of the residual obtained by subtracting the exponentials from the observed signal.

The matrix $\boldsymbol{V}^N$ has full rank, since it has been assumed in section 2 that the poles are pairwise distinct. Thus, matrix $\boldsymbol{V}^{N\,H}\boldsymbol{V}^N$ is invertible. In order to minimize g with respect to the pair $(z_0 \dots z_{K-1}, \boldsymbol{\alpha}(t))$, we just use the decomposition

$$\begin{aligned} g(z_0 \dots z_{K-1}, \boldsymbol{\alpha}(t)) &= \boldsymbol{x}(t)^H \boldsymbol{x}(t) - \boldsymbol{x}(t)^H \boldsymbol{V}^N \left(\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N\,H}\boldsymbol{x}(t) \\ &+ \left(\boldsymbol{\alpha}(t) - \left(\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N\,H}\boldsymbol{x}(t)\right)^H \left(\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right) \left(\boldsymbol{\alpha}(t) - \left(\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N\,H}\boldsymbol{x}(t)\right). \end{aligned}$$

The last term of this equation is always non-negative, and can be zeroed by defining

$$\boxed{\boldsymbol{\alpha}(t) = \left(\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N\,H}\boldsymbol{x}(t).} \tag{1.5}$$

It appears that the vector of complex amplitudes $\boldsymbol{\alpha}(t)$ is estimated in the same way as with the ordinary least squares method.

Function g is therefore minimal when the K -tuple $(z_0 \dots z_{K-1})$ maximizes function $\mathcal{J}$ defined by

$$\boxed{\mathcal{J}(z_0, \dots, z_{(K-1)}) = \boldsymbol{x}(t)^H \boldsymbol{V}^N \left(\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N\,H}\boldsymbol{x}(t).} \tag{1.6}$$

As this optimization problem does not have a closed-form solution in the general case, it must be solved numerically. In summary, the maximum likelihood principle leads to estimating the parameters of the model in three stages:

complex poles are obtained by maximizing function $\mathcal{J}$ (equation (1.6)),

complex amplitudes are obtained by calculating the right side of equation (1.5),

the standard deviation is then given by equation (1.4).

Unfortunately, it turns out that the first step of this estimation method, which requires the optimization of a function of K complex variables, is difficult to implement, because the function to be maximized has many local maxima. In addition, it is extremely costly in terms of computation time. This is why we generally use more reliable and faster methods to estimate complex poles. However, once the poles are estimated, the maximum likelihood principle can be used to determine the complex amplitudes and the standard deviation of the noise.

3.2 Maximum likelihood and Fourier resolution

Let us now take a look at the particular case where all the poles are on the unit circle ($\forall k,\ \delta_k = 0$). The results of section 3.1 showed that the maximum likelihood principle leads to an optimization problem which does not have a simple closed-form solution in the general case. However, such a solution exists in the particular case where $K = 1$, as well as an approximate solution if $K > 1$.

Let us first examine the case of a single complex exponential ($K = 1$). Then equation (1.6) is simplified as $\mathcal{J}(z_0) = \widehat{R}_x(z_0)$, where $\widehat{R}_x$ is the periodogram of the signal $x(t)$ observed on the time window $\{t-l+1 \dots t+n-1\}$:

$$\widehat{R}_x(e^{i2\pi f_0}) = \frac{1}{N}\left|X(e^{i2\pi f_0})\right|^2$$

where $X(e^{i2\pi f_0}) = \boldsymbol{v}(e^{i2\pi f_0})^H \boldsymbol{x}(t) = \sum_{\tau=0}^{N-1} x(t-l+1+\tau)\, e^{-i2\pi f_0 \tau}$. Similarly, equation (1.5) is simplified as $\alpha_0(t) = \frac{1}{N} X(e^{i2\pi f_0})$. Finally, equation (1.4) is simplified as $\sigma^2 = \frac{1}{N}\left(\|\boldsymbol{x}(t)\|^2 - \widehat{R}_x(e^{i2\pi f_0})\right)$.

These results lead to the following conclusion:

> The maximum likelihood principle leads in the case of a complex sinusoid to detect the frequency at which the periodogram reaches its maximum. The corresponding complex amplitude is proportional to the value of the *Discrete Fourier Transform* (DFT) of the signal at this frequency. The noise variance is estimated as the signal average power after subtracting the sinusoid.

Let us now address the general case $K \geq 1$, for which the maximization of function $\mathcal{J}(z)$ no longer has an exact closed-form solution. We then introduce the following hypothesis:

$$N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left|f_{k_2} - f_{k_1}\right|}.$$

Matrix $\boldsymbol{V}^{N\,H}\boldsymbol{V}^N$ is a positive definite Hermitian matrix of dimension $K \times K$, whose entries can be calculated in closed-form: $\left\{\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right\}_{(k_1,k_2)} = \sum_{\tau=0}^{N-1} (\overline{z_{k_1}} z_{k_2})^\tau$. We then obtain

$$\begin{array}{lcll} \frac{1}{N}\left\{\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right\}_{(k_1,k_2)} & = & e^{i\pi(N-1)(f_{k_2}-f_{k_1})}\, \frac{\sin(\pi N(f_{k_2}-f_{k_1}))}{N\sin(\pi(f_{k_2}-f_{k_1}))} & \text{if } k_1 \neq k_2 \\ \frac{1}{N}\left\{\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right\}_{(k,k)} & = & 1 & \text{if } k_1 = k_2 = k \end{array}$$

Therefore, when $N >> \frac{1}{\min\limits_{k_1 \neq k_2} \left|f_{k_2} - f_{k_1}\right|}$, $\frac{1}{N}\boldsymbol{V}^{N\,H}\boldsymbol{V}^N = \boldsymbol{I}_K + O\left(\frac{1}{N}\right)$, thus

$$\left(\boldsymbol{V}^{N\,H}\boldsymbol{V}^N\right)^{-1} = \frac{1}{N}\boldsymbol{I}_K + O\left(\frac{1}{N^2}\right).$$

Then equation (1.6) is simplified as

$$\mathcal{J}(z_0, \dots, z_{K-1}) = \frac{1}{N}\left\|\boldsymbol{V}^{N\,H}\boldsymbol{x}(t)\right\|^2 + O\left(\frac{1}{N^2}\right) = \sum_{k=0}^{K-1} \widehat{R}(z_k) + O\left(\frac{1}{N^2}\right).$$

Similarly, equation (1.5) is simplified as $\boldsymbol{\alpha}(t) = \frac{1}{N}\boldsymbol{V}^{N\,H}\boldsymbol{x}(t) + O\left(\frac{1}{N^2}\right)$, hence

$$\alpha_k(t) = \frac{1}{N} X(e^{i2\pi f_k}) + O\left(\frac{1}{N^2}\right).$$

Finally, equation (1.4) is simplified as $\sigma^2 = \frac{1}{N}\left(\|\boldsymbol{x}(t)\|^2 - \sum_{k=0}^{K-1} \widehat{R}(e^{i2\pi f_k})\right) + O\left(\frac{1}{N^2}\right)$.

Thus, the joint maximization of $\mathcal{J}$ with respect to $z_0, \ldots, z_{K-1}$ leads to determine the K frequencies corresponding to the K largest values of the periodogram. The corresponding complex amplitudes are proportional to the value of the DFT of the signal at these frequencies. Remember that these results are only valid if all the poles are on the unit circle and are based on the assumption $N >> \frac{1}{\min\limits_{k_1 \neq k_2} |f_{k_2} - f_{k_1}|}$.

We thus observe the limit of the Fourier analysis in terms of spectral resolution: the parameters are estimated correctly provided that the length of the observation window is sufficiently large compared to the inverse of the smallest frequency difference between two neighboring poles. It is this limit that the HR methods presented in section 4 allow to overcome. So, HR methods are able to distinguish two close sinusoids, that Fourier analysis does not allow to distinguish. In applications, HR methods can be used with shorter windows than those usually used with Fourier analysis.

Figure 1.1: Jean Baptiste Joseph FOURIER (1768-1830)

4 High resolution methods

We begin here by introducing the oldest high-resolution methods, which are based on linear prediction techniques (section 4.1), before addressing in section 4.2 the more recent subspace methods.

4.1 Linear prediction techniques

The first two high-resolution methods presented in this chapter are based on a fundamental result related to linear recurrence equations, presented in section 4.1.1.

4.1.1 Linear recurrence equations

Let $p_0 \in \mathbb{C}^*$, $K \in \mathbb{N}^*$, and $\{z_0, \ldots, z_{K-1}\}$ be K distinct and non-zero complex numbers. We define the polynomial of degree K whose dominant coefficient is p_0 and whose roots are z_k:

$$P[z] = p_0 \prod_{k=0}^{K-1} (z - z_k) = \sum_{\tau=0}^{K} p_{K-\tau}\, z^{\tau}.$$

The following theorem characterizes the signal model.

Theorem 1. *A complex discrete signal $\{s(t)\}_{t\in\mathbb{Z}}$ satisfies the recurrence equation*

$$\boxed{\sum_{\tau=0}^{K} p_\tau\, s(t-\tau) = 0} \tag{1.7}$$

for all $t \in \mathbb{Z}$ if and only if there are scalars $\alpha_0, \ldots, \alpha_{K-1} \in \mathbb{C}$ such that $s(t) = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$.

Proof. First of all, it is straightforward to check that the set of signals which satisfy the relation (1.7) forms a vector space E over $\mathbb{C}$. Next, we will prove that this vector space has dimension less than or equal to K. Consider the application

$$\begin{array}{rccc} f: & E & \to & \mathbb{C}^K \\ & s[t] & \mapsto & [s[0], \ldots, s[K-1]]^\top \end{array}$$

We can notice that f is a linear map. Let $s \in E$ be a signal such that $f(s) = \mathbf{0}$. Then s is zero over the interval $[0, K-1]$. By using the recurrence (1.7), we deduce that s is also zero over the interval $[K, +\infty[$. Finally, using the recurrence (1.7) and the fact that $p_K \neq 0$, we show that s is also zero over the interval $]-\infty, -1]$. Consequently, $s \equiv 0$, so the linear map f is injective. We conclude that the vector space E is at most of dimension K.

Now we will show that any signal of the form $s[t] = {z_k}^t$ where $k \in \{0, \ldots, K-1\}$ belongs to the vector space E. Indeed, if $s[t] = {z_k}^t$, then $\forall t \in \mathbb{Z}$, $\sum_{k=0}^{K} p_k\, s[t-k] = {z_k}^{t-K} \sum_{k=0}^{K} p_k\, {z_k}^{K-k} = {z_k}^{t-K}\, P[z_k] = 0$, therefore $s[t]$ satisfies the relationship (1.7).

Finally, consider the family of vectors $\{{z_k}^t\}_{\{k\in\{0,\ldots,K-1\}\}}$. The square matrix whose columns are extracted from these vectors and whose rows correspond to times $\{0 \ldots K-1\}$ is a Vandermonde matrix (*cf.* definition 2 of the appendix 8.2 page 23). According to proposition 9 in appendix 8.2, it is invertible, since the poles z_k are pairwise distinct. Therefore, the family $\{{z_k}^t\}_{\{k\in\{0,\ldots,K-1\}\}}$ is linearly independent. However it contains precisely K vectors of E. This vector space is therefore exactly of dimension K, and this family forms a basis of it. Thus, a signal $s[t]$ belongs to E if and only if it is of the form (1.2). □ □

4.1.2 Prony method

The work of Baron de Prony is at the origin of the development of high resolution methods. He proposed an estimation method inspired by the previous result on the linear recurrence equations Riche de Prony [1795]. This method was originally intended to estimate noiseless real exponentials; however we apply it here to the estimation of noisy complex exponentials. Prony's method consists in first determining the polynomial $P[z]$ using linear prediction techniques, then extracting the roots of this polynomial. We define the prediction error

$$\varepsilon(t) \triangleq \sum_{\tau=0}^{K} p_\tau\, x(t-\tau). \tag{1.8}$$

In particular, by substituting equations (1.2) and (1.7) into equation (1.8), we get $\varepsilon(t) = \sum_{\tau=0}^{K} p_\tau\, b(t-\tau)$. The prediction error therefore characterizes only the noise which is superimposed on the signal. Let us address the

Figure 1.2: Gaspard-Marie RICHE de PRONY (1755-1839)

particular case $n = K+1$, and suppose that $l \geq K+1$. Thus, the signal is observed on the window $\{t-l+1 \ldots t+K\}$. By applying equation (1.8) at times $\{t-l+K+1,\ t-l+K+2,\ \ldots,\ t+K\}$, we get the system of equations

$$\left\{\begin{array}{ccccccccc} p_0\,x(t-l+K+1) & + & p_1\,x(t-l+K) & + & \ldots & + & p_K\,x(t-l+1) & = & \varepsilon(t-l+K+1) \\ p_0\,x(t-l+K+2) & + & p_1\,x(t-l+K+1) & + & \ldots & + & p_K\,x(t-l+2) & = & \varepsilon(t-l+K+2) \\ \vdots & + & \vdots & + & \ldots & + & \vdots & = & \vdots \\ p_0\,x(t+K) & + & p_1\,x(t+K-1) & + & \ldots & + & p_K\,x(t) & = & \varepsilon(t+K) \end{array}\right. \tag{1.9}$$

Then define $\boldsymbol{p} = [p_K,\ p_{(K-1)}, \ldots, p_0]^H$, $\boldsymbol{\varepsilon}(t) = [\varepsilon(t-l+K+1), \varepsilon(t-l+K+2), \ldots, \varepsilon(t+K)]^H$ and

$$\boldsymbol{X}(t) = \left[\begin{array}{cccc} x(t-l+1) & \cdots & x(t-1) & x(t) \\ x(t-l+2) & \cdots & x(t) & x(t+1) \\ \vdots & \cdots & \vdots & \vdots \\ x(t-l+K+1) & \cdots & x(t+K-1) & x(t+K) \end{array}\right] \tag{1.10}$$

so that the system of equations (1.9) can be condensed in the form $\boldsymbol{p}^H \boldsymbol{X}(t) = \boldsymbol{\varepsilon}(t)^H$.

Prony's method consists in minimizing the power of the prediction error $\frac{1}{l}\|\boldsymbol{\varepsilon}\|^2$ with respect to $\boldsymbol{p}$, subject to the constraint $p_0 = 1$. However it is possible to write $\frac{1}{l}\|\boldsymbol{\varepsilon}\|^2 = \boldsymbol{p}^H \widehat{\boldsymbol{R}}_{xx}(t)\,\boldsymbol{p}$, where matrix $\widehat{\boldsymbol{R}}_{xx}(t) = \frac{1}{l}\,\boldsymbol{X}(t)\,\boldsymbol{X}(t)^H$ has dimension $(K+1)\times(K+1)$. As matrix $\boldsymbol{X}(t)$ has $K+1$ rows and $l \geq K+1$ columns, we can assume that matrix $\widehat{\boldsymbol{R}}_{xx}(t)$ is invertible.

Theorem 8 in appendix 8.1 page 23 allows to prove [1] that the solution of this optimization problem is [2]

$$\boldsymbol{p} = \frac{1}{\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1}\, \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$$

[1] As the data are complex, it is necessary to decompose the vector $\boldsymbol{p}$ into its real part and its imaginary part to be able to apply theorem 8, which deals exclusively with the real data case.

[2] The scalar $\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$ is non-zero, since the vector $\boldsymbol{e}_1$ is unitary and matrix $\widehat{\boldsymbol{R}}_{xx}(t)$ is positive definite.

where $\boldsymbol{e}_1 \triangleq [1,\, 0 \ldots 0]^\top$ is a vector of dimension $K+1$. Thus, the Prony estimation method includes the following steps:

- Construct matrix $\boldsymbol{X}(t)$ and calculate $\widehat{\boldsymbol{R}}_{xx}(t)$;
- Compute $\boldsymbol{p} = \frac{1}{\boldsymbol{e}_1^H \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1} \widehat{\boldsymbol{R}}_{xx}(t)^{-1} \boldsymbol{e}_1$;
- Determine the poles $\{z_0, \ldots, z_{K-1}\}$ as the roots of the polynomial $P[z] = \sum_{k=0}^{K} p_k\, z^{K-k}$.

4.1.3 Pisarenko method

The Pisarenko method is a variant of the Prony method. It consists in minimizing the power of the prediction error $\frac{1}{l}\|\boldsymbol{\varepsilon}\|^2 = \boldsymbol{p}^H \widehat{\boldsymbol{R}}_{xx}(t)\, \boldsymbol{p}$ subject to the constraint that vector $\boldsymbol{p}$ has norm 1. Theorem 8 in appendix 8.1 page 23 allows us to prove that the solution of this optimization problem is the eigenvector of matrix $\widehat{\boldsymbol{R}}_{xx}(t)$ associated with the smallest eigenvalue.

Thus the Pisarenko method Pisarenko [1973] consists of the following stages:

- calculate and diagonalize $\widehat{\boldsymbol{R}}_{xx}(t)$;
- determine $\boldsymbol{p}$ as the eigenvector associated with the smallest eigenvalue;
- extract the roots of polynomial $P[z]$.

The Prony and Pisarenko methods are the oldest HR methods. As we will show in section 6.2, they do not prove to be very robust in practice, this is why the subspace methods, proposed more recently, are generally preferred to them.

4.2 Subspace methods

In the same spirit as the Pisarenko method, modern HR methods (for example Schmidt [1986], Roy et al. [1986], Hua and Sarkar [1990]) are based on a decomposition of matrix $\widehat{\boldsymbol{R}}_{xx}(t)$.

4.2.1 Singular structure of the data matrix

Suppose now that $n \geq K+1$ and $l \geq K+1$, and construct the data matrix of the noiseless signal $s(t)$ on the same model as matrix $\boldsymbol{X}(t)$ in equation (1.10), according to a Hankel structure:

$$\boldsymbol{S}(t) = \begin{bmatrix} s(t-l+1) & \cdots & s(t-1) & s(t) \\ s(t-l+2) & \cdots & s(t) & s(t+1) \\ \vdots & \cdots & \vdots & \vdots \\ s(t-l+n) & \cdots & s(t+n-2) & s(t+n-1) \end{bmatrix}. \tag{1.11}$$

The following proposition characterizes the signal model.

Proposition 2 (Factorization of the data matrix)**.** *The following assertions are equivalent:*

1. *The signal $s(t)$ satisfies the model defined in equation* (1.1) *on the interval $\{t-l+1, \ldots, t+n-1\}$;*
2. *The matrix $\boldsymbol{S}(t)$ defined in equation* (1.11) *can be factorized as*

$$\boxed{\boldsymbol{S}(t) = \boldsymbol{V}^n\, \boldsymbol{A}(t)\, \boldsymbol{V}^{l\top}} \tag{1.12}$$

where the diagonal matrix $\boldsymbol{A}(t) = \mathrm{diag}(z_0^{t-l+1}\alpha_0, \ldots, z_{(K-1)}^{t-l+1}\alpha_{(K-1)})$ has dimension $K \times K$, $\boldsymbol{V}^n$ has dimension $n \times K$, and $\boldsymbol{V}^l$ has dimensions $l \times K$.

Proof of Proposition 2: Let us prove each of the two implications.

Proof of 1. ⇒ 2. The signal $s(t)$ is defined as a sum of complex exponentials:$s(t) = \sum_{k=0}^{K-1} s_k(t)$, where $s_k(t) = \alpha_k {z_k}^t$. Consequently, matrix $\boldsymbol{S}(t)$ defined in equation (1.11) can be decomposed in the same way: $\boldsymbol{S}(t) = \sum_{k=0}^{K-1} \boldsymbol{S}_k(t)$, where matrices $\boldsymbol{S}_k(t)$ are constructed in the same way as $\boldsymbol{S}(t)$ in equation (1.11), from the signals $s_k(t)$. However it is straightforward to check that $\boldsymbol{S}_k(t) = \alpha_k z_k^{t-l+1} \boldsymbol{v}^n(z_k) \boldsymbol{v}^l(z_k)^\top$, where $\boldsymbol{v}^n(z) = [1, z, \ldots, z^{n-1}]^\top$ and $\boldsymbol{v}^l(z) = [1, z, \ldots, z^{l-1}]^\top$. Equation (1.12) follows directly.

Proof of 1. ⇒ 2. If matrix $\boldsymbol{S}(t)$ is defined by equation (1.12), the reverse reasoning allows to show that $\boldsymbol{S}(t)$ can be written in the form (1.11), where $s(t)$ is defined in equation (1.1). □ □

Proposition 2 induces a result on which all subspace methods are based:

Corollary 3. *Matrix $\boldsymbol{S}(t)$ defined in equation* (1.11) *has rank less than or equal to K. More precisely, it has rank K if and only if $n \geq K$, $l \geq K$, all the poles z_k are distinct and non-zero, and all the amplitudes α_k are non-zero. In this case, its range space is spanned by matrix $\boldsymbol{V}^n$.*

Proof of Corollary 3: It turns out that matrices $\boldsymbol{V}^n$ and $\boldsymbol{V}^l$ have full rank equal to K. Indeed, if $\boldsymbol{V}_0$ is the Vandermonde matrix made up of the first K rows of $\boldsymbol{V}^n$ or $\boldsymbol{V}^l$, proposition 9 shows that $\boldsymbol{V}_0$ is invertible, since the poles z_k are pairwise distinct. Consequently, $\text{rank}(\boldsymbol{V}^n) \geq K$ and $\text{rank}({\boldsymbol{V}^l}) \geq K$. However $\dim(\boldsymbol{V}^n) = n \times K$ and $\dim(\boldsymbol{V}^l) = l \times K$, therefore $\text{rank}(\boldsymbol{V}^n) = \text{rank}({\boldsymbol{V}^l}) = K$. Furthermore, matrix $\boldsymbol{A}(t)$, of dimensions $K \times K$, is invertible, hence of rank K.

