Reduction and interpretation of matrices of frequency response functions by Bayesian independent component analysis

G. Brogna, J. Antoni, N. Totaro, L. Gagliardini, O. Sauvage

Version of Record: https://www.sciencedirect.com/science/article/pii/S0022460X1930330X
Manuscript_e97fafb1e07bdee46eaa071e691456c6

Reduction and Interpretation of Matrices of Frequency Response Functions by Bayesian Independent Component Analysis

G. Brogna[a,c,*], J. Antoni[a], N. Totaro[a], L. Gagliardini[b], O. Sauvage[c]

[a] *Univ Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, F-69621, Villeurbanne, FRANCE*
[b] *Groupe PSA, NVH Department, Route de Gisy, F-78140 Vélizy-Villacoublay, FRANCE*
[c] *Groupe PSA, Scientific and Future Technologies Department/StelLab, F-78140 Vélizy-Villacoublay, FRANCE*

Abstract

This work seeks an effective data reduction method for matrices of Frequency Response Functions (FRF) in a way that preserves, as much as possible, the physical interpretation of FRFs in the full targeted frequency range. Also, this reduction method is wished able to cope with the different sources of uncertainties linked to the definition of the mechanical system whose FRFs are processed. It is shown that a Bayesian formulation of Independent Component Analysis (ICA) serves this purpose. It is used here to decompose a FRF matrix as a sum of frequency independent matrices multiplied by a frequency dependent scalar component. On the one hand, the independence property of this processing allows the scalar component to be concentrated in a narrow frequency range, on the other hand the chosen Bayesian approach presents itself as the most natural way to take into account uncertainties in the input FRFs whether they are due to measurement errors or structural uncertainties. Moreover, the probabilistic framework is shown to provide credible intervals on the estimation of the decomposition factors, thus allowing some considerations on the reliability of the processing and the development of a straightforward thresholding method to enhance the data reduction. A first application on measured automotive vibro-acoustic transfer functions shows the reduction performance of the approach and its interest when trying to analyse the measurements. A second application on non-parametric random FRFs computed through a stochastic finite element model illustrates the capacity of the proposed approach to take into account the uncertainty of the FRFs data and to propagate it to the factors of the decomposition.

Keywords: Data reduction, ICA, Bayesian inference, FRFs

1. Introduction

It is well known that the response of any linear time-invariant system at a degree of freedom to a monochromatic unitary excitation can be simply described by a frequency response function (FRF), which is an intrinsic property of the system. By the virtue of the principle of superposition, the response of the system to a non-unitary excitation can then be obtained by multiplying the Fourier transform of the excitation signal by the FRF. Since in vibro-acoustic problems the linear time-invariant system assumption can often be made with a reasonable degree of approximation, this approach is extremely widespread.
Analysis of measured or computed matrices of FRFs is often performed as a way of assessing the dynamical behaviour of a system. In the industrial context, FRFs based methods are used even for complex and very large systems, i.e. with multiple degrees of freedom (DOFs). For instance, a common automotive finite element model may feature up to ten millions of DOFs and aimed studies may still yield matrices of FRFs concerning tens of pertinent excitation and response DOFs. This work addresses two main typical problems for large industrial FRF matrix applications:

- Dealing with large matrices makes the analysis and processing cumbersome. Unnecessary and meaningless information should be left out.

*Corresponding author
Email address: gianluigi.brogna@insa-lyon.fr (G. Brogna)

- Complex systems are characterized by an uncertainty of their dynamical properties due for instance to the production process. A model able to quickly provide samples representative of this uncertainty could be useful.

The first issue is here tackled by introducing a drastic reduction of the FRF matrix, under the constraint of physical interpretability on the whole frequency range. The second by performing the data reduction in an adapted probabilistic context.

1.1. The data-reduction problem

Literature covers the first problem quite well. Concerning FRFs matrices, the problem first arose in the context of Laser Doppler vibrometer measurements, in which it is current to deal with large data matrices. Several solutions can be found among the order reduction methods [1, 2]. Old but now largely used techniques comprehend the Padé approximations [3] and Chebyschev polynomials expansions. Bistritz et al [4] showed the relation between the two approaches for linear, time-invariant single-input/single-output systems. Other approaches rely on space-frequency separation: for instance Principal Component Analysis (PCA) [5] is a common way of performing data reduction and variable separation on measured signals, whereas Proper Generalized Decomposition [6] exploits a variable separation assumption to constrain the solution of partial differential equation problems, in order to make them more easily solved. While these methods perform well on the data reduction front, the former does not concern itself with the physical interpretation, whereas the latter needs an assumed model described through differential equations. Thus they are not satisfactory when trying to understand the dynamical behaviour of a system through its measured FRFs.

If understanding the mechanical system is of primary importance, then modal analysis should be preferred. It can itself be considered as a regression of measured FRFs on a modal model and, indeed, it is useful to interpret the behaviour of the system as linked to well-defined mode shapes. It also leads to a good reduction of the input data in the low frequency band, where few modes are enough to describe the system. However, its reduction capability decreases rapidly with the increase of the complexity of the system and of the frequency range. For instance, a car body presents more than 1700 structural modes between 0 and 350 Hz, making the modal approach powerful for numerical solving but useless with respect to a physically meaningful data reduction. Due to these limitations, other techniques of reduction have been proposed. Halvorsen et al. [7] and Dippery et al. [8] both based their reduction on a singular value decomposition of the FRF matrix. This approach - applicable only on matrices, i.e. multi-input/multi-output data - favours the reduction over the physical interpretation of the extracted components. Later, Arruda et al. [9] targeted single-input/multi-output data and proposed to perform data reduction in the spatial domain through two-dimensional Fourier series and in the frequency domain using Chebyschev polynomials. This approach is particularly adapted to the reduction of Laser Doppler vibrometer measurements of simple surfaces since the sinusoidal nature of the measured mobilities makes the Fourier series a good basis for the approximation of displacement fields.

From what just presented, it seems that a trade-off between data reduction efficiency and physical interpretability is unavoidable: good data reduction algorithms - such as PCA or singular value decomposition - do not yield physically meaningful components, whereas powerful physical models - as in modal analysis - become cumbersome when the frequency range shifts towards the high frequencies. In this paper, the authors wish to propose an approach that might overcome this trade-off.

The reduction of FRFs matrices is here tackled in the form of a decomposition based on Bayesian Independent Component Analysis (ICA). The FRF matrix itself is decomposed as a sum of frequency independent complex matrices multiplied by a frequency dependent complex scalar component. Thus, the proposed form is similar to the modal decomposition, where the frequency independent matrices would be the complex *mode shapes* and the frequency dependent scalars would be the *modal coordinates*. The great advantage is that, unlike modal decomposition, the proposed ICA based approach reveals to be more parsimonious, providing few spectrally concentrated scalar components up to frequency ranges where the number of modes would have been prohibitive. This yields effective data reduction and also a potentially greatly simplified analysis of the mechanical system.

ICA techniques generally aim at describing the input variables as a linear sum of some hidden unknown variables (less than or equal to the number of input variables). Strictly related to singular value decomposition and PCA, ICA does not enforce just the orthogonality (i.e. statistical decorrelation) of the extracted

components, but also their statistical independence. This approach is largely reminiscent of recent works in compressive sensing [10]. If one of the most used principles is still the PCA (or Karhunen-Loève Transform in signal processing) [11, 12], Saito et al. [13, 14, 15] pointed out the importance of sparsity and independence when choosing the basis: sparsity is particularly important for data compression whereas independence is more adapted to modelling. In most common and regular signals, the two constraints yield similar basis [13], meaning that enforcing the statistical independence ensure both data compression and a possible physical interpretation of the extracted components. On these premises, ICA, which was first developed for real data, has been extended to the complex domain [16, 17, 18, 19] and has already been used in mechanics, for instance to determine hidden independent acoustic sources [20] or to perform Operational Modal Analysis [21, 22]. A wide review of ICA approaches can be found in [23].

1.2. The uncertainty problem

One added difficulty when processing real systems FRFs is to take into account their uncertainty. For instance, as early as 1993, Kompella and Bernhard [24] measured FRFs on several nominally identical end-of-line vehicles. Just accounting for measurement errors, climate variations and production uncertainties leads to as much as 10 dB dispersion over [0 500] Hz for the vibro-acoustic FRFs and [0 1000] Hz for the acoustic ones. For the FRFs framework, it is proposed in this work to divide the uncertainties in "additive" and "multiplicative":

- Additive uncertainties are related to the additive measurement error and include the random measurement error and the bias error.
- Multiplicative uncertainties are related to the structural uncertainties. Among them, it is convenient to separate the manufacturing variability, intended here as the intrinsic variability necessary for marketing purposes. In the car case, for instance, it would be the variability due to manufacturing options among cars belonging to the same model. This allows to distinguish it from the production uncertainty, which is related only to the production process and can theoretically be reduced without any influence on the marketing options.