We can deduce from this remark that matrix $\boldsymbol{S}(t)$ is also of rank K. To do this, let us first show that $\text{Ker}(\boldsymbol{S}(t)) = \text{Ker}({\boldsymbol{V}^l}^\top)$. Indeed,

- $\forall \boldsymbol{y} \in \mathbb{C}^n$, ${\boldsymbol{V}^l}^\top \boldsymbol{y} = \boldsymbol{0} \Rightarrow \boldsymbol{V}^n \boldsymbol{A}(t) {\boldsymbol{V}^l}^\top \boldsymbol{y} = \boldsymbol{0}$;
- $\forall \boldsymbol{y} \in \mathbb{C}^n$, $\boldsymbol{V}^n \boldsymbol{A}(t) {\boldsymbol{V}^l}^\top \boldsymbol{y} = \boldsymbol{0} \Rightarrow \left(\boldsymbol{V}^{nH} \boldsymbol{V}^n\right) \boldsymbol{A}(t) {\boldsymbol{V}^l}^\top \boldsymbol{y} = \boldsymbol{0}$. However matrix $\left(\boldsymbol{V}^{nH} \boldsymbol{V}^n\right) \boldsymbol{A}(t)$ is invertible, hence ${\boldsymbol{V}^l}^\top \boldsymbol{y} = \boldsymbol{0}$.

The rank-nullity theorem then implies: $\text{rank}(\boldsymbol{S}(t)) = n - \dim(\text{Ker}(\boldsymbol{S}(t))) = n - \dim(\text{Ker}({\boldsymbol{V}^l}^\top)) = \text{rank}({\boldsymbol{V}^l}^\top) = K$. □ □

The singular structure of the data matrix induces an equivalent structure for the correlation matrix, defined below.

4.2.2 Singular structure of the correlation matrix

The subspace methods are based on the particular structure of the signal correlation matrix $\boldsymbol{C}_{ss}(t) = \boldsymbol{S}(t) \boldsymbol{S}(t)^H$, and in particular on its eigensubspaces, that we will now study. Let us define $\boldsymbol{R}_{ss}(t) = \frac{1}{l} \boldsymbol{C}_{ss}(t)$. Equation (1.12) shows that

$$\boxed{\boldsymbol{R}_{ss}(t) = \boldsymbol{V}^n \boldsymbol{P}(t) \boldsymbol{V}^{nH}} \tag{1.13}$$

where

$$\boldsymbol{P}(t) = \frac{1}{l} \boldsymbol{A}(t) {\boldsymbol{V}^l}^\top \overline{\boldsymbol{V}^l} \boldsymbol{A}(t)^H \tag{1.14}$$

is a symmetric positive definite matrix. Thus, equation (1.13) shows that under the same assumptions as for $\boldsymbol{S}(t)$, matrix $\boldsymbol{R}_{ss}(t)$ has rank K.

The range space of matrix $\boldsymbol{R}_{ss}(t)$, of dimension K, is spanned by matrix $\boldsymbol{V}^n$. This vector space is called *signal subspace* in the literature.

Then let $\{\boldsymbol{w}_m\}_{m=0\ldots n-1}$ be an orthonormal basis of eigenvectors of matrix $\boldsymbol{R}_{ss}(t)$, associated to the eigenvalues $\lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1} \geq 0$. Since matrix $\boldsymbol{R}_{ss}(t)$ has only of rank K, we actually have $\lambda_m = 0 \; \forall m \geq K$. We denote by $\boldsymbol{W}(t)$ matrix $[\boldsymbol{w}_0 \ldots \boldsymbol{w}_{K-1}]$, and by $\boldsymbol{W}_\perp(t)$ matrix $[\boldsymbol{w}_K \ldots \boldsymbol{w}_{n-1}]$. We can then check that ${}^\top\mathcal{I}\text{m}(\boldsymbol{W}(t)) = {}^\top\mathcal{I}\text{m}(\boldsymbol{V}^n)$.

Indeed, vectors $\{\boldsymbol{w}_m\}_{m=0\ldots K-1}$ are eigenvectors of matrix $\boldsymbol{R}_{ss}(t) = \boldsymbol{V}^n \boldsymbol{P}(t) \boldsymbol{V}^{nH}$ associated with nonzero eigenvalues. So $\forall\, k \in \{0 \ldots K-1\}$, $\boldsymbol{w}_m \in {}^{\top}\mathcal{I}\mathrm{m}(\boldsymbol{V}^n)$. Thus ${}^{\top}\mathcal{I}\mathrm{m}(\boldsymbol{W}(t)) \subset {}^{\top}\mathcal{I}\mathrm{m}(\boldsymbol{V}^n)$. However matrices $\boldsymbol{W}(t)$ and $\boldsymbol{V}^n$ have the same rank K, therefore ${}^{\top}\mathcal{I}\mathrm{m}(\boldsymbol{W}(t)) = {}^{\top}\mathcal{I}\mathrm{m}(\boldsymbol{V}^n)$.

Matrix $\boldsymbol{W}(t)$ is another basis of the signal subspace, generally distinct from $\boldsymbol{V}^n$.

We then define matrix $\boldsymbol{X}(t)$ from the samples of the noisy signal $x(t)$, in the same way as matrix $\boldsymbol{S}(t)$ in equation (1.11), and we consider the correlation matrix

$$\boldsymbol{C}_{xx}(t) = \boldsymbol{X}(t)\, \boldsymbol{X}(t)^H. \tag{1.15}$$

Then let $\widehat{\boldsymbol{R}}_{xx}(t) = \frac{1}{l}\, \boldsymbol{C}_{xx}(t)$ (as in section 4.1.2). Since the additive noise $b(t)$ is white and centered, of variance σ^2, matrix $\boldsymbol{R}_{xx}(t) = \mathbb{E}[\widehat{\boldsymbol{R}}_{xx}(t)]$ is such that

$$\boldsymbol{R}_{xx}(t) = \boldsymbol{R}_{ss}(t) + \sigma^2 \boldsymbol{I}_n. \tag{1.16}$$

Using equation (1.16), we show that the family $\{\boldsymbol{w}_m\}_{m=0\ldots n-1}$ defined above is also an orthonormal basis of eigenvectors of matrix $\boldsymbol{R}_{ss}(t)$, associated with the eigenvalues

$$\overline{\lambda}_m = \begin{cases} \lambda_m + \sigma^2 \ \forall m \in \{0, \ldots K-1\} \\ \sigma^2 \ \forall m \in \{K, \ldots n-1\} \end{cases}.$$

Thus, all the eigenvectors of matrix $\boldsymbol{R}_{ss}(t)$ are also eigenvectors of $\boldsymbol{R}_{xx}(t)$, and the corresponding eigenvalues of $\boldsymbol{R}_{xx}(t)$ are equal to those of $\boldsymbol{R}_{ss}(t)$ plus σ^2. Consequently, the signal subspace, defined as the range space of matrix $\boldsymbol{R}_{ss}(t)$, is also the principal subspace of dimension K of matrix $\boldsymbol{R}_{xx}(t)$, i.e. the eigensubspace of $\boldsymbol{R}_{xx}(t)$ associated with the K largest eigenvalues, all strictly greater than σ^2. The $n-K$ eigenvalues associated with the orthogonal complement of the signal subspace, called *noise subspace*, are all equal to σ^2. Thus, it is possible to estimate the signal subspace and the noise subspace by calculating the *EigenValue Decomposition* (EVD) of matrix $\widehat{\boldsymbol{R}}_{xx}(t)$, or even the *Singular Value Decomposition* (SVD) of $\boldsymbol{X}(t)$. By concatenating the K principal eigen or singular vectors of one of these matrices, we thus obtain matrix $\boldsymbol{W}(t) = [\boldsymbol{w}_0 \ldots \boldsymbol{w}_{K-1}]$ of dimensions $n \times K$ spanning the signal subspace, and by concatenating the $n-K$ other vectors, we obtain matrix $\boldsymbol{W}_{\perp}(t) = [\boldsymbol{w}_K \ldots \boldsymbol{w}_{n-1}]$ of dimensions $n \times (nK)$ spanning the noise subspace.

The idea of decomposing the data space into two subspaces (signal and noise) is the source of several high-resolution methods, including the MUSIC method, presented in section 4.2.4, and the ESPRIT method, presented in section 4.2.5.

4.2.3 Complement: analogy between the spectrum in the matrix sense and in the Fourier sense

We examine here the particular case where all poles are on the unit circle ($\forall k$, $\delta_k = 0$) and where all frequencies f_k are both multiple of $\frac{1}{n}$ and $\frac{1}{l}$ (we will consider the results that we will get as asymptotic results). We denote by $X\left(e^{i2\pi\frac{\nu}{n}}\right)$ the DFT of the signal observed on the window $\{t, \ldots, t+n-1\}$:

$$X\left(e^{i2\pi\frac{\nu}{n}}\right) = \sum_{\tau=0}^{n-1} x(t+\tau)\, e^{-i2\pi\frac{\nu}{n}}.$$

We then define the periodogram of the signal as follows

$$\widehat{R}_x\left(e^{i2\pi\frac{\nu}{n}}\right) = \frac{1}{n}\left|X\left(e^{i2\pi\frac{\nu}{n}}\right)\right|^2.$$

Since all frequencies f_k are multiple of $\frac{1}{n}$, the discrete spectrum $R_x\left(e^{i2\pi\frac{\nu}{n}}\right) \triangleq \mathbb{E}\left[\widehat{R}_x\left(e^{i2\pi\frac{\nu}{n}}\right)\right]$ is such that

$$R_x\left(e^{i2\pi\frac{\nu}{n}}\right) = \sigma^2 + \sum_{k=0}^{K-1} a_k^2\, \mathbf{1}_{\left\{e^{i2\pi\frac{\nu}{n}} = z_k\right\}}. \tag{1.17}$$

Furthermore, since all frequencies f_k are multiple of $\frac{1}{l}$, we can verify that

$$\boldsymbol{P}(t) = \text{diag}(a_0^2, \ldots, a_{(K-1)}^2).$$

Therefore, $\boldsymbol{R}_{xx}(t) = \sigma^2 \boldsymbol{I}_n + \sum_{k=0}^{K-1} a_k^2 \boldsymbol{v}(z_k)\, \boldsymbol{v}(z_k)^H$. It is then assumed without loss of generality that the a_k are sorted in decreasing order. However, the family $\{\boldsymbol{v}(z_k)\}_{k=0\ldots K-1}$ is orthogonal. Therefore $\forall m \in \{0 \ldots K-1\}$, $\boldsymbol{R}_{xx}(t)\boldsymbol{v}(z_k) = (a_k^2 + \sigma^2)\boldsymbol{v}\boldsymbol{b}(z_k)$. Thus, $\{\boldsymbol{v}(z_k)\}_{k=0\ldots K-1}$ forms a family of eigenvectors of matrix $\boldsymbol{R}_{xx}(t)$, associated to the eigenvalues $\{a_k^2 + \sigma^2\}_{k=0\ldots K-1}$. By completing this family with the other vectors of the n-th order Fourier basis, we obtain a basis of eigenvectors of matrix $\boldsymbol{R}_{xx}(t)$ (all other eigenvectors being associated with the same eigenvalue σ^2).

Consequently, the spectrum of the eigenvalues of matrix $\boldsymbol{R}_{xx}(t)$ matches the discrete spectrum given in equation (1.17) (whose periodogram forms an estimator).

4.2.4 MUltiple SIgnal Characterization (MUSIC)

The MUSIC method, developed by R. O. Schmidt Schmidt [1981], is based on the following remark: the poles $\{z_k\}_{k=0\ldots K-1}$ are the only solutions of equation

$$\|\boldsymbol{W}_\perp(t)^H \boldsymbol{v}(z)\|^2 = 0 \tag{1.18}$$

where $\boldsymbol{v}(z) = [1, z, , \ldots, z^{n-1}]^\top$. Indeed, z is solution if and only if $\boldsymbol{v}(z) \in \text{Span}(\boldsymbol{W}(t)) = \text{Span}(\boldsymbol{V}^n)$. So every pole z_k is a solution, and there can be no other because otherwise the signal subspace would be of dimension strictly larger than K. So the *root-MUSIC* Barabell [1983] method consists of the following stages:

- calculate and diagonalize matrix $\widehat{\boldsymbol{R}}_{xx}(t)$;
- deduce a basis of the noise subspace $\boldsymbol{W}_\perp(t)$;
- extract the roots of equation (1.18).

In the particular case where the noise subspace has dimension 1, it is equivalent to the Pisarenko method presented in section 4.1.3.

In practice, real signals do not strictly correspond to the model, and equation (1.18) is not rigorously verified. This is why the *spectral-MUSIC* Schmidt [1986] method rather consists in searching for the K highest peaks of function $\widehat{S}(z) = \frac{1}{\|\boldsymbol{W}_\perp^H \boldsymbol{v}(z)\|^2}$. The *spectral-MUSIC* method is illustrated in figure I.3, where it is applied to a piano note.

The ESPRIT method, presented below, avoids the optimization of function $\widehat{S}(z)$, or the resolution of equation (1.18), and provides the values of the complex poles in a more direct way.

4.2.5 Estimation of Signal Parameters via Rotational Invariance Techniques

The ESPRIT Roy et al. [1986] method is based on a particular property of the signal subspace: the rotational invariance. Let $\boldsymbol{V}_\downarrow^n$ be the matrix of dimensions $(n-1) \times K$ which contains the first $n-1$ rows of $\boldsymbol{V}^n$, and $\boldsymbol{V}_\uparrow^n$ the matrix of dimensions $(n-1) \times K$ which contains the last $n-1$ rows of $\boldsymbol{V}^n$. Similarly, let $\boldsymbol{W}(t)_\downarrow$ be the matrix of dimensions $(n-1) \times K$ which contains the first $n-1$ rows of $\boldsymbol{W}(t)$, and $\boldsymbol{W}(t)_\uparrow$ the matrix of dimensions $(n-1) \times K$ which contains the $n-1$ last rows of $\boldsymbol{W}(t)$. Then we have

$$\boxed{\boldsymbol{V}_\uparrow^n = \boldsymbol{V}_\downarrow^n \boldsymbol{D}} \tag{1.19}$$

where $\boldsymbol{D} = \text{diag}(z_0, \ldots, z_{(K-1)})$. Now the columns of $\boldsymbol{V}^n$ and those of $\boldsymbol{W}(t)$ form two bases of the same vector space of dimension K. Thus, there is an invertible matrix $\boldsymbol{G}(t)$ of dimension $K \times K$ such that

$$\boldsymbol{V}^n = \boldsymbol{W}(t)\, \boldsymbol{G}(t) \tag{1.20}$$

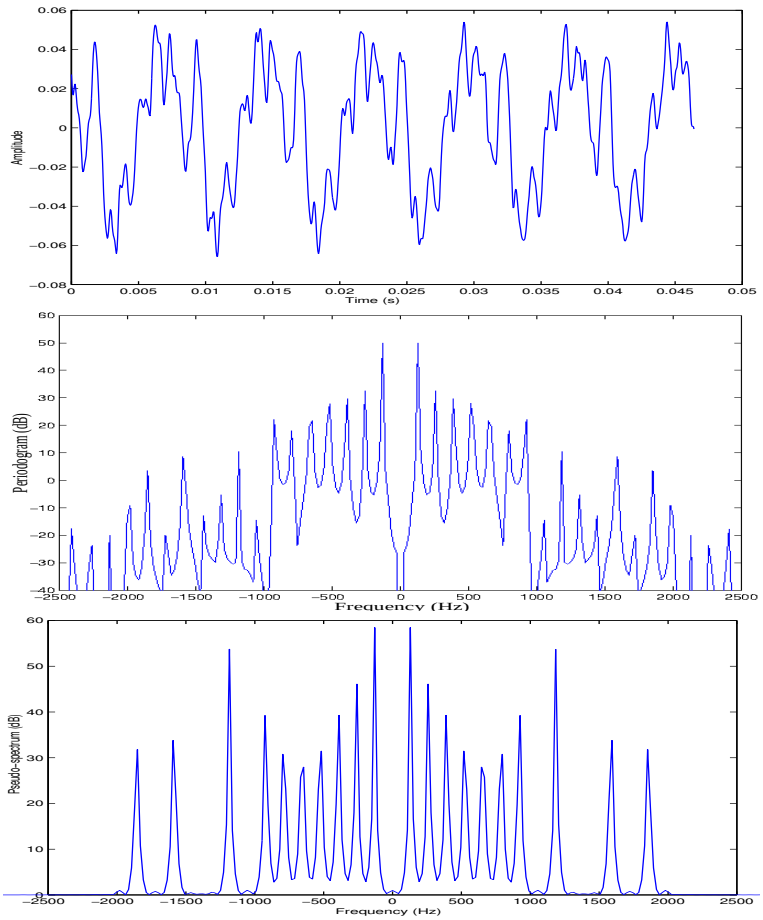


Figure 1.3: Waveform ($t \mapsto x(t)$), periodogram ($f \mapsto 20\log_{10}|S(e^{i2\pi f})|$) and pseudo-spectrum ($f \mapsto 20\log_{10}|\widehat{S}(e^{i2\pi f})|$ with $K = 20$ and $n = 256$) of a piano note

where $\boldsymbol{G}(t)$ is defined as the transition matrix from the first basis to the second one. By substituting equation (1.20) into equation (1.19), we show that

$$\boxed{\boldsymbol{W}(t)_{\uparrow} = \boldsymbol{W}(t)_{\downarrow}\,\boldsymbol{\Phi}(t)} \tag{1.21}$$

where $\boldsymbol{\Phi}(t)$, called *spectral matrix*, is defined by its EVD:

$$\boxed{\boldsymbol{\Phi}(t) = \boldsymbol{G}(t)\,\boldsymbol{D}\,\boldsymbol{G}(t)^{-1}.} \tag{1.22}$$

In particular, the eigenvalues of $\boldsymbol{\Phi}(t)$ are the poles $\{z_k\}_{k=0\ldots K-1}$.

By multiplying equation (1.21) on the left by $\boldsymbol{W}(t)_{\downarrow}^{H}$, we get

$$\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\uparrow} = \boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\downarrow}\,\boldsymbol{\Phi}(t). \tag{1.23}$$

However, if $\operatorname{rank}(\boldsymbol{W}(t)_{\downarrow}) = K$, matrix $\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\downarrow}$ is invertible. Indeed, we trivially note that $\forall \boldsymbol{x} \in \mathbb{C}^n$, $\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\downarrow}\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow \boldsymbol{W}(t)_{\downarrow}\boldsymbol{x} = 0$. So $\dim(\ker(\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\downarrow})) = \dim(\ker(\boldsymbol{W}(t)_{\downarrow}))$. The rank-nullity theorem allows us to conclude that $\operatorname{rank}(\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\downarrow}) = \operatorname{rank}(\boldsymbol{W}(t)_{\downarrow}) = K$, so $\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\downarrow}$ is invertible. Therefore equation (1.23) implies $\boldsymbol{\Phi}(t) = \left(\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\downarrow}\right)^{-1}\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\uparrow}$.

Finally, the ESPRIT algorithm is composed of four steps:

- calculate and diagonalize matrix $\widehat{\boldsymbol{R}}_{xx}(t)$;
- deduce a basis of the signal subspace $\boldsymbol{W}(t)$;
- extract from $\boldsymbol{W}(t)$ matrices $\boldsymbol{W}(t)_{\downarrow}$ and $\boldsymbol{W}(t)_{\uparrow}$;
- compute the spectral matrix $\boldsymbol{\Phi}(t) = \left(\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\downarrow}\right)^{-1}\boldsymbol{W}(t)_{\downarrow}^{H}\boldsymbol{W}(t)_{\uparrow}$
- diagonalize $\boldsymbol{\Phi}(t)$ and deduce the estimated poles.

Theoretical and experimental studies have shown that the ESPRIT method is the most efficient of the HR methods presented above. (*cf.* section 6.2).

5 Estimation of the other parameters

The high resolution methods exposed in the previous sections estimate only the poles z_k. We are now interested in the estimation of the other model parameters.

5.1 Estimation of the modeling order

Until now, the order of the ESM model was assumed to be known, which is generally not the case in practice. Many methods have been proposed in the literature for estimating the number of sinusoids present in white noise. The most classic ones are the maximum likelihood method Bienvenu and Kopp [1983] and the *Information Theoretic Criteria* (ITC) Wax and Kailath [1985], among which the *Akaike Information Criterion* (AIC) criterion Akaike [1973] and the *Minimum Description Length* (MDL) criterion by Schwartz Schwarz [1978] and Rissanen Rissanen [1978]. Another technique in the context of ITC is the *Efficient Detection Criteria* (EDC) criterion Zhao et al. [1986a], which is also robust to a multiplicative white noise Gini and Bordoni [2003]. These various ITC criteria are based on the similarity of the eigenvalues in the noise subspace, and not on the existence of a gap between the signal and noise subspaces Liavas and Regalia [2001]. A criterion for selecting the modeling order based on this gap, formulated in terms of *maximally stable* decomposition, was developed in Liavas et al. [1999]. Other approaches are based on Wishart matrices Grouffaud et al. [1996] and on the cross validation method Kundu and Mitra [2000].

However, in the case where the noise is colored, all these methods tend to overestimate the order of the model. Thus, specific methods have been designed to deal with the case of colored noise, among which new ITC criteria Zhao et al. [1986b], Zhang and Wong [1993], a technique based on a model of noise *Auto-Covariance Function* (ACF) of finite support Fuchs [1992], and a maximum a posteriori criterion Bishop and Djuric [1996].

Among all these methods, we present here the most classic ones, namely the three main ITC criteria: AIC,MDL and EDC (which is a robust generalization of AIC and MDL). These methods consist in minimizing a cost function composed of a first common term and a second term which forms a penalizing factor:

$$\boxed{\mathrm{ITC}(p) = -(n-p)\,l\,\ln\left(\frac{\left(\prod_{q=p+1}^{n}\sigma_q^2\right)^{\frac{1}{n-p}}}{\frac{1}{n-p}\sum_{q=p+1}^{n}\sigma_q^2}\right) + p\,(2n-p)\,C(l)}$$

where the scalars σ_q^2 are the eigenvalues of matrix $\widehat{\boldsymbol{R}}_{xx}(t)$ sorted in decreasing order, and $C(l)$ is a function of variable l. The AIC criterion is defined by setting $C(l) = 1$, and the MDL criterion is defined by setting $C(l) = \frac{1}{2}\ln(l)$. The EDC criteria are obtained for all functions $l \mapsto C(l)$ such that $\lim_{l\to+\infty}\frac{C(l)}{l} = 0$ and $\lim_{l\to+\infty}\frac{C(l)}{\ln(\ln(l))} = +\infty$. These criteria lead to maximizing the ratio of the geometric mean of the eigenvalues of the noise subspace to their arithmetic mean. However this ratio is maximum and equal to 1 when all these eigenvalues are equal; it therefore measures the whiteness of the noise (in theory the eigenvalues are all equal to σ^2). The penalty term $C(l)$ avoids overestimating p. In practice, these methods are relatively satisfactory when processing signals that fit well the signal model, but their performance collapses when this model is less well fitted, in particular when the noise is colored.