Thus, for instance, a population of FRF matrices obtained by performing the same measurement protocol on several cars (of the same model) would present measurement errors, production uncertainty and manufacturing variability. If the cars were to be nominally identical, then the manufacturing variability would be left out.

To address the existence of such uncertainties in FRF matrices, the proposed ICA is formulated in a Bayesian context. The main advantage resides in the estimation of the joint posterior probability density function of the model variables: Bayesian inference not only provides a point estimation of the model parameters, but from their full joint probability density function, any kind of statistical analysis (e.g. median, percentiles and credible intervals) can also be easily performed. However, this same full joint posterior distribution can rarely be computed analytically and sampling algorithms - e.g. Markov Chain Monte Carlo (MCMC) methods - are often needed. Although computationally heavy, MCMC algorithms are nowadays hardly prohibitive. A wide literature exists about the application of Bayesian techniques to various domains. In mechanics it has been used as a way of regularizing complex source identification problems both acoustical [25] and vibratory [26, 27] and it is now a widely accepted approach to perform model updating [28, 29] and structural health monitoring [30]. In this paper, it is instead used to "retro-propagate" the dispersion of the processed FRF database to the terms of the decomposition model. Consequently, any FRF sample issued from the identified decomposition model can be considered as a possible outcome of the process whose samples are in the input database. This also means that the type of uncertainties covered by the proposed approach is completely dependent on the database. For instance, if the FRFs in the database are measured on nominally identical vehicles, only the production uncertainty and measurement random error can be covered by the model. On the other hand, if the FRFs are measured on vehicles sharing the same model but different options, then the model will also cover the manufacturing variability.

Therefore, resorting to the Bayesian context to deal with the FRFs uncertainty yields two main results:

1. The identified independent components not only help understanding the dynamic behaviour of the system, but through their uncertainty, they also provide some information on how the production protocol influences this behaviour.

2. The proposed approach yields a surrogate model through which random samples of the FRFs can be easily generated, presenting itself as a way to build FRFs *metamodels* alternative to more common techniques already largely used in mechanics such as polynomial chaos expansion [31, 32], response surface methodology [33] or Gaussian Process emulators (*kriging*) [34, 35]. The fundamental advantage lies in the ease of physical interpretation enabled by the independent components. The main drawback is that the proposed approach is completely data driven and it does not allow the interpolation of FRFs for non-measured DOFs, as instead does a modal model or any *metamodel* spawning in the space domain.

For what concerns the Bayesian ICA algorithm itself, the main sources can be found in the blind source separation literature. In this framework, a Bayesian formulation for ICA has already been proposed by Knuth and Djafari [36, 37] and applied by Févotte for audio source separation [38]. This work extends Févotte's formulation to complex random variables and it modifies the sampling algorithm in order to take into account the uncertainty of the processed database when more samples of the same FRF matrix are available.

1.3. Content

The second section of this work defines the FRF matrix decomposition and the working variables. The Bayesian ICA is introduced along with some basics of Bayesian inference. Then, the actual algorithm is provided. Some applications of the algorithm are presented in the third section. First of all, it is used on a measured automotive vibro-acoustic FRF matrix: the data reduction capacity is pointed out, along with physical interpretations linked to the independent nature of the extracted components. Then, the algorithm is applied to a set of several samples of the "same" automotive vibro-acoustic FRF matrix. Due to the lack of measured data, for this application the FRFs are computed through a non-parametric stochastic finite element model [39, 40]. It is shown that the FRFs uncertainty is easily propagated to the independent components and it can thus be taken into account when interpreting the results from a mechanical point of view. Finally, some perspectives and considerations on the approach are provided in the conclusions.
Even though the present application concerns automotive vibro-acoustic FRF matrices, the proposed algorithm is indicated to perform data reduction and ease the interpretation of any complex dynamical system.

2. Definition of the problem

2.1. FRF matrix decomposition and definition of the working variables

Let Ω be a domain in which the equation of motion is verified and a unique solution exists which linearly links the accelerations (or acoustic pressures) and the loads, in a frequency band $[\omega_l, \omega_u]$. In this domain, excitations on some degrees of freedom yield to unique responses at the observation degrees of freedom (DOFs), as:

$$\forall \omega \in [\omega_l, \omega_u], \quad \mathbf{p}(\omega) = \mathbf{H}(\omega)\mathbf{f}(\omega) \tag{1}$$

where $\mathbf{p}(\omega) \in \mathbb{C}^{N_o \times 1}$ is the column vector stacking the spectra of the responses at N_o observation DOFs, $\mathbf{f}(\omega) \in \mathbb{C}^{N_e \times 1}$ is the column vector stacking the spectra of the excitations at N_e DOFs and $\mathbf{H}(\omega) \in \mathbb{C}^{N_o \times N_e}$ is the matrix of the linear, time-invariant frequency response functions between the excitation DOFs and the observation ones.
The objective of this work is to reduce the FRF matrix $\mathbf{H}(\omega)$. To this aim, it is proposed to decompose it as:

$$\forall \omega \in [\omega_l, \omega_u], \quad \mathbf{H}(\omega) = \sum_{p=1}^{N_p} \mathbf{S}_p c_p(\omega) + \mathbf{R}(\omega) \tag{2}$$

where N_p is the number of retained components (its determination is explained afterwards), the matrices $\mathbf{S}_p \in \mathbb{C}^{N_o \times N_e}$ are frequency independent and contain magnitude and phase relations between the excitation and observation DOFs, while the spectral dependence is expressed for each pattern through the complex scalar $c_p(\omega)$ - to be interpreted as a filter - whereas $\mathbf{R}(\omega) \in \mathbb{C}^{N_o \times N_e}$ is a residue.

The decomposition in Eq. (2) is not unique: any basis can be chosen to yield the same formulation (with different performances). Several techniques exist to impose constraints and solve the indeterminacy. One of the most widely used is the Karhunen-Loève decomposition, whose discrete form is the PCA [41, 42, 43].

The use of this approach needs for the input FRF matrix $\mathbf{H}(\omega)$ to be considered as a stochastic matrix with random trajectory along the frequency, meaning that the matrices at every fixed frequency are samples of the same stochastic process. It is thus decomposed as in Eq. (2) with the constraint of decorrelation among the stochastic processes $c_p(\omega)$, which are then sorted by decreasing variance and called principal components. The number of principal components is less than or equal to the number of variables in the matrix, thus performing data reduction.

ICA needs exactly the same framework, however, unlike the Karhunen-Loève decomposition, it does not impose the decorrelation but rather the independence of the stochastic processes $c_p(\omega)$, which are then called independent components. For a much larger dissertation on ICA the reader may refer to Hyvärinen's book [23], but for the scope of this paper it is enough to know that the independence constraint, much stronger than decorrelation, can be rigorously formulated with respect to the separability of the joint probability density function (PDF) of the components. The components $c_p(\omega)$ are independent if and only if:

$$\forall \omega \in [\omega_l, \omega_u], \quad [c_1(\omega), \ldots, c_{N_p}(\omega)] = \prod_{p=1}^{N_p} [c_p(\omega)] \tag{3}$$

where $[c_p(\omega)]$ stands for the PDF of the component $c_p(\omega)$ and $[c_1(\omega), \ldots, c_{N_p}(\omega)]$ for the joint PDF of all components.

It can be argued that given a system, the spectral evolution of the transfer functions $\mathbf{H}(\omega)$ is not random but deterministic, so that the application of the probabilistic constraint above might be out of its scope. In this work, a small paradigm shift is needed: the frequency evolution of the FRFs for one specific system is indeed deterministic, but considering it a sample of a random process - i.e. the $c_p(\omega)$ are seen as trajectories of a multivariate random process - ICA can be applied yielding a basis that presents both sparsity (i.e. spectrally localized components) and thus a remarkable physical interpretation.

In practice, to estimate the elements in Eq. (2) from finite length measurements, discrete matrix forms are here preferred. So, if the frequency band $[\omega_l, \omega_u]$ is discretised in N_f frequency points, the three-dimensional FRF tensor can be defined as $\underline{\mathbf{H}} \in \mathbb{C}^{N_o \times N_f \times N_e}$. It is then mapped into the matrix $\overline{\mathbf{Y}} \in \mathbb{C}^{N_e \cdot N_o \times N_f}$:

$$\overline{\mathbf{Y}} = \begin{pmatrix} \underline{\mathbf{H}}_{:,:,1} \\ \underline{\mathbf{H}}_{:,:,2} \\ \vdots \\ \underline{\mathbf{H}}_{:,:,N_e} \end{pmatrix} \tag{4}$$

The following decomposition can then be written:

$$\overline{\mathbf{Y}} = \overline{\mathbf{\Lambda}}\,\overline{\mathbf{C}} + \overline{\mathbf{E}}. \tag{5}$$