5.2 Estimation of amplitudes, phases and standard deviation of noise

The maximum likelihood principle developed in section 3.1 suggests using the least squares method to estimate the complex amplitudes (*cf.* equation (1.5)):

$$\boldsymbol{\alpha}(t) = \boldsymbol{V}^{N\dagger}\boldsymbol{x}(t),$$

from which $a_k = |\alpha_k|$ and $\phi_k = \arg(\alpha_k)$ are deduced. Remember that according to the Gauss-Markov theorem, the least squares estimator is an unbiased linear estimator, with minimum variance among all unbiased linear estimators, since the additive noise is white. In the case where the additive noise is colored, the optimal estimator is obtained by the weighted least squares method (we can refer to Stoica et al. [2000] for detailed information on the estimation of amplitudes by the weighted least squares method) .

Finally, the maximum likelihood principle suggests estimating the standard deviation by calculating the power of the residual (*cf.* equation (1.4)):

$$\sigma^2 = \frac{1}{N}\left\|\boldsymbol{x}(t) - \boldsymbol{V}^N\boldsymbol{\alpha}(t)\right\|^2.$$

6 Performance of the estimators

6.1 Cramer-Rao bound

The Cramér-Rao bound is a fundamental tool in probability theory, because it makes it possible to analyze the performance of an estimator, by comparing its variance to an optimal value, which in a way acts as a quality benchmark. In the particular case of the ESM signal model, a study of the Cramér-Rao bound was proposed in Hua and Sarkar [1990]. The general Cramér-Rao bound theorem is summarized below (*cf.* Kay [1993]). It is based on the assumption of a regular statistical model.

Definition 1 (Regular statistical model)**.** *Let us consider a statistical model dominated by a measure μ and parameterized by $\boldsymbol{\theta} \in \Theta$, where Θ is an open set of $\mathbb{R}^q$. Let $\boldsymbol{x}$ denote the random vector of dimension N. Then the parameterization is said to be regular if the following conditions are satisfied:*

1. *the probability density $p(\boldsymbol{x};\, \boldsymbol{\theta})$ is continuously differentiable, μ-almost everywhere, with respect to $\boldsymbol{\theta}$.*
2. *the Fisher information matrix*

$$\boldsymbol{F}(\boldsymbol{\theta}) \triangleq \int_H \boldsymbol{l}(\boldsymbol{x};\, \boldsymbol{\theta})\, \boldsymbol{l}(\boldsymbol{x};\, \boldsymbol{\theta})^\top\, p(\boldsymbol{x};\, \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{x}$$

defined from the score function $\boldsymbol{l}(\boldsymbol{x};\, \boldsymbol{\theta}) \triangleq \boldsymbol{\nabla}_{\boldsymbol{\theta}} \ln p(\boldsymbol{x};\, \boldsymbol{\theta})\, \mathbf{1}_{p(\boldsymbol{x};\, \boldsymbol{\theta})>0}$ is positive definite for any value of parameter $\boldsymbol{\theta}$, and continuous with respect to $\boldsymbol{\theta}$.

Theorem 4 (Cramér-Rao bound)**.** *Let us consider a regular statistical model parameterized by $\boldsymbol{\theta} \in \Theta$. Let $\widehat{\boldsymbol{\theta}}$ be an unbiased estimator of $\boldsymbol{\theta}$ ($\forall \boldsymbol{\theta} \in \Theta$, $\mathbb{E}_{\boldsymbol{\theta}}[\widehat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$). Then the dispersion matrix $\boldsymbol{D}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) \triangleq \mathbb{E}_{\boldsymbol{\theta}}\left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)^\top\right]$ is such that matrix $\boldsymbol{D}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) - \boldsymbol{F}(\boldsymbol{\theta})^{-1}$ is positive semidefinite.*

In particular, the diagonal entries of matrix $\boldsymbol{D}(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) - \boldsymbol{F}(\boldsymbol{\theta})^{-1}$ are non-negative. Consequently, the variances of the coefficients of $\widehat{\boldsymbol{\theta}}$ are greater than the diagonal entries of matrix $\boldsymbol{F}(\boldsymbol{\theta})^{-1}$. Thus the Cramér-Rao estimation bounds for all the scalar parameters are obtained in three stages:

- calculation of the Fisher information matrix;
- inversion of this matrix;
- extraction of its diagonal entries.

As mentioned in section 3.1, the vector $\boldsymbol{x}(t)$ containing the N samples of the observed signal is a Gaussian random vector of expected value $\boldsymbol{s}(t)$ and covariance matrix $\boldsymbol{R}_{bb}$. Below, the dependence of $\boldsymbol{s}(t)$ and $\boldsymbol{R}_{bb}$ on the parameters of the model will be mentioned explicitly. On the other hand, in order to simplify the notation,

we will omit the dependence of $\boldsymbol{s}(t)$ with respect to time. It is known that the Fisher information matrix of a Gaussian random vector is expressed simply as a function of the model parameters, as shown in the following proposition [Kay, 1993, pp. 525].

Proposition 5 (Fisher's information matrix for a Gaussian density)**.** *For a family of complex Gaussian probability laws of covariance matrix $\boldsymbol{R}_{bb}(\boldsymbol{\theta})$ and of mean $\boldsymbol{s}(\boldsymbol{\theta})$, where $\boldsymbol{R}_{bb} \in C^1(\Theta, \mathbb{C}^{N\times N})$ and $\boldsymbol{s} \in C^1(\Theta, \mathbb{C}^N)$, the entries of the Fisher information matrix $\left\{\boldsymbol{F}_{(i,j)}(\boldsymbol{\theta})\right\}_{1\leq i,\, j\leq k}$ are given by the extended Bangs-Slepian formula:*

$$\boxed{\boldsymbol{F}_{(i,j)}(\boldsymbol{\theta}) = \operatorname{trace}\left(\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial\theta_i}\boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{R}_{bb}(\boldsymbol{\theta})}{\partial\theta_j}\right) + 2\mathcal{R}e\left(\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial\theta_i}^H \boldsymbol{R}_{bb}^{-1}\frac{\partial \boldsymbol{s}(\boldsymbol{\theta})}{\partial\theta_j}\right).} \tag{1.24}$$

By applying formula (1.24) to the ESM model, we obtain a closed-form expression of the Fisher information matrix. We deduce the following theorem, proved in Hua and Sarkar [1990]:

Proposition 6. *The Cramér-Rao bounds for parameters (ϕ_k, δ_k, f_k) are independent of $a_{k'}$ for all $k' \neq k$, but proportional to $\frac{1}{a_k^2}$. The bound for parameter a_k is independent of all $a_{k'}$. Finally, the bounds for all parameters are independent of all the phases $\phi_{k'}$, and are unchanged by a translation of the set of frequencies $f_{k'}$.*

In addition, the Cramér-Rao bounds can be calculated in closed-form under certain assumptions, as was done in Rao and Zhao [1993].

Proposition 7. *Suppose that all the damping factors are zero, and let us make N tend towards $+\infty$. Then the Cramér-Rao bounds for the parameters of the ESM model admit the following first-order expansions:*

- $\mathrm{CRB}\{\sigma\} = \frac{\sigma^2}{4N} + O\left(\frac{1}{N^2}\right)$;
- $\mathrm{CRB}\{f_k\} = \frac{6\sigma^2}{4\pi^2 N^3 a_k^2} + O\left(\frac{1}{N^4}\right)$;
- $\mathrm{CRB}\{a_k\} = \frac{2\sigma^2}{N} + O\left(\frac{1}{N^2}\right)$;
- $\mathrm{CRB}\{\phi_k\} = \frac{2\sigma^2}{N a_k^2} + O\left(\frac{1}{N^2}\right)$.

We note in particular that the Cramér-Rao bounds related to the frequencies f_k are of order $\frac{1}{N^3}$, which is unusual in parametric estimation. Furthermore, it is known that the maximum likelihood principle provides asymptotically efficient estimators Kay [1993]. Thus, the variances of the estimators given in section 3.1 are asymptotically equivalent to the Cramér-Rao bounds given in proposition 7. The case of HR methods is discussed below.

6.2 Performance of HR methods

The performance of an estimator is generally expressed in terms of bias and variance. It is also possible to measure its efficiency, defined as the ratio between its variance and the Cramér-Rao bound. In particular, an estimator is said to be efficient if its efficiency is equal to 1.

In the case of HR methods, calculating the bias and variance in closed-form unfortunately turns out to be impossible, because the extraction of the roots of a polynomial, or of the eigenvalues of a matrix, induces a complex relationship between the statistics of the signal and those of the estimators. However, asymptotic results have been obtained thanks to the perturbation theory. These results are based either on the hypothesis $N \to +\infty$ (in the case where all poles are on the unit circle), or on the hypothesis of a high *Signal to Noise Ratio* (SNR) (SNR $\to +\infty$). Under each of these two hypotheses, it has been shown that all HR methods presented in this chapter are unbiased. Furthermore, under the assumption $N \to +\infty$, the variances of the Prony and Pisarenko methods were calculated in Stoica and Nehorai [1988], and those of MUSIC and ESPRIT in P. Stoica and T. Söderström [1991]. Under the hypothesis SNR$\to +\infty$, the variance of Prony's method was calculated in Kot et al. [1987], that of MUSIC in Eriksson et al. [1993], and that of ESPRIT in Hua and Sarkar [1991], Eriksson et al. [1993].

The mathematical developments proposed in all these articles are quite complex, and are strongly related to the estimation method considered, this is why they are not reproduced as part of this document. Only the main results are summarized here. First of all, it has been proved in Kot et al. [1987], Stoica and Nehorai [1988] that the Prony and Pisarenko methods are very inefficient, in the statistical sense: their variances are much greater than the Cramér-Rao bounds. In addition, they increase faster than the Cramér-Rao bounds when the SNR decreases. On the other hand, the MUSIC and ESPRIT methods have an asymptotic efficiency close to 1. More precisely, it has been proved in P. Stoica and T. Söderström [1991], Eriksson et al. [1993] (in the case of unmodulated sinusoids) that these two methods achieve almost identical performances, but ESPRIT is slightly better than MUSIC. The study carried out in Hua and Sarkar [1991] (in the more general case of exponentially modulated sinusoids) goes in the same direction: ESPRIT is less sensitive to noise than MUSIC.

7 Conclusion

In this chapter, we have shown that the estimation of frequencies and damping factors by the maximum likelihood method leads to a difficult optimization problem. When all the poles of the signal are on the unit circle, it can be approximated by detecting the K main peaks of the periodogram. This result is only valid when the length of the observation window is sufficiently large compared to the inverse of the smallest frequency difference between neighboring poles. The main interest of HR methods is that they overcome this limit of Fourier analysis in terms of spectral resolution. The first methods of this family, proposed by Prony and Pisarenko, are based on the linear recurrence equations which characterize the signal model. On the other hand, more modern techniques, including the MUSIC and ESPRIT methods, rely on the decomposition of the data space into two eigensubspaces of the covariance matrix, called signal subspace and noise subspace. The statistical study of these various estimation techniques has shown that the ESPRIT method is the most efficient. The amplitudes and phases of the complex exponentials can then be estimated by the least squares method.

8 Appendices

8.1 Constrained optimization

Let V be a Hilbert space and F be a closed subset of V defined by $F = \{\boldsymbol{y} \in V / f_1(\boldsymbol{y}) = 0 \dots f_M(\boldsymbol{y}) = 0\}$, where functions $\{f_m\}_{m=1\dots M}$ are continuous from V to $\mathbb{R}$. Let J be a real function on V and $\boldsymbol{p}$ be a local minimum of J over F. We also assume that functions J and $\{f_m\}_{m=1\dots M}$ are differentiable at $\boldsymbol{p}$.

Theorem 8 (Lagrange multipliers)**.** *There is $\mu_0, \dots \mu_M \in \mathbb{R}$ not all zero such that $\mu_0 J'(\boldsymbol{p}) + \sum_{m=1}^{M} \mu_m f_m'(\boldsymbol{p}) = 0$. If in addition the vectors $\{f_m'(\boldsymbol{p})\}_{m=1\dots M}$ are linearly independent, then there are coefficients $\lambda_1 \dots \lambda_M \in \mathbb{R}$ called Lagrange multipliers such that $J'(\boldsymbol{p}) + \sum_{m=1}^{M} \lambda_m f_m'(\boldsymbol{p}) = 0$.*

8.2 Vandermonde matrices

Definition 2 (Vandermonde matrix)**.** *Let $K \in \mathbb{N}^*$ and $z_0, \dots, z_{K-1} \in \mathbb{C}$. We call Vandermonde matrix a matrix $\boldsymbol{V}$ of dimension $K \times K$ of the form*

$$\boldsymbol{V} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ {z_0}^{K-1} & {z_1}^{K-1} & \dots & {z_{K-1}}^{K-1} \end{bmatrix}.$$

Proposition 9 (Vandermonde determinant)**.** *The determinant of the Vandermonde matrix is*

$$\det(\boldsymbol{V}) = \prod_{0 \le k_1 < k_2 \le K-1} (z_{k_2} - z_{k_1}).$$

Figure 1.4: Joseph-Louis LAGRANGE (1736-1813)

Chapter 2

Nonnegative matrix factorization

Non-negative Matrix Factorization (NMF) refers to a set of techniques that have been used to model the spectra of sound sources in various audio applications, including source separation. Sound sources have a structure in time and frequency: music consists of basic units like notes and chords played by different instruments, speech consists of elementary units such as phonemes, syllables or words, and environmental sounds consist of sound events produced by various sound sources. NMF models this structure by representing the spectra of sounds as a sum of components with fixed spectrum and time-varying gain, so that each component in the model represents these elementary units in the sound.

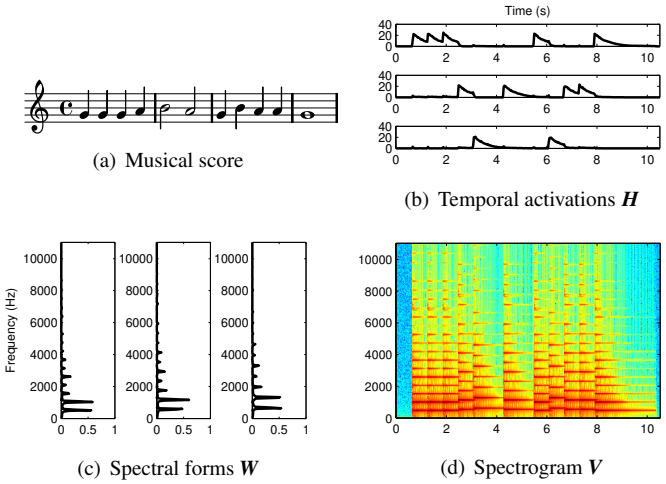
Modeling this structure is beneficial in source separation, since inferring the structure makes it possible to use contextual information for source separation. NMF is typically used to model the magnitude or power spectrogram of audio signals, and its ability to represent the structure of audio sources makes separation possible even in single-channel scenarios.

This chapter presents the use of NMF-based single-channel techniques. In Section 2, several deterministic and probabilistic frameworks for NMF are presented, along with various NMF algorithms. In Section 3, some advanced NMF models are introduced, including regularizations and nonstationary models. Finally, Section 4 summarizes the key concepts introduced in this chapter.

1 Introduction

The NMF was introduced by Lee and Sung to decompose non-negative two-dimensional data into a linear combination of elements in a dictionary [Lee and Seung, 1999].

Given a data matrix $\boldsymbol{V}$ of dimensions $F \times N$ whose coefficients are non-negative, the NMF problem consists in calculating an approximation $\widehat{\boldsymbol{V}}$ of matrix $\boldsymbol{V}$ truncated at rank $K < \min(F, N)$, expressed as a product $\widehat{\boldsymbol{V}} = \boldsymbol{W}\boldsymbol{H}$, where the two matrices $\boldsymbol{W}$ of dimensions $F \times K$ and $\boldsymbol{H}$ of dimensions $K \times N$ have non-negative entries. The columns of the matrix $\boldsymbol{W}$ form the elements of the dictionary and the rows of $\boldsymbol{H}$ contain the coefficients of the decomposition. The dimension K is generally chosen such that $FK + KN \ll FN$, so as to reduce the dimension of the data. The NMF can be considered as a supervised or unsupervised learning technique. In the case of supervised learning, the dictionary $\boldsymbol{W}$ is previously estimated from training data and matrix $\boldsymbol{H}$ only must be calculated from matrix $\boldsymbol{V}$. In the case of unsupervised learning, the two matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ must be computed jointly from $\boldsymbol{V}$. In audio applications, $\boldsymbol{V}$ is often the amplitude or power spectrogram, f denotes the frequency channel and n the time window. Figure 2.1 represents the musical score, spectrogram and unsupervised NMF of the melody of "Au clair de la lune". This figure clearly shows the interest of such a decomposition: it shows the spectra of musical notes in matrix $\boldsymbol{W}$ and their temporal activations in matrix $\boldsymbol{H}$, which makes it possible to consider both transcription and musical note separation applications.



(a) Musical score (b) Temporal activations $\boldsymbol{H}$

(c) Spectral forms $\boldsymbol{W}$ (d) Spectrogram $\boldsymbol{V}$

Figure 2.1: Decomposition of "Au clair de la Lune" spectrogram (from [Bertin, 2009, pp. 40–41])

2 NMF theory and algorithms

Let $\boldsymbol{V} = [v(n,f)]_{fn}$ denote the $F \times N$ nonnegative time-frequency representation of a signal $x(t)$, where $n \in \{0,\dots,N-1\}$ is the time frame index, and $f \in \{0,\dots,F-1\}$ is the frequency index. For instance, if $\boldsymbol{X}$ is the $F \times N$ complex-valued *Short Time Fourier Transform* (STFT) of x, then $\boldsymbol{V}$ can be the magnitude spectrogram $|\boldsymbol{X}|$ or the power spectrogram $|\boldsymbol{X}|^2$ [Smaragdis and Brown, 2003]. Other choices may include perceptual frequency scales, such as constant-Q [Fuentes et al., 2013] or *Equivalent Rectangular Bandwidth* (ERB) [Vincent et al., 2010] representations.

NMF [Lee and Seung, 1999] approximates the nonnegative matrix $\boldsymbol{V}$ with another nonnegative matrix $\widehat{\boldsymbol{V}} = [\widehat{v}(n,f)]_{fn}$ with entries $v(n,f) = \sigma_x^2(n,f)$, defined as the product

$$\widehat{\boldsymbol{V}} = \boldsymbol{W}\boldsymbol{H} \tag{2.1}$$

of a $F \times K$ nonnegative matrix $\boldsymbol{W}$ and a $K \times N$ nonnegative matrix $\boldsymbol{H}$ of lower rank $K < \min(F,N)$. This factorization can also be written $\widehat{\boldsymbol{V}} = \sum_k \widehat{\boldsymbol{V}}_k$, where $\widehat{\boldsymbol{V}}_k = [\widehat{v}_k(n,f)]_{fn} = \boldsymbol{w}_k \boldsymbol{h}_k^\top$, for all $k \in \{1,\dots,K\}$, is the k-th rank-1 matrix component. The k-th column vector $\boldsymbol{w}_k = [w_k(f)]_f$ can be interpreted as its spectrum, and the k-th row vector $\boldsymbol{h}_k^\top = [h_k(n)]_n$ comprises its *activation coefficients* over time. We also write $\widehat{v}_k(n,f) = w_k(f)h_k(n)$. All the parameters of the model, as well as the observed magnitude or power spectra, are elementwise nonnegative.

In this section, we first present the standard criteria for computing the NMF model parameters (Section 2.1), then we introduce probabilistic frameworks for NMF (Section 2.2), and we describe several algorithms designed for computing an NMF (Section 2.3).

2.1 Criteria for computing the NMF model parameters

Since NMF is a rank reduction technique, it involves an approximation: $\widehat{\boldsymbol{V}} \approx \boldsymbol{V}$. Computing the NMF can thus be formalized as an optimization problem: we want to minimize a measure $C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}})$ of divergence between matrices $\boldsymbol{V}$ and $\widehat{\boldsymbol{V}}$. The most popular measures in the NMF literature include the squared *Euclidean* (EUC) distance [Lee and Seung, 1999], the *Kullback-Leibler* (KL) divergence [Lee and Seung, 2001], and the *Itakura-Saito* (IS) divergence [Févotte et al., 2009]. The various NMFs computed by minimizing each of these three measures are named accordingly: EUC-NMF, KL-NMF, and IS-NMF. Actually, these three measures fall under the umbrella of *β-divergences* [Nakano et al., 2010, Févotte and Idier, 2011]. Formally, they are defined for any real-valued β as

$$C^\beta(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) = \sum_{nf} d^\beta(v(n,f) \mid \widehat{v}(n,f)), \tag{2.2}$$

where

- $\forall \beta \notin \{0,1\},\ d^{\beta}(x \mid y) = \frac{1}{\beta(\beta-1)}\left(x^{\beta} + (\beta-1)y^{\beta} - \beta x y^{\beta-1}\right)$,
- $\beta = 2$ corresponds to the squared EUC distance: $d^{\mathrm{EUC}}(x \mid y) = \frac{1}{2}|x-y|^2$,
- $\beta = 1$ corresponds to the KL divergence: $d^{\mathrm{KL}}(x \mid y) = x\log(\frac{x}{y}) - x + y$,
- $\beta = 0$ corresponds to the IS divergence: $d^{\mathrm{IS}}(x \mid y) = \frac{x}{y} - \log(\frac{x}{y}) - 1$.