Equations (2) and (5) are equivalent, being $\overline{\mathbf{\Lambda}} \in \mathbb{C}^{N_e \cdot N_o \times N_p}$ a reshaping and stacking of the matrices $\mathbf{S}_p$, $\overline{\mathbf{C}} \in \mathbb{C}^{N_p \times N_f}$ a stacking of the components $c_{p,n}$ for $p \in [1, \ldots, N_p]$ and $n \in [1, \ldots, N_f]$ and $\overline{\mathbf{E}} \in \mathbb{C}^{N_e \cdot N_o \times N_f}$ the residual error. In this discrete formulation, while PCA bases itself on the diagonalization of the empirical covariance of $\overline{\mathbf{Y}}$, ICA has to seek other characteristics of the signals. For instance, among the tensorial ICA techniques, the JADE algorithm [44] uses the fourth order statistical moment of $\overline{\mathbf{Y}}$ as an image of the independence of the sources.

Hereafter, a Bayesian approach to the identification of $\overline{\mathbf{\Lambda}}$ and $\overline{\mathbf{C}}$ is preferred, in order to easily take into account the structural dispersion of the input FRFs.

2.2. Basic principles of Bayesian inference

Bayesian sampling algorithms derive from the Bayes' rule [45]. Being $\boldsymbol{\theta}$ the vector storing the unknown random variables and X the measurements, the Bayes' rule reads as:

$$[\boldsymbol{\theta}|X] \propto [X|\boldsymbol{\theta}] \cdot [\boldsymbol{\theta}] \tag{6}$$

where:

- $[\boldsymbol{\theta}|X]$ is the joint posterior probability density function (PDF) of the parameters, i.e. the distribution of the unknown parameters given the measurements X.

- $[X|\boldsymbol{\theta}]$ is the likelihood function, which expresses the probability of observing the measurements X given the parameters $\boldsymbol{\theta}$. This notation is adopted for the sake of simplicity, but note that in the likelihood function the variables are the parameters $\boldsymbol{\theta}$ and X is fixed to the measured values.
- $[\boldsymbol{\theta}]$ is the prior PDF of the parameters, which expresses what is known about the parameters before any measurement.

The aim of the inference is to obtain the joint posterior PDF. Its direct computation is usually impossible, so that more sophisticated sampling algorithms are necessary (e.g. acceptance-rejection sampling [46], Metropolis-Hastings sampling [47, 48]). In this work, Gibbs sampling [49] is used. Gibbs sampling, which belongs to the large family of MCMC methods, instead of performing the inference estimating the joint posterior PDF of all the parameters, uses the simpler conditional posterior PDF of one parameter given all the others. In particular, going through the parameters and drawing each from its conditional posterior PDF given the other parameters at their most recent values, the algorithm generates a Markov chain whose long-run distribution is the searched joint posterior PDF of all the parameters. It is particularly indicated for hierarchical models [50] and it has also been used by Knuth [36] and Djafari [37] to formulate ICA in a Bayesian context.

2.3. Bayesian formulation of ICA

The Bayesian approach to ICA consists in solving the probabilistic hierarchical problem corresponding to Eq. (5) and presented in figure 1, which reads:

$$\mathbf{Y} = \boldsymbol{\Lambda}\mathbf{C} + \mathbf{E}. \tag{7}$$

The terms in Eq. (7) are all random variables and in particular $\mathbf{Y}$ is the random variable whose measured sample is $\overline{\mathbf{Y}}$. The aim is to infer the joint posterior PDF of the random variables, i.e. their joint PDF given the measured sample $\overline{\mathbf{Y}}$ and some prior knowledge.
First of all, the likelihood function of the measurements $\overline{\mathbf{Y}}$ is defined. Classically, the noise $\mathbf{E}$ follows a circularly symmetric complex Gaussian distribution[1] (noted $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$) and it is considered homoskedastic, i.e. it has zero mean and the same variance for each measured variable. Under these assumptions, knowing the model $(\boldsymbol{\Lambda}, \mathbf{C})$ and the variance of the noise σ_y^2, the likelihood function can be written as:

$$[\overline{\mathbf{y}}_v|\boldsymbol{\Lambda}, \mathbf{C}, \sigma_y^2] \sim \mathcal{N}_{\mathbb{C}}(\hat{\mathbf{y}}_v, \sigma_y^2 \cdot \mathbb{I}_{N_e \cdot N_o \cdot N_f}) \tag{8}$$

with $\overline{\mathbf{y}}_v = \mathrm{vec}(\overline{\mathbf{Y}})$ and $\hat{\mathbf{y}}_v = \mathrm{vec}(\boldsymbol{\Lambda}\mathbf{C})$, where the operator $\mathrm{vec}(\bullet)$ stands for the vectorization of a matrix. As pointed out at the beginning of the paragraph, in the likelihood function (8) the random variables (and unknowns) are $\hat{\mathbf{y}}_v$ and σ_y^2, whereas $\overline{\mathbf{y}}_v$ is the known measurement (even if the Gaussian distribution is written with respect to $\overline{\mathbf{y}}_v$). Each other unknown parameter is also a random variable and as such it is supposed to follow a prior PDF. Prior PDFs are chosen directly by the analyst, who can use them as a way to convey his knowledge on the objects of inference prior to any measurement. In practice, if the analyst does not wish to convey any knowledge, uninformative prior distributions can be used (e.g. uniform distributions or Jeffreys priors [51]). Otherwise, *conjugate* prior distributions[2] are often used, since they greatly simplify Bayesian posterior computations. This restriction to *conjugate* distributions is not compulsory, although it is a compromise often made in order to flexibly convey prior information while still keeping the computational effort acceptable.
In the present application, ICA forces the prior PDFs of the components $\mathbf{C}$ to be non-Gaussian [36, 37]: gaussianity would lead to PCA instead of ICA [23]. Knowing this, the following prior distributions are chosen for the random variables:

[1] Since the noise can be considered due to a large number of uncontrolled phenomena, the central limit theorem is used to justify the Gaussian assumption.

[2] Given the likelihood $[X|\boldsymbol{\theta}]$, a prior distribution $[\boldsymbol{\theta}]$ is called *conjugate* if $[\boldsymbol{\theta}]$ and the posterior distribution $[\boldsymbol{\theta}|X] \propto [X|\boldsymbol{\theta}] \cdot [\boldsymbol{\theta}]$ belong to the same family of distributions.

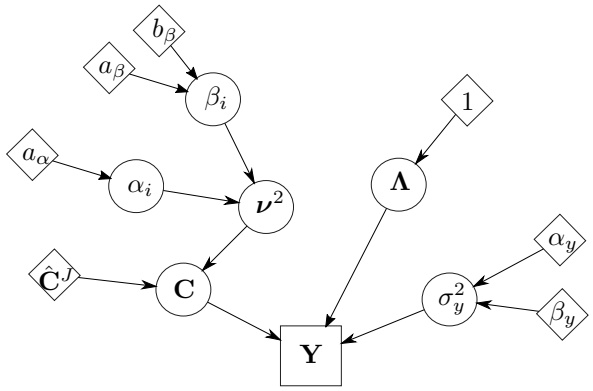


Figure 1: Bayesian hierarchical model for ICA. Diamonds stand for hyper-parameters chosen by the user; rounds for the random variables whose joint posterior PDF has to be inferred; the square stands for the random variable whose measurement is available.

- The N_p components are modelled a priori as independent identically distributed sequences of N_f complex values. They are also considered mutually independent, so that the whole prior probability density function factorizes as:

$$[\mathbf{C}|\boldsymbol{\theta}_\mathbf{C}] = \prod_{i=1}^{N_p} \prod_{n=1}^{N_f} [c_{i,n}|\theta_{c_i}] \tag{9}$$

where $\boldsymbol{\theta}_\mathbf{C}$ stands for the set of parameters defining the prior distribution chosen for $\mathbf{C}$. As it has been said, the prior distribution for the components must be non-Gaussian if ICA is wished. A Student t prior is chosen in this work. This choice is purely practical, since Student t distributions can be easily implemented in a Gibbs sampler through the Scaled Mixture of Gaussians (SMOG) approach [52]. Following this approach, the Student t distributions can be written as:

$$\begin{aligned} \forall i = 1, \ldots, N_p, \forall n = 1, \ldots, N_f, [c_{i,n}] &\sim \mathcal{N}_\mathbb{C}(\hat{c}_{i,n}^J, \nu_{i,n}^2) \\ [\nu_{i,n}^2] &\sim \text{InvGamma}(\alpha_i, \beta_i) \end{aligned} \tag{10}$$

where the mean of the prior distribution $\hat{c}_{i,n}^J$ is an a priori coefficient as returned for instance by a non-probabilistic ICA algorithm, whereas the variance $\nu_{i,n}^2$ changes for each $n \in [1, \ldots, N_f]$ following an Inverse Gamma distribution (noted InvGamma). Since in the SMOG approach the prior knowledge is updated by one single observation, having a good prior estimation (here by JADE estimation) makes the convergence faster and more stable.