It can be easily proved that $\forall x > 0$, the function $y \mapsto d^{\beta}(x \mid y)$ is convex with respect to y if and only if $\beta \in [1,2]$ [Févotte and Idier, 2011]. This means that minimizing $C^{\beta}(\boldsymbol{V} \mid \boldsymbol{WH})$ with respect to $\boldsymbol{H}$ with $\boldsymbol{W}$ fixed, or conversely with respect to $\boldsymbol{W}$ with $\boldsymbol{H}$ fixed, is a convex optimization problem if and only if $\beta \in [1,2]$. This convexity property is particularly convenient in a pretrained framework, where matrix $\boldsymbol{W}$ is fixed, and only matrix $\boldsymbol{H}$ is estimated from the observed data. Optimization algorithms are then insensitive to initialization, which might explain the better success of KL-NMF and EUC-NMF compared with IS-NMF in the NMF literature.

However, in the context of learning-free separation, whatever the value of β, minimizing $C^{\beta}(\boldsymbol{V} \mid \boldsymbol{WH})$ jointly with respect to $\boldsymbol{W}$ and $\boldsymbol{H}$ is not a convex optimization problem. Indeed, this factorization is not unique, since we also have $\widehat{\boldsymbol{V}} = \boldsymbol{W}'\boldsymbol{H}'$ with $\boldsymbol{W}' = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{\Pi}$ and $\boldsymbol{H}' = \boldsymbol{\Pi}^{\top}\boldsymbol{\Lambda}^{-1}\boldsymbol{H}$, where $\boldsymbol{\Lambda}$ can be any $K \times K$ diagonal matrix with positive diagonal entries, and $\boldsymbol{\Pi}$ can be any $K \times K$ permutation matrix. Note that the nonuniqueness of the model is actually ubiquitous in source separation and is generally not considered as a problem: sources are recovered up to a scale factor and a permutation. In the case of NMF however, other kinds of indeterminacies may also exist [Laurberg et al., 2008]. Due to the existence of local minima, optimization algorithms become sensitive to initialization (cf. Section 2.3). In practice, with a random initialization, there is no longer any guarantee to converge to a solution that is helpful for source separation. For this reason, advanced NMF models have later been proposed to enforce some specific desired properties in the decomposition (cf. Section 3).

2.2 Probabilistic frameworks for NMF

Computing an NMF can also be formalized as a parametric estimation problem based on a probabilistic model, that involves both observed and latent (hidden) random variables. Typically, observed variables are related to matrix $\boldsymbol{V}$, whereas latent variables are related to matrices $\boldsymbol{W}$ and $\boldsymbol{H}$. The main advantages of using a probabilistic framework are the facility of exploiting some a priori knowledge that we may have about matrices $\boldsymbol{W}$ and $\boldsymbol{H}$, and the existence of well-known statistical inference techniques, such as the *Expectation-Maximization* (EM) algorithm.

Popular probabilistic models of nonnegative time-frequency representations include Gaussian models that are equivalent to EUC-NMF [Schmidt and Laurberg, 2008] (Section 2.2.1), and count models, such as the celebrated *Probabilistic Latent Component Analysis* (PLCA) [Shashanka et al., 2008] (Section 2.2.2) and the Poisson NMF model based on the Poisson distribution [Virtanen et al., 2008] (Section 2.2.3), that are both related to KL-NMF. However these probabilistic models do not account for the fact that matrix $\boldsymbol{V}$ has been generated from a time-domain signal $x(t)$. As a result, they can be used to estimate a nonnegative time-frequency representation, but they are not able to account for the phase, that is necessary to reconstruct a time-domain signal.

Other probabilistic frameworks focus on power or magnitude spectrograms, and intend to directly model the STFT $\boldsymbol{X}$ instead of the nonnegative time-frequency representation $\boldsymbol{V}$, in order to permit the resynthesis of time-domain signals. The main advantage of this approach is the ability to account for the phase and, in source separation applications, to provide a theoretical ground for using time-frequency masking techniques. Such models include Gaussian models that are equivalent to IS-NMF [Févotte et al., 2009] (Section 2.2.4) and the Cauchy NMF model based on the Cauchy distribution [Liutkus et al., 2015], that both fall under the umbrella of α-stable models [Liutkus and Badeau, 2015] (Section 2.2.5).

2.2.1 Gaussian noise model

A simple probabilistic model for EUC-NMF was presented by Schmidt and Laurberg [2008]: $\boldsymbol{V} = \boldsymbol{WH} + \boldsymbol{U}$, where matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ are seen as deterministic parameters, and the entries of matrix $\boldsymbol{U}$ are Gaussian, independent and identically distributed (i.i.d.): $u(n,f) \sim \mathcal{N}(u(n,f) \mid 0, \sigma_u^2)$. Then the log-likelihood of matrix $\boldsymbol{V}$ is $\log p(\boldsymbol{V} \mid$

$\boldsymbol{W}, \boldsymbol{H}) = -\frac{1}{2\sigma_u^2}\|\boldsymbol{V} - \boldsymbol{W}\boldsymbol{H}\|_F^2 + \text{cst} = -\frac{1}{\sigma_u^2}C^2(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) + \text{cst}$ where cst denotes a constant additive term, as defined in (2.2) with $\beta = 2$. Therefore *Maxmimum Likelihood* (ML) estimation of $(\boldsymbol{W}, \boldsymbol{H})$ is equivalent to EUC-NMF. The main drawback of this generative model is that it does not enforce the nonnegativity of $\boldsymbol{V}$, whose entries might take negative values.

2.2.2 Probabilistic latent component analysis

PLCA [Shashanka et al., 2008] is a count model that views matrix $\widehat{\boldsymbol{V}}$ as a probability distribution (normalized so that $\sum_{nf}\widehat{v}(n,f) = 1$). The observation model is the following one: the probability distribution $P(n,f) = \widehat{v}(n,f)$ is sampled M times to produce M independent time-frequency pairs (n_m, f_m), $m \in \{1,\dots,M\}$. Then matrix $\boldsymbol{V}$ is generated as a histogram: $v(n,f) = \frac{1}{M}\sum_m \delta_{(n_m,f_m)}(n,f)$, that also satisfies $\sum_{nf}\widehat{z}(n,f) = 1$. The connection with NMF is established by introducing a latent variable k that is also sampled M times to produce k_m, $m \in \{1,\dots,M\}$. More precisely, it is assumed that (k_m, n_m) are first sampled together according to distribution $P(k,n) = h_k(n)$, and that f_m is then sampled given k_m according to distribution $P(f \mid k) = w_k(f)$, resulting in the joint distribution $P(n,f,k) = P(k,n)P(f \mid k) = \widehat{v}_k(n,f)$. Then $P(n,f)$ is the marginal distribution resulting from the joint distribution $P(n,f,k)$: $P(n,f) = \sum_k P(n,f,k) = \widehat{v}(n,f)$. Finally, note that another convenient formulation of PLCA is to simply state that $\boldsymbol{v}(n) \sim \mathcal{M}(\boldsymbol{v}(n) \mid \|\boldsymbol{v}(n)\|_1, \boldsymbol{W}\boldsymbol{h}(n))$ where $\boldsymbol{v}(n)$ and $\boldsymbol{h}(n)$ are the n-th columns of matrices $\boldsymbol{V}$ and $\boldsymbol{H}$, respectively, $\mathcal{M}$ denotes the multinomial distribution, and $\boldsymbol{h}(n)$ and the columns of matrix $\boldsymbol{W}$ are vectors that sum to 1.

In Section 2.3, it will be shown that this probabilistic model is closely related to KL-NMF. Indeed, the update rules obtained by applying the EM algorithm are formally equivalent to KL-NMF multiplicative update rules (cf. Section 2.3.3).

2.2.3 Poisson NMF model

The Poisson NMF model [Virtanen et al., 2008] is another count model, that assumes that the observed nonnegative matrix $\boldsymbol{V}$ is generated as the sum of K independent, nonnegative latent components $\boldsymbol{V}_k$. The entries $v_k(n,f)$ of matrix $\boldsymbol{V}_k$ are assumed independent and *Poisson*-distributed: $v_k(n,f) \sim \mathcal{P}(v_k(n,f) \mid \widehat{v}_k(n,f))$. The Poisson distribution is defined for any positive integer $\widehat{v}$ as $\mathcal{P}(\widehat{v} \mid \lambda) = \frac{e^{-\lambda}\lambda^{\widehat{v}}}{\widehat{v}!}$, where λ is the intensity parameter and $\widehat{v}!$ is the factorial of $\widehat{v}$. A nice feature of the Poisson distribution is that the sum of K independent Poisson random variables with intensity parameters λ_k is a Poisson random variable with intensity parameter $\lambda = \sum_k \lambda_k$. Consequently, $v(n,f) \sim \mathcal{P}(v(n,f) \mid \widehat{v}(n,f))$. The NMF model $\widehat{\boldsymbol{V}} = \boldsymbol{W}\boldsymbol{H}$ can thus be computed by maximizing $P(\boldsymbol{V} \mid \boldsymbol{W}, \boldsymbol{H}) = \prod_{nf}\mathcal{P}(v(n,f) \mid \widehat{v}(n,f))$. It can be noticed that $\log P(\boldsymbol{V} \mid \boldsymbol{W}, \boldsymbol{H}) = -C^1(\boldsymbol{V} \mid \widehat{\boldsymbol{V}})$, as defined in (2.2) with $\beta = 1$. Therefore ML estimation of $(\boldsymbol{W}, \boldsymbol{H})$ is equivalent to KL-NMF.

2.2.4 Gaussian composite model

The Gaussian *composite model*introduced by Févotte et al. [2009] exploits a feature of the Gaussian distribution that is similar to that of the Poisson distribution: a sum of K independent Gaussian random variables of means μ_k and variances σ_k^2 is a Gaussian random variable of mean $\mu = \sum_k \mu_k$ and variance $\sigma^2 = \sum_k \sigma_k^2$. The main difference with the Poisson NMF model is that, instead of modeling the nonnegative time-frequency representation $\boldsymbol{V}$, IS-NMF aims to model the complex STFT $\boldsymbol{X}$, such that $\boldsymbol{V} = |\boldsymbol{X}|^2$. The observed complex matrix $\boldsymbol{X}$ is thus generated as the sum of K independent complex latent components $\boldsymbol{X}_k$ (cf. Fig 2.2). The entries of matrix $\boldsymbol{X}_k$ are assumed independent and complex Gaussian distributed: $x_k(n,f) \sim \mathcal{N}_c(x_k(n,f) \mid 0, \widehat{v}_k(n,f))$. Here the complex Gaussian distribution is defined as $\mathcal{N}_c(x \mid \mu, \sigma^2) = \frac{1}{\pi\sigma^2}\exp(-\frac{|x-\mu|^2}{\sigma^2})$, where μ and σ^2 are the mean and variance parameters. Consequently, $x(n,f) = \sum_k x_k(n,f) \sim \mathcal{N}_c(x(n,f) \mid 0, \widehat{v}(n,f))$. The NMF model $\widehat{\boldsymbol{V}} = \boldsymbol{W}\boldsymbol{H}$ can thus be computed by maximizing $p(\boldsymbol{X} \mid \boldsymbol{W}, \boldsymbol{H}) = \prod_{nf}\mathcal{N}_c(x(n,f) \mid 0, \widehat{v}(n,f))$. It can be noticed that $\log p(\boldsymbol{X} \mid \boldsymbol{W}, \boldsymbol{H}) = -C^0(\boldsymbol{V} \mid \widehat{\boldsymbol{V}})$, as defined in (2.2) with $\beta = 0$. Therefore ML estimation of $(\boldsymbol{W}, \boldsymbol{H})$ is equivalent to IS-NMF.

In a source separation application, the main practical advantage of this Gaussian composite model is that the *Minimum Mean Square Error* (MMSE) estimates of the sources are obtained by time-frequency masking, in a way that is closely related to Wiener filtering. Indeed, suppose now that the observed signal $x(t)$ is the sum of J unknown source signals $s_j(t)$, so that $x(n,f) = \sum_j s_j(n,f)$, and that each source follows an IS-NMF model:

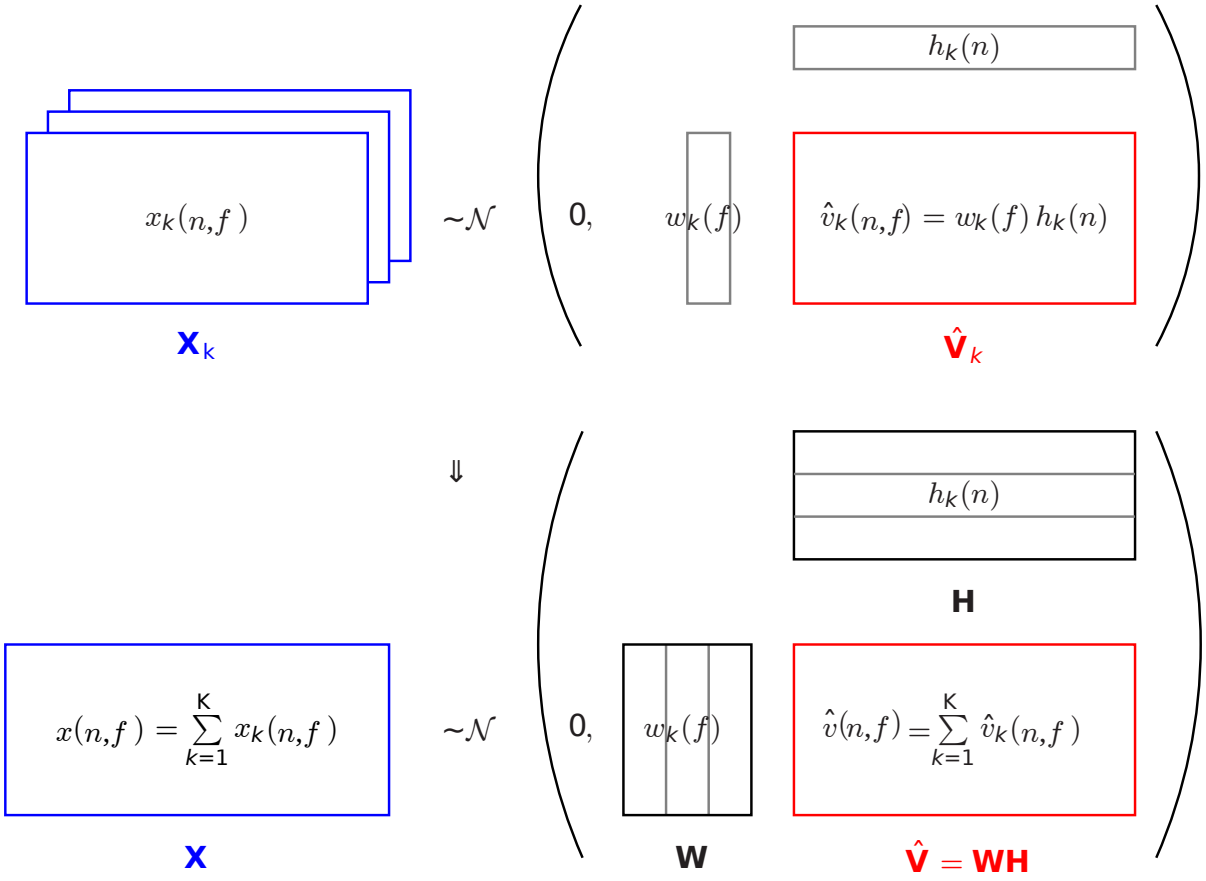


Figure 2.2: Gaussian composite model (IS-NMF) by Févotte et al. [2009]

$s_j(n,f) \sim \mathcal{N}_c(s_j(n,f) \mid 0, \widehat{v}_j(n,f))$ where $\widehat{v}_j(n,f) = \sigma^2_{s_j}(n,f)$ denotes the entries of matrix $\widehat{\boldsymbol{V}}_j = \boldsymbol{W}_j\boldsymbol{H}_j$. Then the minimum of the mean square error (MSE) criterion $\sum_{nf} \mathbb{E}\{|s_j(n,f) - \widehat{s}_j(n,f)|^2 \mid x(n,f)\}$ is reached when

$$\forall n,f,\ \widehat{s}_j(n,f) = \mathbb{E}\{s_j(n,f) \mid x(n,f)\} = m_j(n,f)x(n,f), \tag{2.3}$$

where the time-frequency mask $m_j(n,f)$ is defined as

$$m_j(n,f) = \frac{\widehat{v}_j(n,f)}{\sum_{j'} \widehat{v}_{j'}(n,f)}. \tag{2.4}$$

2.2.5 α-stable NMF models

Despite its nice features, the Gaussian model introduced in the previous section presents two drawbacks. Firstly, it amounts to assuming the additivity of the source power spectrograms, whereas several experimental studies have shown that the additivity of magnitude spectrograms is a better fit (see Liutkus and Badeau [2015] and references therein). Secondly, the IS divergence is not convex, which leads to increased optimization issues due to the existence of local minima. In order to circumvent these problems, a generalization of this model was introduced by Liutkus and Badeau [2015], based on isotropic complex α-*stable* distributions denoted $S\alpha S_c$, which stands for complex symmetric α-stable. This is a family of heavy-tailed probability distributions defined for any $\alpha \in]0,2]$, which do not have a closed-form expression, except in the particular cases $\alpha = 2$, which corresponds to the complex

Gaussian distribution, and $\alpha = 1$, which corresponds to the isotropic complex *Cauchy* distribution. In the general case, the distribution is defined by its characteristic function: $x \sim S\alpha S_c(x \mid \sigma) \Leftrightarrow \phi_z(\theta) = \mathbb{E}\{e^{\jmath\Re(\bar{\theta}x)}\} = e^{-\sigma^\alpha|\theta|^\alpha}$ for any complex-valued θ, where $\sigma > 0$ is the scale parameter (which corresponds to the standard deviation in the Gaussian case). These probability distributions enjoy the same nice feature shared by the Poisson and Gaussian distributions: a sum of K independent isotropic complex α-stable random variables of scale parameters σ_k is an isotropic complex α-stable random variable of scale parameter $\sigma^\alpha = \sum_k \sigma_k^\alpha$.

In this context, the observed STFT matrix $\boldsymbol{X}$ is again modeled as the sum of K independent latent components $\boldsymbol{X}_k$. The entries of matrix $\boldsymbol{X}_k$ are independent and isotropic complex α-stable: $x_k(n,f) \sim S\alpha S_c(x_k(n,f) \mid \sigma_k(n,f))$, where $\widehat{v}_k(n,f) = \sigma_k^\alpha(n,f)$ is called an α-spectrogram. Thus $x(n,f) = \sum_k x_k(n,f) \sim S\alpha S_c(x(n,f) \mid \sigma(n,f))$, with

$$\widehat{v}(n,f) = \sigma^\alpha(n,f) = \sum_k \sigma_k^\alpha(n,f) = \sum_k \widehat{v}_k(n,f). \tag{2.5}$$

When the distribution has a closed-form expression (i.e., $\alpha = 1$ or 2), the NMF model $\widehat{\boldsymbol{V}} = \boldsymbol{W}\boldsymbol{H}$ can still be estimated in the ML sense, otherwise different inference methods are required. In the Cauchy case [Liutkus et al., 2015], it has been experimentally observed that Cauchy NMF is much less sensitive to initialization than IS-NMF and produces meaningful basis spectra for source separation.

In a source separation application, we again suppose that the observed signal $x(t)$ is the sum of J unknown source signals $s_j(t)$, so that $x(n,f) = \sum_j s_j(n,f)$, and that each source follows an isotropic complex α-stable NMF model: $s_j(n,f) \sim S\alpha S_c(s_j(n,f) \mid \widehat{v}_j^{1/\alpha}(n,f))$ where $\widehat{v}_j(n,f)$ denotes the entries of matrix $\widehat{\boldsymbol{V}}_j = \boldsymbol{W}_j\boldsymbol{H}_j$. Then the MSE criterion is no longer defined for all $\alpha \in (0,2]$, but in any case, the posterior mean $\widehat{s}_j(n,f) = \mathbb{E}\{s_j(n,f) \mid x(n,f)\}$ is still well-defined, and admits the same expression as in (2.3) and (2.4).

2.2.6 Choosing a particular NMF model

When choosing a particular NMF model for a given source separation application, several criteria may be considered, including the following ones:

Robustness to initialization: Cauchy NMF has proved to be more robust to initialization than all other probabilistic NMF models. Besides, in the context of pretrained source separation, the Gaussian noise model (related to EUC-NMF) and the PLCA/Poisson NMF models (related to KL-NMF) lead to a convex optimization problem with a unique minimum, which is not the case of the Gaussian composite model (related to IS-NMF).

Source reconstruction: only the α-stable NMF models, including IS-NMF and Cauchy NMF, provide a theoretical ground for using Wiener filtering in order to reconstruct time-domain signals.

Existence of closed-form update rules: ML estimation of the model parameters is tractable for all NMF models except α-stable models with $\alpha \neq 1, 2$.

2.3 Algorithms for NMF

In the literature, various algorithms have been designed for computing an NMF, including the famous multiplicative update rules [Lee and Seung, 2001], the alternated least squares method [Finesso and Spreij, 2004], and the projected gradient method [Lin, 2007]. In Section 2.3.1, we present the multiplicative update rules, that form the most celebrated NMF algorithm, and we summarize their convergence properties. Then in the following sections, we present some algorithms dedicated to the probabilistic frameworks introduced in Section 2.2.

2.3.1 Multiplicative update rules

The basic idea of *multiplicative update* rules is that the nonnegativity constraint can be easily enforced by updating the previous values of the model parameters by multiplication with a nonnegative scale factor. A heuristic way of deriving these updates consists in decomposing the gradient of the cost function $C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}})$, e.g., the β-divergence introduced in (2.2), as the difference of two nonnegative terms: $\nabla_{\boldsymbol{W}} C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) = \nabla_{\boldsymbol{W}}^+ C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) - \nabla_{\boldsymbol{W}}^- C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}})$, where

$\nabla^+_{\boldsymbol{W}} C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) \geq 0$ and $\nabla^-_{\boldsymbol{W}} C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) \geq 0$, meaning that all the entries of these two matrices are nonnegative. Then matrix $\boldsymbol{W}$ can be updated as $\boldsymbol{W} \leftarrow \boldsymbol{W} \circ (\nabla^-_{\boldsymbol{W}} C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) / \nabla^+_{\boldsymbol{W}} C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}))^\eta$, where $\circ$ denotes elementwise matrix product, / denotes elementwise matrix division, the matrix exponentiation must be understood elementwise, and $\eta > 0$ is a stepsize similar to that involved in a gradient descent [Badeau et al., 2010]. The same update can be derived for matrix $\boldsymbol{H}$, and then matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ can be updated in turn, until convergence[1]. Note that the decomposition of the gradient as a difference of two nonnegative terms is not unique, and different choices can be made, leading to different multiplicative update rules. In the case of the β-divergence, the standard multiplicative update rules are expressed as follows [Févotte and Idier, 2011]:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \circ \left(\frac{(\boldsymbol{V} \circ (\boldsymbol{W}\boldsymbol{H})^{\beta-2})\boldsymbol{H}^\top}{(\boldsymbol{W}\boldsymbol{H})^{\beta-1}\boldsymbol{H}^\top} \right)^\eta \tag{2.6}$$

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \circ \left(\frac{\boldsymbol{W}^\top(\boldsymbol{V} \circ (\boldsymbol{W}\boldsymbol{H})^{\beta-2})}{\boldsymbol{W}^\top(\boldsymbol{W}\boldsymbol{H})^{\beta-1}} \right)^\eta \tag{2.7}$$

where matrix division and exponentiation must be understood elementwise. By using the auxiliary function approach, Nakano et al. [2010] proved that the cost function $C^\beta(\boldsymbol{V} \mid \widehat{\boldsymbol{V}})$ is nonincreasing under these updates when the stepsize η is given by $\eta = \frac{1}{2-\beta}$ for $\beta < 1$, $\eta = 1$ for $1 \leq \beta \leq 2$, and $\eta = \frac{1}{\beta-1}$ for $\beta > 2$. In addition, Févotte and Idier [2011] proved that the same cost function is nonincreasing under (2.6)–(2.7) for $\eta = 1$ and for all $\beta \in [0, 2]$ (which includes the popular EUC, KL and IS-NMF). They also proved that these updates correspond to a *Majorization-Minimization* (MM) algorithm when the stepsize η is expressed as a given function of β, which is equal to 1 for all $\beta \in [1, 2]$, and they correspond to a majorization-equalization algorithm for $\eta = 1$ and $\beta = 0$. However, contrary to a widespread belief [Lee and Seung, 2001], the decrease of the cost function is not sufficient to prove the convergence of the algorithm to a local or global minimum. Badeau et al. [2010] analyzed the convergence of multiplicative update rules by means of Lyapunov's stability theory. In particular, it was proved that:

- There is $\eta^{\max} > 0$ such that these rules are exponentially or asymptotically stable for all $\eta \in (0, \eta^{\max})$. Moreover, $\forall \beta$, the upper bound $\eta^{\max}$ is such that $\eta^{\max} \in (0, 2]$, and if $\beta \in [1, 2]$, $\eta^{\max} = 2$.
- These rules are unstable if $\eta \notin [0, 2]$, $\forall \beta$.

In practice, the step size η permits us to control the convergence rate of the algorithm.

Note that, due to the nonuniqueness of NMF, there is a scaling and permutation ambiguity between matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ (cf. Section 2.1). Therefore, when $\boldsymbol{W}$ and $\boldsymbol{H}$ are to be updated in turn, numerical stability can be improved by renormalizing the columns of $\boldsymbol{W}$ (resp. the rows of $\boldsymbol{H}$), and scaling the rows of $\boldsymbol{H}$ (resp. the columns of $\boldsymbol{W}$) accordingly, so as to keep the product $\boldsymbol{W}\boldsymbol{H}$ unchanged.

Finally, a well-known drawback of most NMF algorithms is the sensitivity to initialization, that is due to the multiplicity of local minima of the cost function (cf. Section 2.1). Many initialization strategies were thus proposed in the literature [Cichocki et al., 2009]. In the case of IS-NMF multiplicative update rules, a *tempering* approach was proposed by Bertin et al. [2009]. The basic idea is the following one: since the β-divergence is convex for all $\beta \in [1, 2]$, but not for $\beta = 0$, the number of local minima is expected to increase when β goes from 2 to 0. Therefore a simple solution for improving the robustness to initialization consists in making parameter β vary from 2 to 0 over the iterations of the algorithm. Nevertheless, the best way of improving the robustness to initialization in general is to select a robust NMF criterion, such as that involved in Cauchy NMF (cf. Section 2.2.5).

2.3.2 The EM algorithm and its variants

As mentioned in Section 2.2, one advantage of using a probabilistic framework for NMF is the availability of classical inference techniques, whose convergence properties are well-known. Classical algorithms used in the NMF literature include the EM algorithm [Shashanka et al., 2008], the space-alternating generalized EM algorithm [Févotte et al., 2009], variational Bayesian (VB) inference [Badeau and Drémeau, 2013], and *Markov chain Monte Carlo* [Simsekli and Cemgil, 2012].

[1] This iterative algorithm can stop, e.g., when the decrease of the β-divergence, or when the distance between the successive iterates of matrices $\boldsymbol{W}$ and $\boldsymbol{H}$, goes below a given threshold.

Below, we introduce the basic principles of the space-alternating generalized EM algorithm [Fessler and Hero, 1994], which includes the regular EM algorithm as a particular case. We then apply the EM algorithm to the PLCA framework described in Section 2.2.2, and the space-alternating generalized EM algorithm to the Gaussian composite model described in Section 2.2.4.

Consider a random observed dataset $\mathcal{X}$, whose probability distribution is parameterized by a parameter set $\boldsymbol{\theta}$, that is partitioned as $\boldsymbol{\theta} = \{\boldsymbol{\theta}_k\}_k$. The space-alternating generalized EM algorithm aims to estimate parameters $\boldsymbol{\theta}_k$ iteratively, while guaranteeing that the likelihood $p(\mathcal{X} \mid \boldsymbol{\theta})$ is nondecreasing over the iterations. It requires choosing for each subset $\boldsymbol{\theta}_k$ a *hidden data* space which is complete for this particular subset, i.e., a latent dataset $\mathcal{X}_k$ such that $p(\mathcal{X}, \mathcal{X}_k \mid \boldsymbol{\theta}) = p(\mathcal{X} \mid \mathcal{X}_k, \{\boldsymbol{\theta}_{k'}\}_{k' \neq k}) p(\mathcal{X}_k \mid \boldsymbol{\theta})$. The algorithm iterates over both the iteration index and over k. For each iteration and each k, it is composed of an expectation step (*E-step*) and a maximization step (*M-step*):

- E-step: evaluate $Q_k(\boldsymbol{\theta}_k) = \mathbb{E}\{\log p(\mathcal{X}_k \mid \boldsymbol{\theta}_k, \{\boldsymbol{\theta}_{k'}\}_{k' \neq k}) \mid \mathcal{X}, \boldsymbol{\theta}\}$;
- M-step: compute $\boldsymbol{\theta}_k = \operatorname{argmax}_{\boldsymbol{\theta}_k} Q_k(\boldsymbol{\theta}_k)$.

The regular EM algorithm corresponds to the particular case $K = 1$, where $\mathcal{X}$ is a deterministic function of the complete data space.

2.3.3 Application of the EM algorithm to PLCA

Shashanka et al. [2008] applied the EM algorithm to the PLCA model described in Section 2.2.2. The observed dataset is $\mathcal{X} = \{n_m, f_m\}_m$, the parameter set is $\boldsymbol{\theta} = \{\boldsymbol{W}, \boldsymbol{H}\}$, and the complete data space is $\{n_m, f_m, k_m\}_m$. Then:

- The E-step consists in computing $P(k \mid n, f) = \frac{P(n,f,k)}{\sum_{k'} P(n,f,k')} = \frac{\widehat{v}_k(n,f)}{\widehat{v}(n,f)}$, that appears in the expression $Q(\boldsymbol{\theta}) = \sum_{nf} v(n,f) \sum_k P(k \mid n, f) \log\left(w_k(f) h_k(n)\right)$.
- The M-step consists in maximizing $Q(\boldsymbol{\theta})$ with respect to $w_k(f)$ and $h_k(n)$, subject to $\forall k$, $\sum_f w_k(f) = 1$ and $\sum_{kn} h_k(n) = 1$. Given that $\sum_{nf} v(n,f) = 1$, we get:

$$h_k(n) \leftarrow \frac{\sum_f v(n,f) P(k \mid n, f)}{\sum_{k'n'f} v(n',f) P(k' \mid n', f)} = h_k(n) \sum_f w_k(f) \frac{v(n,f)}{\widehat{v}(n,f)}, \tag{2.8}$$

$$w_k(f) \leftarrow \frac{\sum_n v(n,f) P(k \mid n, f)}{\sum_{nf'} v(n,f') P(k \mid n, f')} = \frac{\tilde{w}_k(f)}{\sum_{f'} \tilde{w}_k(f')}, \tag{2.9}$$

  where $\tilde{w}_k(f) = w_k(f) \sum_n h_k(n) \frac{v(n,f)}{\widehat{v}(n,f)}$.

It is easy to check that this algorithm is identical to the multiplicative update rules for KL divergence, as described in (2.6)–(2.7) with $\eta = \beta = 1$, up to a scaling factor in $\boldsymbol{H}$ due to the normalization of matrix $\boldsymbol{V}$ [Shashanka et al., 2008].

2.3.4 Application of the space-alternating generalized EM algorithm to the Gaussian composite model

Févotte et al. [2009] applied the *Space Alternating Generalized EM* (SAGE) algorithm to the Gaussian composite model described in Section 2.2.4. The observed dataset is $\mathcal{X} = \boldsymbol{X}$, the k-th parameter set is $\boldsymbol{\theta}_k = \{\boldsymbol{w}_k, \boldsymbol{h}_k\}$, and the k-th complete latent dataset is $\mathcal{X}_k = \boldsymbol{X}_k$. Then:

- The E-step consists in computing $\boldsymbol{V}_k = \frac{\widehat{\boldsymbol{V}}_k^2}{\widehat{\boldsymbol{V}}^2} \circ \boldsymbol{V} + \frac{\widehat{\boldsymbol{V}}_k \circ (\widehat{\boldsymbol{V}} - \widehat{\boldsymbol{V}}_k)}{\widehat{\boldsymbol{V}}}$, that appears in the expression $Q_k(\boldsymbol{\theta}_k) = -C^0(\boldsymbol{V}_k \mid \widehat{\boldsymbol{V}}_k)$, where criterion C^0 was defined in (2.2) with $\beta = 0$.
- The M-step computes $h_k(n) \leftarrow \frac{1}{F} \sum_f \frac{v_k(n,f)}{w_k(f)}$ and $w_k(f) \leftarrow \frac{1}{N} \sum_n \frac{v_k(n,f)}{h_k(n)}$.

Note that it has been experimentally observed by Févotte et al. [2009] that this space-alternating generalized EM algorithm converges more slowly than the IS-NMF multiplicative update rules described in (2.6)–(2.7) for $\eta = 1$ and $\beta = 0$.

3 Advanced NMF models

The basic NMF model presented in Section 1 has proved successful for addressing a variety of audio source separation problems. Nevertheless, the source separation performance can still be improved by exploiting prior knowledge that we may have about the source signals. For instance, we know that musical notes and voiced sounds have a harmonic spectrum (or, more generally, an inharmonic or a sparse spectrum), and that both their spectral envelope and their temporal power profile have smooth variations. On the opposite, percussive sounds rather have a smooth spectrum, and a sparse temporal power profile. It may thus be desirable to impose properties such as *harmonicity*, *smoothness*, and *sparsity* on either the spectral matrix $\boldsymbol{W}$ or the activation matrix $\boldsymbol{H}$ in the NMF $\widehat{\boldsymbol{V}} = \boldsymbol{W}\boldsymbol{H}$. For that purpose, it is possible to apply either hard constraints, e.g., by parameterizing matrix $\boldsymbol{W}$ or $\boldsymbol{H}$, or soft constraints, e.g., by adding a regularization term to the criterion (2.2), or by introducing the prior distributions of $\boldsymbol{W}$ or $\boldsymbol{H}$ in the probabilistic frameworks introduced in Section 2.2 (Bayesian approach). Examples of such regularizations are described in Section 3.1. Note that another possible way of exploiting prior information is to use a predefined dictionary $\boldsymbol{W}$ trained on a training dataset.

In other respects, audio signals are known to be nonstationary, therefore it is useful to consider that some characteristics such as the fundamental frequency or the spectral envelope may vary over time. Such nonstationary models will be presented in Section 3.2.

3.1 Regularizations

In this section, we present a few examples of NMF regularizations, including sparsity (Section 3.1.1), group-sparsity (Section 3.1.2), harmonicity and spectral smoothness (Section 3.1.3), and inharmonicity (Section 3.1.4).

3.1.1 Sparsity

Since NMF is well suited to the problem of separating audio signals formed of a few repeated audio events, it is often desirable to enforce the sparsity of matrix $\boldsymbol{H}$.

The most straightforward way of doing so is to add to the NMF criterion a sparsity-promoting regularization term. Ideally, sparsity is measured by the ℓ_0 norm, which counts the number of nonzero entries in a vector. However, optimizing a criterion involving the ℓ_0 norm raises intractable combinatorial issues. In the optimization literature, the ℓ_1 norm is often preferred, because it is the tightest convex relaxation of the ℓ_0 norm. Therefore the criterion $C^\beta(\boldsymbol{V} \mid \widehat{\boldsymbol{V}})$ in (2.2) may be replaced with

$$C(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) = C^\beta(\boldsymbol{V} \mid \widehat{\boldsymbol{V}}) + \lambda \sum_k \|\boldsymbol{h}_k\|_1, \tag{2.10}$$

where $\lambda > 0$ is a tradeoff parameter to be tuned manually, as suggested, e.g., by Hurmalainen et al. [2015].

However, if the NMF is embedded in a probabilistic framework such as those introduced in Section 2.2, sparsity is rather enforced by introducing an appropriate prior distribution of matrix $\boldsymbol{H}$. In this case, $\boldsymbol{H}$ is estimated by maximizing its posterior probability given $\boldsymbol{V}$, or equivalently the *Maximum a Posteriori* (MAP) criterion $\log p(\boldsymbol{V} \mid \boldsymbol{W}, \boldsymbol{H}) + \log p(\boldsymbol{H})$, instead of the log-likelihood $\log p(\boldsymbol{V} \mid \boldsymbol{W}, \boldsymbol{H})$. For instance, Kameoka et al. [2009] consider a generative model similar to the Gaussian noise model presented in Section 2.2.1, where the sparsity of matrix $\boldsymbol{H}$ is enforced by means of a generalized Gaussian prior:

$$p(\boldsymbol{H}) = \prod_{kn} \frac{1}{2\Gamma\left(1 + \frac{1}{p}\right)\sigma} e^{-\frac{|h_k(n)|^p}{\sigma^p}}, \tag{2.11}$$

where $\Gamma(\cdot)$ denotes the gamma function, parameter p promotes sparsity if $0 < p < 2$, and the case $p = 2$ corresponds to the standard Gaussian distribution.

In the PLCA framework described in Section 2.2.2, the entries of $\boldsymbol{H}$ are the discrete probabilities $P(k, n)$. By noticing that the entropy $\mathbb{H}\{\boldsymbol{H}\}$ of this discrete probability distribution is related to the sparsity of matrix $\boldsymbol{H}$ (the lower $\mathbb{H}\{\boldsymbol{H}\}$, the sparser $\boldsymbol{H}$), a suitable sparsity-promoting prior is the so-called *entropic prior* [Shashanka et al., 2008], defined as $p(\boldsymbol{H}) \propto e^{-\beta\mathbb{H}\{\boldsymbol{H}\}}$, where $\beta > 0$.

3.1.2 Group sparsity

Now suppose that the observed signal $x(t)$ is the sum of J unknown source signals $s_j(t)$ for $j \in \{1, \ldots, J\}$, whose spectrograms $\boldsymbol{V}_j$ are approximated as $\widehat{\boldsymbol{V}}_j = \boldsymbol{W}_j \boldsymbol{H}_j$, as in Section 2.2.4. Then the spectrogram $\boldsymbol{V}$ of $x(t)$ is approximated with the NMF $\sum_j \widehat{\boldsymbol{V}}_j = \boldsymbol{W}\boldsymbol{H}$, where $\boldsymbol{W} = [\boldsymbol{W}_1, \ldots, \boldsymbol{W}_J]$ and $\boldsymbol{H} = [\boldsymbol{H}_1^\top, \ldots, \boldsymbol{H}_J^\top]^\top$. In this context, it is natural to expect that if a given source j is inactive at time n, then all the entries in the n-th column of $\boldsymbol{H}_j$ are zero. Such a property can be enforced by using *group sparsity*. A well-known group sparsity regularization term is the mixed ℓ_2-ℓ_1 norm: $\|\boldsymbol{H}\|_{2,1} = \sum_{jn} \|\boldsymbol{h}_j(n)\|_2$ where $\boldsymbol{h}_j(n)$ is the n-th column of matrix $\boldsymbol{H}_j$, as suggested, e.g., by Hurmalainen et al. [2015]. Indeed, the minimization of the criterion (2.10) involving this regularization term tends to enforce sparsity over both n and j, while ensuring that the whole vector $\boldsymbol{h}_j(n)$ gets close to zero for most values of n and j.

Lefevre et al. [2011] proposed a group sparsity prior for the IS-NMF probabilistic framework described in Section 2.2.4. The idea is to consider a prior distribution of matrix $\boldsymbol{H}$ such that all vectors $\boldsymbol{h}_j(n)$ are independent: $p(\boldsymbol{H}) = \prod_{jn} p(\boldsymbol{h}_j(n))$. Each $p(\boldsymbol{h}_j(n))$ is chosen so as to promote near-zero vectors. Then, as in Section 3.1.1, the NMF parameters are estimated in the MAP sense: $(\boldsymbol{W}, \boldsymbol{H}) = \operatorname{argmax}_{\boldsymbol{W},\boldsymbol{H}} \log p(\boldsymbol{X} \mid \boldsymbol{W}, \boldsymbol{H}) + \sum_{jn} \log p(\boldsymbol{h}_j(n))$, where $p(\boldsymbol{X} \mid \boldsymbol{W}, \boldsymbol{H})$ was defined in Section 2.2.4.

3.1.3 Harmonicity and spectral smoothness

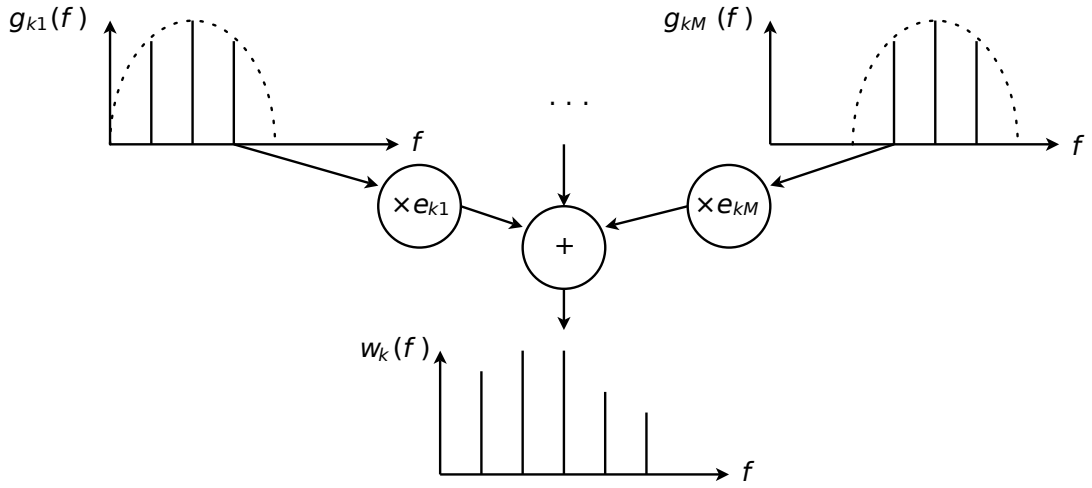


Figure 2.3: Harmonic NMF model by Vincent et al. [2010] and Bertin et al. [2010]

Contrary to sparsity, harmonicity in matrix $\boldsymbol{W}$ is generally enforced as a hard constraint, by using parametric models, whose parameter set includes the fundamental frequency. For instance, Vincent et al. [2010] and Bertin et al. [2010] parameterized the spectrum vector $\boldsymbol{w}_k$ as a nonnegative linear combination of M narrowband, harmonic spectral patterns (cf. Fig. 2.3): $w_k(f) = \sum_m e_{km} g_{km}(f)$, where all spectral patterns $\{g_{km}(f)\}_m$ share the same fundamental frequency $\nu_k^0 > 0$, have smooth spectral envelopes and different spectral centroids, so as to form a filterbank-like decomposition of the whole spectrum, and $\{e_{km}\}_m$ are the nonnegative coefficients of this decomposition. In this way, it is guaranteed that $w_k(f)$ is a harmonic spectrum of fundamental frequency ν_k^0, with a smooth spectral envelope. If the signal of interest is a music signal, then the order K and the fundamental frequencies ν_k^0 can typically be preset according to the semitone scale; otherwise they have to be estimated along with the other parameters. Two methods were proposed for estimating the coefficients e_{km} and the activations in matrix $\boldsymbol{H}$ from the observed spectrogram: a space-alternating generalized EM algorithm based on a Gaussian model [Bertin et al., 2010] (cf. Section 2.3.2) and multiplicative update rules (with a faster convergence speed) based either on the IS divergence [Bertin et al., 2010], or more generally on the β-divergence [Vincent et al., 2010].

Hennequin et al. [2010] proposed a similar parameterization of the spectrum vector $\boldsymbol{w}_k$, considering that harmonic spectra are formed of a number M of distinct partials:

$$w_k(f) = \sum_m a_k^m \, g_{km}(f), \tag{2.12}$$

where $a_k^m \geq 0$, $g_{km}(f) = g\left(\nu_f - \nu_k^m\right)$, $\nu_f = \frac{f}{F} f_s$ and $\nu_k^m = m\,\nu_k^0$, and $g(\cdot)$ is the spectrum of the analysis window used for computing the spectrogram. Multiplicative update rules based on the β-divergence were proposed for estimating this model. Since this parametric model does not explicitly enforce the smoothness of the spectral envelope, a regularization term promoting this smoothness was added to the β-divergence, resulting in a better decomposition of music spectrograms [Hennequin et al., 2010].