- Low values of $\nu_{i,n}^2$ may be good for the stability of the algorithm since they give more weight to the prior components $\hat{c}_{i,n}^J$. They can be favoured using an exponential prior for α_i:

$$\forall i = 1, \ldots, N_p, \quad [\alpha_i] = a_\alpha e^{-a_\alpha \alpha_i} \tag{11}$$

- For the same reason, a Gamma distribution is best chosen for β_i:

$$\forall i = 1, \ldots, N_p, \quad [\beta_i] \sim \text{Gamma}(a_\beta, b_\beta). \tag{12}$$

Moreover, a Gamma distribution for $[\beta_i]$ is *conjugate* with respect to the Inverse Gamma distribution chosen for $[\nu_{i,n}^2]$, which has the role of the likelihood function when inferring $[\beta_i]$ in the hierarchical model (Fig. 1).

- Unlike $\mathbf{C}$, $\boldsymbol{\Lambda}$ is inferred from all N_f samples. As a consequence, enough information is available to perform a stable inference of the posterior distribution directly from the measurements and a non-informative prior can be used:

$$[\mathbf{l}_v] \propto 1 \tag{13}$$

with $\mathbf{l}_v = \text{vec}(\boldsymbol{\Lambda})$.

- The variance σ_y^2 of the error can be supposed following an Inverse Gamma law as:

$$[\sigma_y^2] \sim \text{InvGamma}(\alpha_y, \beta_y) \tag{14}$$

An Inverse Gamma prior is used here, since it is *conjugate* with respect to the Gaussian likelihood function in Eq. (8), yielding to simple posterior computations.

The parameters which have not been assigned a prior PDF in the list above are the hyper-parameters of the model. They are chosen directly by the user and reflect his/her knowledge. For instance, the mean values $\hat{c}_{i,n}^J$ in Eq. (10) are hyper-parameters and reflect the knowledge of the user, in this case obtained applying before-hand an ICA algorithm to the analysed data. On the other hand, the non-informative prior in Eq. (13) reflects the desire of the user to leave some freedom to the convergence of the Gibbs sampler (one could have used the result of JADE as prior as well). Likewise, the hyper-parameters α_y and β_y in Eq. (14) can be chosen to have a wide prior distribution in order to mirror the lack of knowledge of the analyst.
The choice of which parameters to put as hyper-parameters or as random variables may be delicate: putting the hyper-parameters high in the hierarchical model will make the computation heavier and less stable, but placing them at a low level may over-constrain the calculation. In general, the more data are available, the higher the hyper-parameters can be set [50] and still ensure a fast convergence.

Knowing the likelihood and the prior distributions, the conditional posterior distribution of each parameter can be computed using Bayes theorem. Then, a Gibbs sampler allows the posterior marginal distribution of each parameter to be obtained. The algorithm is presented hereafter. One already familiar with the Bayesian formulation of ICA as delivered by Fevotte [38] may notice that the main difference lays in the extension to complex random variables. Other modifications are instead needed to take into account several samples of the processed FRF matrix and they are presented in subsection 2.5.

2.4. Bayesian hierarchical ICA: algorithm

In what follows $\bullet^H$ stands for the Hermitian transpose, $\bullet^t$ stands for the transpose, $\bullet^*$ stands for the complex conjugate and "rest" stands for all the random variables except the one whose PDF is inferred. The definition of the used PDF and the posterior computations are carried out respectively in Appendices A and B. The Gibbs sampling algorithm applied to the ICA is composed of 9 steps, steps 5 and 5 bis are alternative and are discussed later on:

1. Initialize the values of $\mathbf{\Lambda}$, $\mathbf{C}$, $\forall i = 1, \ldots, N_p$ and $\forall n = 1, \ldots, N_f$, $\quad \nu_{i,n}^2$ and $\forall i = 1, \ldots, N_p$, $\quad \beta_i$ and α_i.