3.1.4 Inharmonicity

When modeling some string musical instruments such as the piano or the guitar, the harmonicity assumption has to be relaxed. Indeed, because of the bending stiffness, the partial frequencies no longer follow an exact harmonic progression, but rather a so-called *inharmonic* progression:

$$\nu_k^m = m\nu_k^0 \sqrt{1 + Bm^2}, \tag{2.13}$$

where m is the partial index, $B > 0$ is the inharmonicity coefficient, and $\nu_k^0 > 0$ is the fundamental frequency of vibration of an ideal flexible string [Rigaud et al., 2013]. Then the spectrum vector $\boldsymbol{w}_k$ can be parameterized as in (2.12), and all parameters, including the inharmonicity coefficient B, can be estimated by minimizing the β-divergence criterion by means of multiplicative update rules. However, it was observed that the resulting algorithm is very sensitive to initialization (cf. Section 2.3.1). In order to improve the robustness to initialization, the exact parameterization of frequencies ν_k^m in (2.13) was relaxed by considering these frequencies as free parameters, and by adding the following regularization term to the β-divergence criterion: $\sum_{km} |\nu_k^m - m\nu_k^0 \sqrt{1 + Bm^2}|^2$.

3.2 Nonstationarity

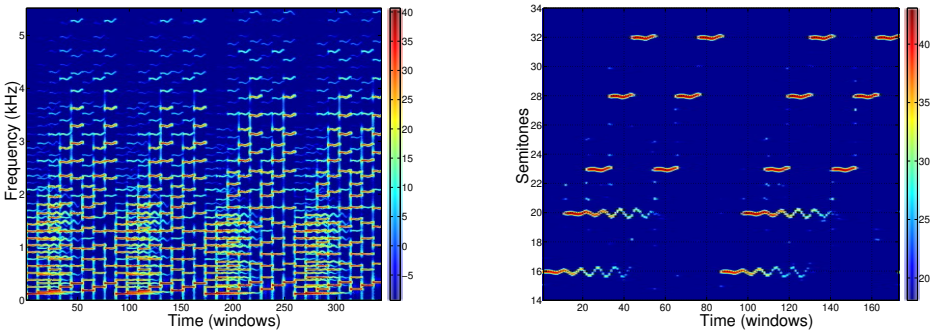
In the previous Section 3.1, several methods have been presented for enforcing the harmonicity and the spectral smoothness of vectors $\boldsymbol{w}_k$ in matrix $\boldsymbol{W}$, by means of either hard or soft constraints. All these methods assumed that the spectra of the audio events forming the observed spectrogram are stationary. However, many real audio signals are known to be nonstationary: the fundamental frequency, as well as the spectral envelope, may vary over time. In this section, we present some models that aim to represent such nonstationary signals, by allowing the fundamental frequency and spectral envelope parameters to vary over time.

3.2.1 Time-varying fundamental frequencies

In Section 3.1.3, a harmonic parameterization of vector $\boldsymbol{w}_k$ was described in (2.12). Hennequin et al. [2010] proposed a straightforward generalization of this model by making the spectral coefficient w_k also depend of time n: $w_k(n, f) = \sum_m a_k^m \, g\left(\nu_f - \nu_k^m(n)\right)$, resulting in a a spectrogram model that is a generalization of NMF: $\widehat{v}(n, f) = \sum_k \widehat{v}_k(n, f)$ with $\widehat{v}_k(n, f) = w_k(n, f)\, h_k(n)$.

Multiplicative update rules based on the β-divergence were proposed for estimating this extended model, along with several regularization terms designed to better fit music spectrograms [Hennequin et al., 2010].

We used this model to decompose the spectrogram of an excerpt (first 4 bars) of the first prelude of Johann Sebastian Bach played by a synthesizer (figure 2.4(a)). A slight vibrato has been added in the played notes to emphasize the variable fundamental frequency estimation. The decomposition uses 72 spectral shapes distributed every semitone. The figure 2.4(b) represents the activations and the fundamental frequencies ν_{kn}^0 obtained. The notes of the prelude appear very clearly, with the vibrato effect.



(a) Original spectrogram

(b) Representation of the activations including the fundamental frequencies (first two measurements of the extract). The color scale is in dB.

Figure 2.4: Decomposition of an excerpt from the first Prelude by Johann Sebastian Bach (figure extracted from Hennequin et al. [2010])

3.2.2 Time-varying spectral envelopes

Beyond the fundamental frequency, the spectral envelope of freely vibrating harmonic tones (such as those produced by a piano or a guitar) is not constant over time: generally, the upper partials decrease faster than the lower ones. Besides, some sounds such as those produced by a didgeridoo are characterized by a strong resonance in the spectrum that varies over time. Similarly, every time fingerings change on a wind instrument, the shape of the resonating body changes and the resonance pattern is different.

In order to properly model such sounds involving time-varying spectra, Hennequin et al. [2011] proposed to make the activations in vector $\boldsymbol{h}_k$ not only depend on time, but also on frequency, in order to account for the temporal variations of the spectral envelope of vector $\boldsymbol{w}_k$. More precisely, the activation coefficient $h_k(n,f)$ is parameterized according to an *Autoregressive Moving Average* (ARMA) model:

$$h_k(n,f) = \sigma_k^2(n) \left| \frac{1 + \sum_{n'} \beta_k(n,n') e^{-2j\pi n' f/F}}{1 - \sum_{n'} \alpha_k(n,n') e^{-2j\pi n' f/F}} \right|^2 \tag{2.14}$$

where $\sigma_k^2(n)$ is the variance parameter at time n, $\alpha_k(n,n')$ denotes the *Autoregressive* (AR) coefficients, and $\beta_k(n,n')$ the *Moving Average* (MA) coefficients, at time n. Then the NMF model is generalized in the following way: $\widehat{v}(n,f) = \sum_k \widehat{v}_k(n,f)$ with $\widehat{v}_k(n,f) = w_k(f)\, h_k(n,f)$. This model is also estimated by minimizing a β-divergence criterion. All parameters, including the ARMA coefficients, are computed by means of multiplicative update rules, without any training. Note that even though the ARMA model of $h_k(n,f)$ in (2.14) is nonnegative, the model coefficients $\alpha_k(n,n')$ and $\beta_k(n,n')$ are not necessarily nonnegative, which means that the multiplicative update rules introduced in Section 2.3.1 were generalized so as to handle these coefficients appropriately [Hennequin et al., 2011].

This algorithm allowed to efficiently represent non-stationary sounds with strong spectral variations, such as the Jew's harp sounds. The Jew's harp is an instrument made of a vibrating metal rod. This rod is placed in the mouth of the instrumentalist who modulates the sound with his mouth. It is thus a harmonic sound (with a fixed fundamental frequency) with a strong resonance varying with time (see the spectrogram in figure 2.5(a)). The decomposition obtained with our algorithm (using a single component) is shown in figures 2.5(b) and 2.5(c) : it shows well the harmonic shape of the spectrum on the one hand and the temporal variations of the resonance on the other hand.

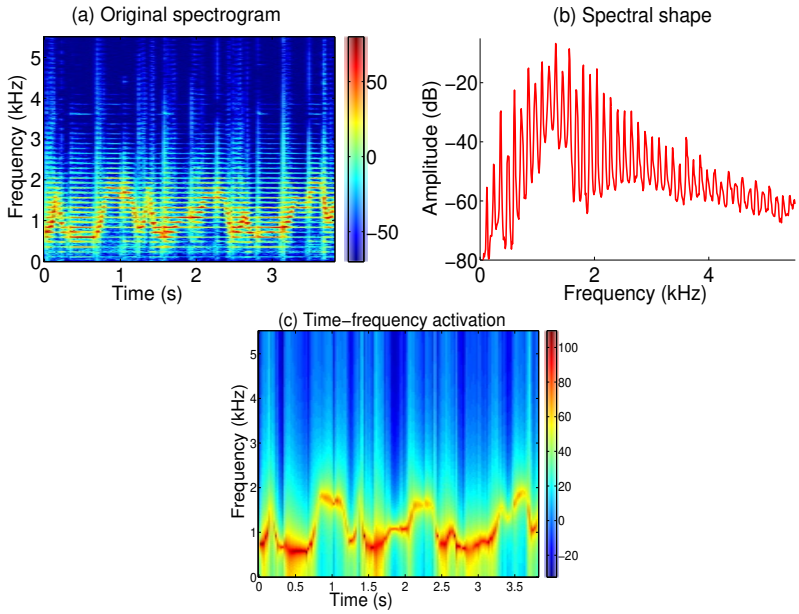


Figure 2.5: Jew's harp sound decomposed with a time-frequency activation parameterized by an ARMA filter of order (1,1) (figure extracted from Hennequin [2010])

3.2.3 Both types of variations

In order to account for the temporal variations of the fundamental frequency and the spectral envelope jointly, Fuentes et al. [2013] proposed a model called harmonic adaptive latent component analysis. This models falls within the scope of the PLCA framework described in Section 2.2.2: matrix $\widehat{\boldsymbol{V}}$ is viewed as a discrete probability distribution $P(n,f) = \widehat{v}(n,f)$, and the spectrogram $\boldsymbol{V}$ is modeled as a histogram: $v(n,f) = \frac{1}{M}\sum_m \delta_{(n_m,f_m)}(n,f)$, where $\{(n_m,f_m)\}_m$ are i.i.d. random vectors distributed according to $P(n,f)$.

In practice, the time-frequency transform used to compute $\boldsymbol{V}$ is a constant-Q transform. Because this transform involves a log-frequency scale, pitch shifting can be approximated as a translation of the spectrum along this log-frequency axis. Therefore all the notes produced by source j at time n are approximately characterized by a unique template spectrum modeled by a probability distribution $P(\mu \mid j,n)$ (where μ is a frequency parameter), which does not depend on the fundamental frequency. However this distribution depends on time n in order to account for possible temporal variations of the spectral envelope. Besides, the variation of the pitch f_0 of source j over time n is modeled by a probability distribution $P(f_0 \mid j,n)$. Hence the resulting distribution of the shifted frequency $f = \mu + f_0$ is $P(f \mid j,n) = \sum_{f_0} P(f - f_0 \mid j,n)P(f_0 \mid j,n)$. Finally, the presence of source j at time n is characterized by a distribution $P(j,n)$. Therefore the resulting spectrogram corresponds to the probability distribution $P(n,f) = \sum_{jf_0} P(f - f_0 \mid j,n)P(f_0 \mid j,n)P(j,n)$.

In order to enforce both the harmonicity and the smoothness of the spectral envelope, the template spectrum $P(\mu \mid j,n)$ is modeled in the same way as in the first paragraph of Section 3.1.3, as a nonnegative linear combination of K narrowband, harmonic spectral patterns $P(\mu \mid k)$: $P(\mu \mid j,n) = \sum_k P(\mu \mid k)P(k \mid j,n)$, where $P(k \mid j,n)$ is the nonnegative weight of pattern k at time n for source j. Finally, the resulting harmonic adaptive latent component analysis model is expressed as

$$P(n,f) = \sum_{f_0 k j} P(f - f_0 \mid k)P(k \mid j,n)P(f_0 \mid j,n)P(j,n). \tag{2.15}$$

4 Summary

In this chapter, we have shown that NMF is a very powerful model for representing speech and music data. We have presented the mathematical foundations, and described several probabilistic frameworks and various algorithms for computing an NMF. We have also presented some advanced NMF models that are able to more accurately represent audio signals, by enforcing properties such as sparsity, harmonicity and spectral smoothness, and by taking the nonstationarity of the data into account. We have shown that coupled factorizations make it possible to exploit some extra information we may have about the observed signal, such as the musical score. Finally, we have presented several methods that perform dictionary learning for NMF.

The benefits of NMF in comparison with other separation approaches are the capability of performing unsupervised source separation, learning source models from a relatively small amount of material (especially in comparison with *Deep Neural Networks* (DNN)), and easily implementing and adapting the source models and the algorithms. The main downside is the complexity of iterative NMF algorithms. Note that beyond source separation, NMF models have also proved successful in a broad range of audio applications, including automatic music transcription [Smaragdis and Brown, 2003], multipitch estimation [Vincent et al., 2010, Bertin et al., 2010, Fuentes et al., 2013, Benetos et al., 2014], and audio inpainting [Smaragdis et al., 2011].

Bibliography

H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, *Proc. of the 2nd International Symposium on Information Theory*, pages 267–281, Budapest, Hongrie, 1973. Akademia Kiado.

R. Badeau and A. Drémeau. Variational Bayesian EM algorithm for modeling mixtures of non-stationary signals in the time-frequency domain (HR-NMF). In *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, pages 6171–6175, May 2013.

R. Badeau, N. Bertin, and E. Vincent. Stability analysis of multiplicative update algorithms and application to non-negative matrix factorization. *IEEE Transactions on Neural Networks*, 21(12):1869–1881, Dec. 2010.

A. J. Barabell. Improving the resolution performance of eigenstructure-based direction-finding algorithms. In *Proc. of ICASSP'83*, pages 336–339, Boston, MA, USA, 1983. IEEE.

E. Benetos, G. Richard, and R. Badeau. Template adaptation for improving automatic music transcription. In *Proceedings of International Society for Music Information Retrieval Conference*, pages 175–180, Oct. 2014.

N. Bertin. *Les factorisations en matrices non-négatives. Approches contraintes et probabilistes, application à la transcription automatique de musique polyphonique.* PhD thesis, École Nationale Supérieure des Télécommunications, Paris, France, Oct. 2009.

N. Bertin, C. Févotte, and R. Badeau. A tempering approach for Itakura-Saito non-negative matrix factorization. With application to music transcription. In *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, pages 1545–1548, Apr. 2009.

N. Bertin, R. Badeau, and E. Vincent. Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 18(3):538–549, Mar. 2010.

G. Bienvenu and L. Kopp. Optimality of high-resolution array processing using the eigensystem method. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(5):1235–1245, Oct. 1983.

W. B. Bishop and P. M. Djuric. Model order selection of damped sinusoids in noise by predictive densities. *IEEE Transactions on Signal Processing*, 44(3):611–619, Mar. 1996.

A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, Sept. 2009.

B. David. *Caractérisations acoustiques de structures vibrantes par mise en atmosphère raréfiée*. PhD thesis, University of Paris VI, 1999.

A. Eriksson, P. Stoica, and T. Söderström. Second-order properties of MUSIC and ESPRIT estimates of sinusoidal frequencies in high SNR scenarios. *IEE Proceedings on Radar, Sonar and Navigation*, 140(4):266–272, Aug. 1993.

J. A. Fessler and A. O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing*, 42(10):2664–2677, Oct. 1994.

C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the beta-divergence. *Neural Computation*, 23(9):2421–2456, Sep. 2011.

C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis. *Neural Computation*, 21(3):793–830, Mar. 2009.

L. Finesso and P. Spreij. Approximate nonnegative matrix factorization via alternating minimization. In *Proceedings of International Symposium on Mathematical Theory of Networks and Systems*, July 2004.

J. J. Fuchs. Estimation of the number of signals in the presence of unknown correlated sensor noise. *IEEE Transactions on Signal Processing*, 40(5):1053–1061, May 1992.

B. Fuentes, R. Badeau, and G. Richard. Harmonic adaptive latent component analysis of audio and application to music transcription. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 21(9):1854–1866, Sept. 2013.

F. Gini and F. Bordoni. On the behavior of information theoretic criteria for model order selection of InSAR signals corrupted by multiplicative noise. *Signal Processing*, 83:1047–1063, 2003.

J. Grouffaud, P. Larzabal, and H. Clergeot. Some properties of ordered eigenvalues of a Wishart matrix: application in detection test and model order selection. In *Proc. of ICASSP'96*, volume 5, pages 2465–2468. IEEE, 1996.

R. Hennequin. Rapport à mi-parcours de travaux de thèse. Télécom ParisTech, Apr. 2010.

R. Hennequin, R. Badeau, and B. David. Time-dependent parametric and harmonic templates in non-negative matrix factorization. In *Proceedings of International Conference on Digital Audio Effects*, Sept. 2010.

R. Hennequin, R. Badeau, and B. David. NMF with time-frequency activations to model non-stationary audio events. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 19(4):744–753, May 2011.

K. Hermus, W. Verhelst, and P. Wambacq. Psychoacoustic modeling of audio with exponentially damped sinusoids. In *Proc. of ICASSP'02*, volume 2, pages 1821–1824. IEEE, 2002.

Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5):814–824, May 1990.

Y. Hua and T. K. Sarkar. On SVD for estimating generalized eigenvalues of singular matrix pencil in noise. *IEEE Transactions on Signal Processing*, 39(4):892–900, Apr. 1991.

A. Hurmalainen, R. Saeidi, and T. Virtanen. Similarity induced group sparsity for non-negative matrix factorisation. In *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, pages 4425–4429, Apr. 2015.

M. Jeanneau, P. Mouyon, and C. Pendaries. Sintrack analysis, application to detection and estimation of flutter for flexible structures. In *Proc. of EUSIPCO*, pages 789–792, Ile de Rhodes, Grèce, Sept. 1998.

J. Jensen, R. Heusdens, and S. H. Jensen. A perceptual subspace approach for modeling of speech and audio signals with damped sinusoids. *IEEE Transactions on Speech and Audio Processing*, 12(2):121–132, Mar. 2004.

H. Kameoka, N. Ono, K. Kashino, and S. Sagayama. Complex NMF: A new sparse representation for acoustic signals. In *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, pages 3437–3440, April 2009.

S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.

F. Keiler and S. Marchand. Survey on extraction of sinusoids in stationary sounds. In *Proc. of DAFx-02*, pages 51–58, Hambourg, Allemagne, Sept. 2002.

A. Kot, S. Parthasarathy, D. Tufts, and R. Vaccaro. The statistical performance of state-variable balancing and Prony's method in parameter estimation. In *Proc. of ICASSP'87*, volume 12, pages 1549–1552, Apr. 1987.

R. Kumaresan and D. W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(6):833–840, Dec. 1982.

D. Kundu and A. Mitra. Detecting the number of signals for an undamped exponential model using cross-validation approach. *Signal Processing*, 80(3):525–534, 2000.

S. Y. Kung, K. S. Arun, and D. B. Rao. State-space and singular value decomposition based approximation methods for harmonic retrieval problem. *J. of Opt. Soc. of America*, 73:1799–1811, Dec. 1983.

C. Lambourg and A. Chaigne. Measurements and modeling of the admittance matrix at bridge in guitars. In *Proc. of SMAC'93*, pages 449–453, Stockholm, Suède, July 1993.

J. Laroche. The use of the Matrix Pencil method for the spectrum analysis of musical signals. *Journal of the Acoustical Society of America*, 94(4):1958–1965, Oct. 1993.

H. Laurberg, M. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen. Theorems on positive data: On the uniqueness of NMF. *Computational Intelligence and Neuroscience*, 2008, 2008. Article ID 764206, 9 pages.

D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788–791, Oct. 1999.

D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Proceedings of Neural Information Processing Systems*, pages 556–562, Dec. 2001.

A. Lefevre, F. Bach, and C. Févotte. Itakura-Saito nonnegative matrix factorization with group sparsity. In *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, pages 21–24, May 2011.

Y. Li, K. Liu, and J. Razavilar. A parameter estimation scheme for damped sinusoidal signals based on low-rank Hankel approximation. *IEEE Transactions on Signal Processing*, 45:481–486, Feb. 1997.

A. P. Liavas and P. A. Regalia. On the behavior of Information Theoretic Criteria for model order selection. *IEEE Transactions on Signal Processing*, 49(8):1689–1695, Aug. 2001.

A. P. Liavas, P. A. Regalia, and J.-P. Delmas. Blind channel approximation: effective channel order determination. *IEEE Transactions on Signal Processing*, 47(12):3336–3344, Dec. 1999.

C.-J. Lin. Projected gradient methods for non-negative matrix factorization. *Neural Computation*, 19(10):2756–2779, Oct. 2007.

A. Liutkus and R. Badeau. Generalized Wiener filtering with fractional power spectrograms. In *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, pages 266–270, Apr. 2015.

A. Liutkus, D. Fitzgerald, and R. Badeau. Cauchy nonnegative matrix factorization. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2015.

S. Marcos, J. Sanchez-Araujo, N. Bertaux, P. Larzabal, and P. Forster. *Les Méthodes à haute résolution : traitement d'antenne et analyse spectrale. Chapitres 4 et 5*. Hermès, Paris, France, 1998. Ouvrage collectif sous la direction de S. Marcos.

M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama. Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with beta-divergence. In *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing*, pages 283–288, 2010.

J. Nieuwenhuijse, R. Heusens, and E. F. Deprettere. Robust exponential modeling of audio signals. In *Proc. of ICASSP'98*, volume 6, pages 3581–3584. IEEE, May 1998.

P. Stoica and T. Söderström. Statistical Analysis of MUSIC and Subspace Rotation Estimates of Sinusoidal Frequencies. *IEEE Transactions on Signal Processing*, 39:1836–1847, Aug. 1991.

V. F. Pisarenko. The retrieval of harmonics from a covariance function. *Geophysical J. Royal Astron. Soc.*, 33: 347–366, 1973.

C. R. Rao and L. C. Zhao. Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals. *IEEE Transactions on Signal Processing*, 41(3):1461–1464, Mar. 1993.

G.-M. Riche de Prony. Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool à différentes températures. *Journal de l'école polytechnique*, 1(22):24–76, 1795.

F. Rigaud, B. David, and L. Daudet. A parametric model and estimation techniques for the inharmonicity and tuning of the piano. *Journal of the Acoustical Society of America*, 133(5):3107–3118, May 2013.

J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.

R. Roy and T. Kailath. Total least squares ESPRIT. In *Proc. of 21st Asilomar Conference on Signals, Systems, and Computers*, pages 297–301, Nov. 1987.

R. Roy, A. Paulraj, and T. Kailath. ESPRIT–A subspace rotation approach to estimation of parameters of cisoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1340–1342, Oct. 1986.

M. N. Schmidt and H. Laurberg. Non-negative matrix factorization with Gaussian process priors. *Computational Intelligence and Neuroscience*, 2008:1–10, 2008. Article ID 361705.

R. O. Schmidt. *A signal subspace approach to multiple emitter location and spectral estimation*. PhD thesis, Stanford University, Stanford, Californie, USA, Nov. 1981.

R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280, Mar. 1986.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

M. Shashanka, B. Raj, and P. Smaragdis. Probabilistic latent variable models as nonnegative factorizations. *Computational Intelligence and Neuroscience*, 2008:1–8, 2008. Article ID 947438.

U. Simsekli and A. T. Cemgil. Markov chain Monte Carlo inference for probabilistic latent tensor factorization. In *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing*, pages 1–6, 2012.

P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 177–180, Oct. 2003.

P. Smaragdis, B. Raj, and M. Shashanka. Missing data imputation for time-frequency representations of audio signals. *Journal of Signal Processing Systems*, 65:361–370, Aug 2011.

P. Stoica and A. Nehorai. Study of the statistical performance of the Pisarenko harmonic decomposition method. *IEE Proceedings Radar and Signal Processing*, 135(2):161–168, Apr. 1988.

P. Stoica, H. Li, and J. li. Amplitude estimation of sinusoidal signals: survey, new results, and an application. *IEEE Transactions on Signal Processing*, 48(2):338–352, 2000.

A.-J. Van der Veen, E. F. Deprettere, and A. L. Swindlehurst. Subspace based signal analysis using singular value decomposition. *Proc. of IEEE*, 81(9):1277–1308, Sept. 1993.

E. Vincent, N. Bertin, and R. Badeau. Adaptive harmonic spectral decomposition for multiple pitch estimation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 18(3):528–537, Mar. 2010.

T. Virtanen, A. T. Cemgil, and S. Godsill. Bayesian extensions to non-negative matrix factorisation for audio signal modelling. In *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, pages 1825–1828, Apr. 2008.

M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, Apr. 1985.

G. Weinreich. Coupled piano strings. *Journal of the Acoustical Society of America*, 62(6):1474–1484, 1977.

Q. T. Zhang and K. M. Wong. Information theoretic criteria for the determination of the number of signals in spatially correlated noise. *IEEE Transactions on Signal Processing*, 41(4):1652–1663, Apr. 1993.

L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. On detection of the number of signals in presence of white noise. *Journal of Multivariate Analysis*, 20(1):1–25, 1986a.

L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. On detection of the number of signals when the noise covariance matrix is arbitrary. *Journal of Multivariate Analysis*, 20(1):26–49, 1986b.

M. Zoltawski and D. Stavrinides. Sensor array signal processing via a Procrustes rotations based eigen-analysis of the ESPRIT data pencil. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(6):832–861, June 1989.

Institut Mines-Télécom

Exercices sur les méthodes à haute résolution

Roland Badeau
roland.badeau@telecom-paristech.fr

On considère le modèle de signal complexe *Exponential Sinusoidal Model* (ESM)

$$s[t] = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} e^{i(2\pi f_k t + \phi_k)},$$

où à chaque fréquence $f_k \in \left]-\frac{1}{2}, \frac{1}{2}\right]$ est associée une amplitude réelle $a_k > 0$, une phase $\phi_k \in]-\pi, \pi]$, et un facteur d'amortissement $\delta_k \in \mathbb{R}$. En définissant les amplitudes complexes $\alpha_k = a_k e^{i\phi_k}$ et les pôles complexes $z_k = e^{\delta_k + i2\pi f_k}$, ce modèle se réécrit sous la forme

$$s[t] = \sum_{k=0}^{K-1} \alpha_k \, {z_k}^t.$$

En pratique, le signal observé $x[t]$ ne satisfait jamais rigoureusement ce modèle. On le modélise comme la somme du signal $s[t]$ et d'un bruit blanc gaussien complexe $b[t]$ de variance σ^2 :

$$x[t] = s[t] + b[t].$$

Remarque : Un bruit blanc gaussien complexe de variance σ^2 est un processus complexe dont la partie réelle et la partie imaginaire sont deux bruits blancs gaussiens de même variance $\frac{\sigma^2}{2}$, indépendants l'un de l'autre.

On suppose que le signal $x[t]$ est observé sur l'intervalle temporel $\{0 \dots N-1\}$ de longueur $N > 2K$. On considère deux entiers n et l tels que $n > K$, $l > K$, et $N = n + l - 1$.

On définit alors la matrice de Hankel de dimension $n \times l$ contenant les N échantillons du signal observé :

$$\boldsymbol{X} = \begin{bmatrix} x[0] & x[1] & \dots & x[l-1] \\ x[1] & x[2] & \dots & x[l] \\ \vdots & \vdots & \vdots & \vdots \\ x[n-1] & x[n] & \dots & x[N-1] \end{bmatrix}.$$

On définit de même les matrices de Hankel $\boldsymbol{S}$ et $\boldsymbol{B}$ de mêmes dimensions $n \times l$, à partir des échantillons des signaux $s[t]$ et $b[t]$ respectivement.

Notations :

— $\boldsymbol{X}^T$: transposé de la matrice $\boldsymbol{X}$,
— $\boldsymbol{X}^*$: conjugué de la matrice $\boldsymbol{X}$,
— $\boldsymbol{X}^H$: conjugué hermitien de la matrice $\boldsymbol{X}$ (c'est-à-dire transposé et conjugué).

1 *Mu*ltiple *Si*gnal *C*lassification (MUSIC)

Question 1 Pour tout $k \in \{0 \dots K-1\}$, on considère la composante $s_k[t] = \alpha_k \, {z_k}^t$. On définit alors la matrice de Hankel de dimensions $n \times l$

$$\boldsymbol{S_k} = \begin{bmatrix} s_k[0] & s_k[1] & \dots & s_k[l-1] \\ s_k[1] & s_k[2] & \dots & s_k[l] \\ \vdots & \vdots & \vdots & \vdots \\ s_k[n-1] & s_k[n] & \dots & s_k[N-1] \end{bmatrix}$$

Roland Badeau roland.badeau@telecom-paristech.fr

Contexte académique } sans modifications
Voir page 4

Pour tout $z \in \mathbb{C}$, on définit le vecteur $\boldsymbol{v}^n(z) = [1,\, z,\, z^2, \ldots,\, z^{n-1}]^T$, de dimension n, et le vecteur $\boldsymbol{v}^l(z) = [1,\, z,\, z^2, \ldots,\, z^{l-1}]^T$, de dimension l. Vérifier alors que $\boldsymbol{S_k} = \alpha_k\, \boldsymbol{v}^n(z_k)\, \boldsymbol{v}^l(z_k)^T$.

Question 2 Utiliser le résultat de la question 1 pour démontrer que $\boldsymbol{S} = \sum_{k=0}^{K-1} \alpha_k\, \boldsymbol{v}^n(z_k)\, \boldsymbol{v}^l(z_k)^T$. Vérifier que cette dernière égalité peut se réécrire sous la forme $\boldsymbol{S} = \boldsymbol{V}^n\, \boldsymbol{A}\, \boldsymbol{V}^{lT}$, où

— $\boldsymbol{V}^n$ est la matrice de Vandermonde de dimensions $n \times K$:

$$\boldsymbol{V}^n = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_{K-1} \\ {z_0}^2 & {z_1}^2 & \ldots & {z_{K-1}}^2 \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{n-1} & {z_1}^{n-1} & \ldots & {z_{K-1}}^{n-1} \end{bmatrix}$$

— $\boldsymbol{V}^l$ est la matrice de Vandermonde de dimensions $l \times K$,
— $\boldsymbol{A} = \mathrm{diag}(\alpha_0,\, \alpha_1, \ldots,\, \alpha_{K-1})$ est une matrice diagonale de dimension $K \times K$.

Question 3 On définit la matrice $\boldsymbol{R}_{ss} = \frac{1}{l}\, \boldsymbol{S}\, \boldsymbol{S}^H$. Démontrer que $\boldsymbol{R}_{ss}$ est une matrice symétrique hermitienne et positive. Vérifier que $\boldsymbol{R}_{ss}$ peut être factorisée sous la forme $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\, \boldsymbol{P}\, \boldsymbol{V}^{nH}$, où $\boldsymbol{P}$ est une matrice symétrique hermitienne et définie positive, de dimension $K \times K$. En déduire que le rang de la matrice $\boldsymbol{R}_{ss}$ est égal à K (on rappelle que les pôles z_k sont distincts deux à deux).

Question 4 Démontrer que la matrice $\boldsymbol{R}_{ss}$ est diagonalisable dans une base orthonormée, et que ses valeurs propres $\{\lambda_i\}_{i=0\ldots n-1}$ sont positives. En les supposant rangées par ordre décroissant et en utilisant le résultat de la question 3, en déduire que

— $\forall i \in \{0 \ldots K-1\},\, \lambda_i > 0$;
— $\forall i \in \{K \ldots n-1\},\, \lambda_i = 0$.

Question 5 On pose $\widehat{\boldsymbol{R}}_{xx} = \frac{1}{l}\, \boldsymbol{X}\, \boldsymbol{X}^H$ et $\boldsymbol{R}_{xx} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{xx}\right]$. De même, on pose $\widehat{\boldsymbol{R}}_{bb} = \frac{1}{l}\, \boldsymbol{B}\, \boldsymbol{B}^H$ et $\boldsymbol{R}_{bb} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{bb}\right]$. En utilisant l'égalité $\boldsymbol{X} = \boldsymbol{S} + \boldsymbol{B}$ et le fait que le bruit est centré, démontrer que $\boldsymbol{R}_{xx} = \boldsymbol{R}_{ss} + \boldsymbol{R}_{bb}$. Prouver que pour un bruit blanc gaussien complexe, $\boldsymbol{R}_{bb} = \sigma^2\, \boldsymbol{I}_n$.

Question 6 Pour tout $i \in \{0 \ldots n-1\}$, on note $\boldsymbol{w}_i$ le vecteur propre de la matrice $\boldsymbol{R}_{ss}$ associé à la valeur propre λ_i. En utilisant le résultat de la question 5, vérifier que $\boldsymbol{w}_i$ est aussi vecteur propre de $\boldsymbol{R}_{xx}$ associé à la valeur propre $\lambda'_i = \lambda_i + \sigma^2$. On en déduit que

— $\forall i \in \{0 \ldots K-1\},\, \lambda'_i > \sigma^2$;
— $\forall i \in \{K \ldots n-1\},\, \lambda'_i = \sigma^2$.

Question 7 On note $\boldsymbol{W}$ la matrice $[\boldsymbol{w}_0 \ldots \boldsymbol{w}_{K-1}]$, et $\boldsymbol{W}_\perp$ la matrice $[\boldsymbol{w}_K \ldots \boldsymbol{w}_{n-1}]$. Démontrer que $\mathrm{Im}(\boldsymbol{W}) = \mathrm{Im}(\boldsymbol{V}^n)$ (on commencera par prouver que $\mathrm{Im}(\boldsymbol{W}) \subset \mathrm{Im}(\boldsymbol{V}^n)$).

Roland Badeau roland.badeau@telecom-paristech.fr

Contexte académique } sans modifications
Voir page 4

Remarque : L'espace engendré par $\boldsymbol{W}_\perp$ est un sous-espace propre de la matrice $\boldsymbol{R}_{xx}$ associé à la valeur propre σ^2. C'est pour cela qu'il est appelé *espace bruit*. L'espace engendré par $\boldsymbol{W}$ est aussi celui engendré par la matrice de Vandermonde $\boldsymbol{V}^n$. Il caractérise donc entièrement les K pôles du signal, c'est pourquoi on l'appelle *espace signal*. En revanche, les valeurs propres de $\boldsymbol{R}_{xx}$ correspondant à l'espace signal sont toutes surélevées de σ^2, ce qui signifie que cet espace contient aussi du bruit.

Question 8 Prouver que les pôles $\{z_k\}_{k\in\{0\dots K-1\}}$ sont les solutions de l'équation $\left\|\boldsymbol{W}_\perp^H \boldsymbol{v}^n(z)\right\|^2 = 0$.

Remarque : Dans la pratique, les signaux réels ne correspondent pas rigoureusement au modèle, et cette équation n'est jamais vérifiée. C'est pourquoi la méthode d'estimation des pôles baptisée spectral-MUSIC consiste à rechercher les K pics les plus élevés de la fonction $z \mapsto \frac{1}{\left\|\boldsymbol{W}_\perp^H \boldsymbol{v}^n(z)\right\|^2}$. Elle est ainsi plus facile à implémenter que la méthode du maximum de vraisemblance, qui requiert l'optimisation d'une fonction de K variables complexes.

2 *E*stimation of *S*ignal *P*arameters via *R*otational *I*nvariance *T*echniques (ESPRIT)

Soit $\boldsymbol{V}_\downarrow^n$ la matrice de dimension $(n-1)\times K$ qui contient les $n-1$ premières lignes de $\boldsymbol{V}^n$, et $\boldsymbol{V}_\uparrow^n$ la matrice de dimension $(n-1)\times K$ qui contient les $n-1$ dernières lignes de $\boldsymbol{V}^n$. De même, soit $\boldsymbol{W}_\downarrow$ la matrice de dimension $(n-1)\times K$ qui contient les $n-1$ premières lignes de $\boldsymbol{W}$, et $\boldsymbol{W}_\uparrow$ la matrice de dimension $(n-1)\times K$ qui contient les $n-1$ dernières lignes de $\boldsymbol{W}$.

Question 1 Vérifier que les matrices $\boldsymbol{V}_\downarrow^n$ et $\boldsymbol{V}_\uparrow^n$ satisfont l'égalité $\boldsymbol{V}_\uparrow^n = \boldsymbol{V}_\downarrow^n \boldsymbol{D}$, où $\boldsymbol{D}$ est une matrice diagonale de dimension $K\times K$ dont on donnera les coefficients diagonaux.

Question 2 Prouver qu'il existe une matrice $\boldsymbol{G}$ inversible de dimension $K\times K$ telle que $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$ (on ne demande pas d'exprimer $\boldsymbol{G}$, mais seulement de démontrer son existence). Vérifier alors que $\boldsymbol{V}_\downarrow^n = \boldsymbol{W}_\downarrow \boldsymbol{G}$ et $\boldsymbol{V}_\uparrow^n = \boldsymbol{W}_\uparrow \boldsymbol{G}$.

Question 3 En déduire qu'il existe une matrice inversible $\boldsymbol{\Phi}$ telle que $\boldsymbol{W}_\uparrow = \boldsymbol{W}_\downarrow \boldsymbol{\Phi}$. Quelles sont les valeurs propres de $\boldsymbol{\Phi}$?

Question 4 En supposant que la matrice $\boldsymbol{W}_\downarrow^H \boldsymbol{W}_\downarrow$ est inversible, exprimer $\boldsymbol{\Phi}$ en fonction de $\boldsymbol{W}_\downarrow$ et $\boldsymbol{W}_\uparrow$.

Question 5 En déduire une méthode d'estimation des pôles $\{z_k\}_{k\in\{0\dots K-1\}}$.

Remarque : Le principal avantage de cette méthode par rapport à spectral-MUSIC est qu'il n'est plus nécessaire d'optimiser une fonction pour déterminer les pôles, ceux-ci étant obtenus par un calcul direct.

Roland Badeau roland.badeau@telecom-paristech.fr

Contexte académique } sans modifications

Institut Mines-Télécom

TP Analyse et synthèse de sons de cloche

Roland Badeau

Les fichiers nécessaires à ce TP sont disponibles sur le Moodle ATIAM. Vous pourrez charger un fichier son sous Matlab en tapant la commande `[x,Fs] = wavread('cloche.wav')`. Pour l'écouter, vous pourrez taper `soundsc(x,Fs)`. En Python, vous pourrez utiliser le notebook qui vous est fourni `template-TP-HR.ipynb`.

1 Introduction

Les cloches sont parmi les instruments de musique les plus anciens et le son qu'elle produisent est souvent évocateur parce qu'il a bercé le quotidien des générations depuis environ 3000 ans, accompagnant petits et grands événements. Cette évocation tient en partie à la structure du spectre sonore : les modes propres de vibration sont en général accordés par les facteurs de cloches (dans le cas des bons instruments) pour que leurs fréquences suivent une série particulière, qui comporte notamment la tierce mineure (Mi bémol si la cloche est en Do). Cette série n'est pas harmonique mais les rapports entre fréquences propres sont tels qu'on perçoit une hauteur bien définie (celle à laquelle on chanterait la note entendue). Notamment la présence de la série 2-3-4, forte au début du son, vient renforcer la sensation de hauteur au voisinage du fondamental. Cette sensation est liée à un effet psychoacoustique (traitement du signal reçu par le cerveau).

Soit f_p la fréquence correspondant à la hauteur perçue. L'analyse de la série des fréquences propres se traduit par la donnée d'un tableau d'environ 15 facteurs $\alpha_n = f_n/f_p$. Un ordre de grandeur est donné ci-après : 0.5 (bourdon), 1 (fondamental), 1.2 (tierce mineure), 1.5 (quinte), 2 (nominal), 2.5, 2.6, 2.7, 3, 3.3, 3.7, 4.2 (faux double octave), 4.5, 5, 5.9. Le timbre du son correspondant dépend de l'amplitude et de la décroissance de chacun de ces partiels.

Ce TP vise à mettre en oeuvre une méthode d'estimation spectrale à haute résolution pour effectuer l'analyse / synthèse de sons de cloche. Comme on peut le constater sur la figure 1, ce type de sons présente une forte décroissance temporelle.

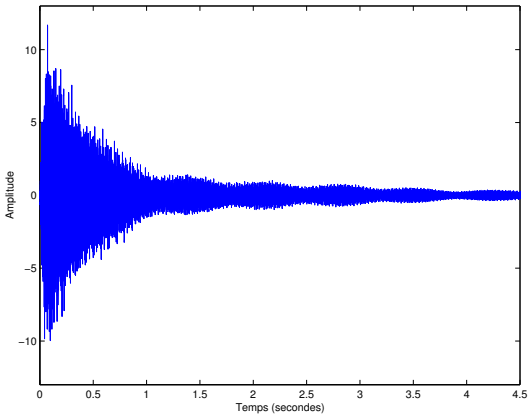


FIGURE 1 – Son de cloche

Pour tenir compte de cette atténuation, on utilise le modèle *Exponential Sinusoidal Model* (ESM) :

$$s[t] = \sum_{k=0}^{K-1} a_k \, e^{\delta_k t} e^{i(2\pi f_k t + \phi_k)},$$

où à chaque fréquence $f_k \in \left]-\frac{1}{2}, \frac{1}{2}\right]$ est associée une amplitude réelle $a_k > 0$, une phase $\phi_k \in]-\pi, \pi]$, et un facteur d'amortisement $\delta_k \in \mathbb{R}$. En définissant les amplitudes complexes $\alpha_k = a_k e^{i\phi_k}$ et les pôles complexes $z_k = e^{\delta_k + i2\pi f_k}$, ce modèle se réécrit sous la forme

$$s[t] = \sum_{k=0}^{K-1} \alpha_k \, {z_k}^t.$$

Le cas $\delta_k < 0$ permet donc de modéliser des sinusoïdes à décroissance exponentielle, ce qui correspond aux solutions des équations de propagation physique. Les paramètres du modèle sont alors $\{\delta_k, f_k, a_k, \phi_k\}_{k\in\{0\ldots K-1\}}$. Pour les estimer, nous utiliserons la méthode ESPRIT présentée dans le cours. Dans un premier temps, nous l'appliquerons à un signal synthétique pour mettre en évidence la supériorité des méthodes à haute résolution sur l'analyse de Fourier en terme de résolution spectrale. Dans un deuxième temps, cette méthode sera appliquée à des sons de cloche.

2 Rappels de Matlab/Python

En Matlab :
- `A'` : conjugué hermitien de la matrice $\boldsymbol{A}$ (inclut la conjugaison complexe) ;
- `A.'` : transposé de la matrice $\boldsymbol{A}$ (sans conjugaison complexe) ;
- `A(l1:l2,c1:c2)` : matrice extraite de $\boldsymbol{A}$ entre les lignes l_1 et l_2 (inclue) et les colonnes c_1 et c_2 (inclue).

En Python :
- `A.conj().T` : conjugué hermitien de la matrice $\boldsymbol{A}$ (inclut la conjugaison complexe) ;
- `A.T` : transposé de la matrice $\boldsymbol{A}$ (sans conjugaison complexe) ;
- `A[l1:l2,c1:c2]` : matrice extraite de $\boldsymbol{A}$ entre les lignes l_1 et l_2 (exclue) et les colonnes c_1 et c_2 (exclue).

3 Signal synthétique

Nous nous intéressons ici à un signal synthétique de longueur N, constitué d'une somme de deux exponentielles complexes, dont les fréquences sont séparées d'un intervalle $\Delta f = \frac{1}{N}$ (ce qui correspond à la limite de la résolution de l'analyse de Fourier). Les phases seront tirées aléatoirement. On ne bruitera pas ce signal, de sorte que le signal observé $x[t]$ sera égal à $s[t]$. On choisira par exemple les paramètres suivants : $N = 63$, $f_0 = \frac{1}{4}$, $f_1 = f_0 + \frac{1}{N}$, $a_0 = 1$, $a_1 = 10$, $\delta_0 = 0$, $\delta_1 = -0.05$. Pour le synthétiser, on pourra utiliser la fonction fournie

```
x = Synthesis(N,delta,f,a,phi);
```

dont les arguments sont N, le vecteur `delta` des atténuations δ_k, le vecteur `f` des fréquences f_k, le vecteur `a` des amplitudes a_k, et le vecteur `phi` des phases ϕ_k.

3.1 Analyse spectrale par transformation de Fourier

Observer le périodogramme de ce signal. On examinera brièvement la séparabilité des deux raies spectrales, sans zero-padding ($N_{\text{fft}} = N$) et avec zero-padding ($N_{\text{fft}} = 1024 > N$).

3.2 Méthodes à haute résolution

On se propose d'écrire des fonctions

`MUSIC(x,n,K)` et `[delta,f] = ESPRIT(x,n,K)`

qui analysent le signal x de longueur N en utilisant les méthodes MUSIC et ESPRIT, avec un espace signal de dimension K et un espace bruit de dimension $n - K$, et des vecteurs de données de longueur n comprise entre $K + 1$ et $N - K + 1$. Pour traiter nos signaux synthétiques, on pourra choisir $n = 32$ et $K = 2$. Les deux méthodes partagent les étapes suivantes :

1. **Calcul de la matrice de corrélation**
   La matrice de corrélation du signal observé est définie par la relation
   $$\widehat{\boldsymbol{R}}_{xx} = \frac{1}{l}\boldsymbol{X}\,\boldsymbol{X}^H$$
   où $\boldsymbol{X}$ est une matrice de Hankel de dimension $n \times l$ contenant les $N = n + l - 1$ échantillons du signal :
   $$\boldsymbol{X} = \begin{bmatrix} x[0] & x[1] & \dots & x[l-1] \\ x[1] & x[2] & \dots & x[l] \\ \vdots & \vdots & \ddots & \vdots \\ x[n-1] & x[n] & \dots & x[N-1] \end{bmatrix}$$
   La matrice $\boldsymbol{X}$ pourra être construite avec la fonction `hankel`.
2. **Estimation de l'espace signal**
   On pourra diagonaliser la matrice $\widehat{\boldsymbol{R}}_{xx}$ à l'aide de la commande `[U1,Lambda,U2] = svd(Rxx)`. La matrice $\widehat{\boldsymbol{R}}_{xx}$ étant symétrique positive, les vecteurs colonnes des matrices $\boldsymbol{U}_1$ et $\boldsymbol{U}_2$ (de dimension $n \times n$) sont vecteurs propres de $\widehat{\boldsymbol{R}}_{xx}$, associés aux n valeurs propres rangées dans la matrice diagonale $\boldsymbol{\Lambda}$ par ordre décroissant (on a alors $\widehat{\boldsymbol{R}}_{xx} = \boldsymbol{U}_1\,\boldsymbol{\Lambda}\,\boldsymbol{U}_1^H = \boldsymbol{U}_2\,\boldsymbol{\Lambda}\,\boldsymbol{U}_2^H$). On pourra ainsi extraire de $\boldsymbol{U}_1$ (ou de $\boldsymbol{U}_2$) une base de l'espace signal $\boldsymbol{W}$ (de dimension $n \times K$).

3.2.1 Algorithme ESPRIT

L'algorithme ESPRIT vise dans un premier temps à estimer les fréquences et les facteurs d'amortissement :

3. **Estimation des fréquences et des facteurs d'amortissement**
   Pour estimer les fréquences, on pourra procéder de la façon suivante :
   — extraire de $\boldsymbol{W}$ les matrices $\boldsymbol{W}_\downarrow$ (obtenue en supprimant la dernière ligne de $\boldsymbol{W}$) et $\boldsymbol{W}_\uparrow$ (obtenue en supprimant la première ligne de $\boldsymbol{W}$) ;
   — calculer $\boldsymbol{\Phi} = \left(\left(\boldsymbol{W}_\downarrow^H\boldsymbol{W}_\downarrow\right)^{-1}\boldsymbol{W}_\downarrow^H\right)\boldsymbol{W}_\uparrow = \boldsymbol{W}_\downarrow^\dagger\,\boldsymbol{W}_\uparrow$, où le symbole † est l'opérateur de pseudo inverse (fonction `pinv` de Matlab ou fonction `numpy.linalg.pinv` de Python).
   — calculer les valeurs propres de $\boldsymbol{\Phi}$ à l'aide de la fonction `eig` de Matlab ou de la fonction `numpy.linalg.eig` de Python (on rappelle que les valeurs propres de $\boldsymbol{\Phi}$ sont les pôles $z_k = e^{\delta_k + i2\pi f_k}$). En déduire $\delta_k = \ln(|z_k|)$ et $f_k = \frac{1}{2\pi}\,\mathrm{angle}(z_k)$.
4. **Estimation des amplitudes et des phases**
   Il s'agit à présent d'écrire une fonction

```
[a,phi] = LeastSquares(x,delta,f)
```

qui estime les amplitudes a_k et les phases ϕ_k par la méthode des moindres carrés, connaissant le signal x, les atténuations δ_k et les fréquences f_k. Les amplitudes complexes sont ainsi déterminées par la relation

$$\boldsymbol{\alpha} = \left((\boldsymbol{V}^{N\,H} \boldsymbol{V}^N)^{-1} \boldsymbol{V}^{N\,H} \right) \boldsymbol{x} = \boldsymbol{V}^{N\,\dagger}\, \boldsymbol{x} \tag{1}$$

où $\boldsymbol{x}$ est le vecteur $[x[0], \ldots, x[N-1]]^T$ et $\boldsymbol{V}^N$ est la matrice de Vandermonde de dimension $N \times K$, dont les coefficients vérifient la relation $\boldsymbol{V}^N_{(t,k)} = {z_k}^t$ pour tous $(t,k) \in \{0 \ldots N-1\} \times \{0 \ldots K-1\}$. Pour calculer la matrice $\boldsymbol{V}^N$, il est possible d'éviter d'utiliser une boucle `for` en remarquant que $\ln\left(\boldsymbol{V}^N_{(t,k)}\right) = t(\delta_k + i2\pi f_k)$. Ainsi, la matrice contenant les coefficients $\ln\left(\boldsymbol{V}^N_{(t,k)}\right)$ s'exprime comme le produit d'un vecteur colonne par un vecteur ligne. On en déduira $a_k = |\alpha_k|$ et $\phi_k = \text{angle}(\alpha_k)$ pour tout $k \in \{0 \ldots K-1\}$.

5. **Application aux signaux synthétiques**

   Appliquer les fonctions `ESPRIT` et `LeastSquares` au signal précédemment synthétisé. Commenter.

3.2.2 Méthode MUSIC

On rappelle que le pseudo-spectre MUSIC est défini par $P(z) = \dfrac{1}{\|\boldsymbol{W}_\perp^H \boldsymbol{v}^n(z)\|^2}$.

6. **Affichage du pseudo-spectre MUSIC**

   Écrire une fonction `MUSIC(x,n,K)` qui affiche le logarithme du pseudo-spectre comme une fonction des deux variables $f \in [0,1]$ et $\delta \in [-0.1, 0.1]$ (on pourra utiliser la fonction `surf` de Matlab ou la fonction `plot_surface` de Python Matplotlib). Appliquer la fonction `MUSIC` au signal précédemment synthétisé, et vérifier que le pseudo-spectre fait bien apparaître les deux pôles $z_k = e^{\delta_k + i2\pi f_k}$.

4 Signaux audio

Nous nous proposons maintenant d'appliquer les fonctions développées dans la partie précédente à des sons de cloche.

4.1 Analyse spectrale par transformation de Fourier

Examiner le périodogramme du signal `cloche.wav` et comparer la série de ses fréquences propres aux valeurs données en introduction.

4.2 Méthode à haute résolution

Il s'agit à présent d'appliquer l'algorithme ESPRIT à ce signal. On posera $K = 54$, $n = 512$ et $l = 2n = 1024$ (d'où $N = n + l - 1 = 1535$).

Afin de garantir que le modèle de signal soit vérifié sur la fenêtre d'analyse (amortissement exponentiel), on extraira un segment de longueur N dont le début sera postérieur au maximum de l'enveloppe de la forme d'onde. On pourra ainsi commencer au $10000^{\text{ème}}$ échantillon.

Appliquer la fonction `ESPRIT` au signal extrait afin d'en estimer les fréquences propres et les atténuations correspondantes. En déduire les amplitudes et les phases à l'aide de la fonction `LeastSquares`. Enfin, écouter le signal resynthétisé à l'aide de la fonction `Synthesis` (sur une durée plus longue que le segment extrait, afin de bien mettre en évidence les résonnances du son), et commenter.

Institut Mines-Télécom

TP Factorisations en matrices positives

Roland Badeau
d'après les sujets écrits par U. Simsekli, B. David et P. Magron

Les fichiers nécessaires à ce TP sont disponibles sur le Moodle ATIAM. Vous pourrez charger un fichier son sous Matlab en tapant la commande `[x,Fs] = wavread('file.wav')`. Pour l'écouter, vous pourrez taper `soundsc(x,Fs)`. Un modèle de notebook Python `NMF_TP.ipynb` est aussi fourni.

1 Factorisation en matrices positives avec la β-divergence

Dans ce sujet de travaux pratiques, nous traiterons de la factorisation en matrices positives (NMF) avec la β-divergence. Le problème que nous cherchons à résoudre est énoncé comme suit :

$$(\mathbf{W}^\star, \mathbf{H}^\star) = \text{argmin}_{\mathbf{W}\geq 0, \mathbf{H}\geq 0} \sum_{f=1}^{F}\sum_{t=1}^{T} d_\beta(x_{ft}\|\hat{x}_{ft}), \tag{1}$$

où x_{ft} est un coefficient de $\mathbf{X} \in \mathbb{R}_+^{F\times T}$, c'est-à-dire la matrice de données *positive*, et $\mathbf{W} \in \mathbb{R}_+^{F\times R}$ et $\mathbf{H} \in \mathbb{R}_+^{R\times T}$ sont les matrices de facteurs *positifs* inconnus. Nous définissons également $\hat{x}_{ft} = \sum_r w_{fr}h_{rt}$. La fonction de coût que nous minimisons s'appelle la β-divergence, qui est définie comme suit :

$$d_\beta(x\|\hat{x}) = \frac{x^\beta}{\beta(\beta-1)} - \frac{x\hat{x}^{\beta-1}}{\beta-1} + \frac{\hat{x}^\beta}{\beta}. \tag{2}$$

Lorsque $\beta = 2$, on obtient la distance Euclidienne au carré (EUC), lorsque $\beta = 1$, on obtient la divergence de Kullback-Leibler (KL), lorsque $\beta = 0$ on obtient la divergence d'Itakura-Saito (IS).

L'un des algorithmes les plus populaires pour la NMF est constitué des règles de mise à jour multiplicatives (MUR). L'algorithme MUR comporte les règles de mise à jour suivantes :

$$\mathbf{W} \leftarrow \mathbf{W} \circ \frac{(\mathbf{X}\circ\hat{\mathbf{X}}^{\beta-2})\mathbf{H}^\top}{\hat{\mathbf{X}}^{\beta-1}\mathbf{H}^\top} \tag{3}$$

$$\mathbf{H} \leftarrow \mathbf{H} \circ \frac{\mathbf{W}^\top(\mathbf{X}\circ\hat{\mathbf{X}}^{\beta-2})}{\mathbf{W}^\top\hat{\mathbf{X}}^{\beta-1}}, \tag{4}$$

où $\circ$ désigne la multiplication coefficient par coefficient et / et $\frac{\cdot}{\cdot}$ désignent la division coefficient par coefficient. En suivant la technique que nous avons utilisée dans le cours, dérivez l'algorithme MUR par vous-même.

2 Variantes autour d'un exemple simple

Il s'agit ici de programmer la NMF telle que décrite ci-dessus. Pour cela, on s'appuiera sur le script à compléter `test_nmf.m` qui gère la partie affichage.

2.1 Construction de l'exemple simple

Nous proposons de synthétiser la NMF d'un exemple musical simple : un DO, suivi d'un MI, suivi d'un accord DO/MI, ce que nous noterons DO-MI-DO/MI.

Cet exemple est obtenu en construisant la matrice $\mathbf{W}$ de la base spectrale et la matrice $\mathbf{H}$ des activations de manière synthétique. Aucun son n'est donc calculé ici. Pour cela on pourra par exemple :

— utiliser la fréquence d'échantillonnage $F_e = 8000$ Hz, l'ordre de TFD : $N_{\text{fft}} = 512$.

— calculer les fréquences fondamentales F_0, du DO et du MI placés sous le LA 440.

Roland Badeau
d'après les sujets écrits par U. Simsekli, B. David et P. Magron

— pour chaque note r, fabriquer un spectre harmonique (non négatif) de la forme $\mathbf{W}_r(f) = \sum_{k=1}^{K_r} a_k w(f - kF_{0,r})$ où w est une fenêtre usuelle (Hann de largeur 80 Hz par exemple) et a_k décrit l'amplitude des raies (par exemple décroissante, de la forme $a_k = exp(-k/K_r)$). K_r est fixé à l'aide du critère de Shannon. En déduire la matrice spectrale $\mathbf{W}_s$.
— pour les activations, on utilisera une fonction $h_r(t) = \sum_p b_{r,p} h(t - p\Delta T)$ où $h(t) = e^{-t/\tau}$ pour $t \in [0, \Delta T]$ avec ΔT de l'ordre de 0.5 secondes et $h(t) = 0$ ailleurs. b_p vaut 0 ou 1. En déduire la matrice $\mathbf{H}_s$. Attention, pour se mettre dans des conditions réalistes, on supposera que la représentation est obtenue avec des trames de longueur $N = N_{\text{fft}}$ et un recouvrement de 75%, soit un taux d'échantillonnage de $F_e/(N_{\text{fft}}/4)$ pour les activations temporelles. τ sera pris de l'ordre de $\Delta T/3$.
— construire alors la matrice temps-fréquence $\mathbf{X}_s = \mathbf{W}_s\mathbf{H}_s$.

2.2 NMF classique

— Utiliser les règles de mise à jour multiplicatives pour factoriser en produit de matrices non-négatives la matrice $\mathbf{X}_s$. Afin d'éviter que les indéterminations de la NMF ne conduisent à des problèmes numériques, à la fin de chaque itération on renormalisera les colonnes de $\mathbf{W}$ et on multipliera les lignes de $\mathbf{H}$ par les facteurs adéquats afin de conserver le produit $\mathbf{WH}$ inchangé. Les matrices $\mathbf{W}$ et $\mathbf{H}$ seront initialisées avec des valeurs aléatoires.
— Tracer le graphe de la fonction de coût en fonction de l'itération. Que constatez vous ?

2.3 Variantes

Testez votre algorithme en faisant varier les données et les paramètres :
— en répétant avec plusieurs tirages d'initialisations (quelles différences constatez-vous ?)
— avec ajout de bruit (attention à respecter la non-négativité, utiliser `rand`)
— avec différentes valeurs de β (0, 1 et 2 typiquement)
— avec plus de notes et des indéterminations du type DO/MI, DO/SOL, DO/MI/SOL
— en fabriquant des notes synthétiques (formes d'onde) par somme de sinusoïdes et en calculant la TFCT du signal temporel obtenu.

3 Transcription automatique à l'aide de la NMF semi-supervisée

Les notes isolées d'un piano vous sont fournies ainsi qu'un morceau réalisé à l'aide du même instrument. La NMF semi-supervisée consiste à définir la matrice spectrale $\mathbf{W}$ à l'aide des notes isolées (fournies sous la forme de `xxx.wav` où `xxx` est le numéro MIDI, 21-108 pour la tessiture du piano), en calculant simplement une estimation de la densité spectrale de puissance associée. La seule mise à jour restante est donc celle de $\mathbf{H}$. On travaillera sur des signaux sous-échantillonnés à 22 kHz.

Le script Matlab ou la fonction Python fourni(e) `build_signal` permet, à l'aide d'une syntaxe simple, de fabriquer un morceau à partir de notes définies par leur numéro MIDI, leur temps d'attaque, leur durée et la vélocité associée.

1. Constituer la base $\mathbf{W}$ correspondant aux 88 notes du piano
2. Fabriquer un signal test correspondant à la gamme C3-C4.
3. Programmer la NMF semi-supervisée à l'aide des résultats de la partie précédente et l'appliquer au signal test. Observer les activations temporelles obtenues.

4. Fabriquer une fonction de détection des attaques à appliquer sur les activations pour connaître les instants où les notes sont jouées.

Roland Badeau
d'après les sujets écrits par U. Simsekli, B. David et P. Magron

ATIAM : Examen de Traitement du Signal Musical Partie Applications et ouvertures

Geoffroy Peeters, Roland Badeau, Mardi 9 janvier 2024
RENDRE DEUX COPIES: l'une pour Geoffroy Peeters (exercices 1 à 3)
et l'autre pour Roland Badeau (exercices 4 et 5)

Les documents ne sont pas autorisés.

1 Cours "Cepstre"

Question 1 Expliquez ce qu'est le cepstre réel et son interprétation dans le cas d'un modèle source/filtre (on attend une réponse textuelle et avec des équations mathématiques).
Question 2 Que représentent les premiers coefficients du cepstre ? Que représente l'énergie aux basses céfrences ? aux hautes céfrences ?
Question 3 Comment peut-on utiliser le cepstre pour estimer la fréquence fondamentale f_0 d'un signal ?
Question 4 Quel est l'avantage de cette méthode d'estimation de f_0 par rapport à celle de l'auto-corrélation temporelle du signal ? Servez-vous de leurs expressions mathématiques dans le domaine fréquentiel.

2 Cours "Chromas"

Question 1 Expliquez ce que sont les Chromas.
Question 2 Pour le calcul des Chromas, la longueur de la fenêtre d'analyse est fonction de la fréquence $f_{min} \in \mathbb{R}^+$ de la note la plus basse considérée. Donner son expression mathématique en considérant une note MIDI notée $m_{min} \in \mathbb{N}^+$ (on attend une réponse textuelle et avec des équations mathématiques).
Question 3 Pourquoi dit-on que les Chromas sont sensibles au timbre des instruments de musique ?

3 Cours "Estimation multi-pitch"

Question 1 Qu'est-ce que la méthode du "produit spectral" pour estimer la fréquence fondamentale ? (on attend une réponse textuelle et avec des équations mathématiques)
Question 2 Quelles en sont ses principales hypothèses/limitations ?
Question 3 Expliquez le principe du lissage spectral ("spectral smoothing") utilisé dans la méthode d'estimation multi-pitch de [Klapuri, 2003, IEEE TASP].
Question 4 Comment le principe de "somme spectrale" est-il repris dans l'algorithme "Deep-Salience" de [Bittner et al., 2017, ISMIR] ?

4 Cours "NMF" : *regularized β-NMF*

In certain NMF applications, we are required to enforce sparsity constraints on the factor matrices W and H. This is often achieved by considering the following optimization problem:

$$(W^\star, H^\star) = \underset{W\geq 0, H\geq 0}{\arg\min} \Big[\sum_{f=1}^{F}\sum_{n=1}^{N} d_\beta(v_{fn}; \hat{v}_{fn}) + \lambda_W \sum_{f=1}^{F}\sum_{k=1}^{K} w_{fk} + \lambda_H \sum_{k=1}^{K}\sum_{n=1}^{N} h_{kn} \Big], \qquad (1)$$

where $\lambda_W > 0$, $\lambda_H > 0$, $\hat{v}_{fn} = \sum_{k=1}^{K} w_{fk} h_{kn}$ and $d_\beta(\cdot;\cdot)$ is the β-divergence that is defined as:

$$d_\beta(v;\hat{v}) = \frac{1}{\beta(\beta-1)} \Big(v^\beta + (\beta-1)\hat{v}^\beta - \beta v \hat{v}^{\beta-1} \Big). \qquad (2)$$

Question 1
What are the roles of $\lambda_W > 0$ and $\lambda_H > 0$ in this problem?
Question 2
Derive the multiplicative update rules for this particular problem.
Question 3
Let us assume that we obtain the optimal factors, $W^\star$ and $H^\star$. How can we use $W^\star$ and $H^\star$ for audio source separation? For music transcription?

5 Cours "Méthodes à haute résolution"

On rappelle que la méthode MUSIC consiste à diagonaliser la matrice de covariance $\boldsymbol{R}_{xx}$ du signal, et à déterminer les pôles $\{z_k\}_{k\in\{0\ldots K-1\}}$ en tant que solutions de l'équation

$$\|\boldsymbol{W}_\perp^H \boldsymbol{v}(z)\|^2 = 0 \qquad (3)$$

où $\boldsymbol{v}(z) = \big[1,\, z, \ldots,\, z^{n-1}\big]^T$, et la matrice $\boldsymbol{W}_\perp$, de dimension $n \times (n-K)$, contient les vecteurs propres de $\boldsymbol{R}_{xx}$ associés aux $n-K$ plus petites valeurs propres (engendrant ainsi l'espace bruit). Dans cet exercice, on suppose que tous les pôles du signal sont sur le cercle unité.
Question 1
Vérifier que l'équation (3) implique qu'ils sont également solutions de l'équation $P(z) = 0$, où

$$P(z) = z^{(n-1)}\, \boldsymbol{v}(1/z^*)^H \left(\boldsymbol{W}_\perp \boldsymbol{W}_\perp^H\right) \boldsymbol{v}(z).$$

Question 2
On définit la matrice $\boldsymbol{P} = \boldsymbol{W}_\perp \boldsymbol{W}_\perp^H$, et on remarquera que

$$P(z) = [z^{n-1},\, z^{n-2}, \ldots,\, z,\, 1]\, \boldsymbol{P} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-1} \end{bmatrix}.$$

Vérifier que la matrice $\boldsymbol{P}$ est à symétrie hermitienne et positive, et démontrer que $P(z)$ est un polynôme de degré au plus $2(n-1)$, dont les racines de module $\neq 1$ peuvent être regroupées par paires (si z est racine, $1/z^*$ l'est aussi).
Question 3
L'algorithme *root*-MUSIC consiste à calculer les racines de $P(z)$. D'après vous, comment pourrait-on en déduire les valeurs des pôles z_k ? (on prêtera attention au fait que le degré du polynôme P est supérieur à K).
Question 4
La méthode *spectral-MUSIC* vue en cours et testée en TP consiste à chercher les K maxima du pseudo-spectre $S(e^{i2\pi f}) = \frac{1}{\|\boldsymbol{W}_\perp^H \boldsymbol{v}(e^{i2\pi f})\|^2}$. Vérifier que les valeurs du pseudo-spectre pour les fréquences $f_k = \frac{k}{N_{\text{fft}}}$, où $k \in \{0, N_{\text{fft}} - 1\}$ et $N_{\text{fft}} \geq 2n-1$, peuvent être obtenues à l'aide de la TFD de longueur N_{fft} du signal constitué des coefficients du polynôme P, complétés par des zéros. Quel est l'inconvénient de cette approche par rapport à *root-MUSIC*?