2. Draw a sample from $[\sigma_y^2|\text{rest}] \sim \text{InvGamma}(\alpha'_y, \beta'_y)$ with

$$\begin{aligned} \alpha'_y &= \alpha_y + N_e \cdot N_o \cdot N_f \\ \beta'_y &= \beta_y + \sum_{j=1}^{N_e \cdot N_o} \sum_{n=1}^{N_f} |\overline{y}_{j,n} - \mathbf{l}_j \mathbf{c}_n|^2 \end{aligned} \tag{15}$$

where $\overline{y}_{j,n}$ stands for the (j,n) value of the matrix $\overline{\mathbf{Y}}$, $\mathbf{l}_j \in \mathbb{C}^{1 \times N_p}$ is the row vector corresponding to the j^{th} row of $\mathbf{\Lambda}$ and $\mathbf{c}_n \in \mathbb{C}^{N_p \times 1}$ is the column vector corresponding to the n^{th} column of $\mathbf{C}$.

3. $\forall j = 1, \ldots, N_e \cdot N_o$, draw a sample from $[\mathbf{l}_j^t|\text{rest}] \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_{l_j}, \mathbf{\Sigma}_{l_j})$ with

$$\begin{aligned} \boldsymbol{\mu}_{l_j} &= \frac{1}{\sigma_y^2} \mathbf{\Sigma}_{l_j} \mathbf{C}^* \overline{\mathbf{y}}_j^t \\ \mathbf{\Sigma}_{l_j} &= \sigma_y^2 ((\mathbf{C}\mathbf{C}^H)^*)^{-1} \end{aligned} \tag{16}$$

where $\overline{\mathbf{y}}_j \in \mathbb{C}^{1 \times N_f}$ is the j^{th} row of matrix $\overline{\mathbf{Y}}$.

4. Normalize the columns of $\mathbf{\Lambda}$ to 1 to solve the indeterminacy on gain.

5. $\forall n = 1, \dots, N_f$, draw a sample of a column $\mathbf{c}_n$ out of $\mathbf{C}$ from $[\mathbf{c}_n|\text{rest}] \sim \mathcal{NC}(\boldsymbol{\mu}_{c_n}, \boldsymbol{\Sigma}_{c_n})$ with

$$\begin{aligned} \boldsymbol{\mu}_{c_n} &= \boldsymbol{\Sigma}_{c_n} \left(\frac{1}{\sigma_y^2} \boldsymbol{\Lambda}^H \overline{\mathbf{y}}_n + \text{diag}\left(\frac{1}{\boldsymbol{\nu}_n^2} \right) \hat{\mathbf{c}}_n^J \right) \\ \boldsymbol{\Sigma}_{c_n} &= \left(\frac{1}{\sigma_y^2} \boldsymbol{\Lambda}^H \boldsymbol{\Lambda} + \text{diag}\left(\frac{1}{\boldsymbol{\nu}_n^2} \right) \right)^{-1} \end{aligned} \tag{17}$$

where $\boldsymbol{\nu}_n^2 \in \mathbb{C}^{N_p \times 1}$ contains the N_p variances for the n^{th} observed sample and $\overline{\mathbf{y}}_n \in \mathbb{C}^{N_e \cdot N_o \times 1}$ is the column vector containing the $N_e \cdot N_o$ observations of the n^{th} measured sample.
This step is the "block" approach to components estimation.

5 bis. $\forall i = 1, \dots, N_p$ and $\forall n = 1, \dots, N_f$, draw a sample from $[c_{i,n}|\text{rest}] \sim \mathcal{NC}(\mu_{c_{i,n}}, \sigma^2_{c_{i,n}})$ with

$$\begin{aligned} \mu_{c_{i,n}} &= \sigma^2_{c_{i,n}} \left(\frac{\overline{y}_{i|-i,n}}{\sigma_i^2} + \frac{\hat{c}_{i,n}^J}{\nu_{i,n}^2} \right) \\ \sigma^2_{c_{i,n}} &= \left(\frac{1}{\sigma_i^2} + \frac{1}{\nu_{i,n}^2} \right)^{-1} \\ \sigma_i^2 &= \frac{\sigma_y^2}{\|\boldsymbol{\lambda}_i\|_2^2} \\ \overline{y}_{i|-i,n} &= \frac{\boldsymbol{\lambda}_i^H \overline{\mathbf{y}}_n}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i} - \sum_{k \neq i} \frac{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_k}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i} c_{k,n} \end{aligned} \tag{18}$$

where $\boldsymbol{\lambda}_i \in \mathbb{C}^{N_e \cdot N_o \times 1}$ is the i^{th} column of the matrix $\boldsymbol{\Lambda}$.
This step contains the "one-by-one" approach to component estimation [38].

6. $\forall i = 1, \dots, N_p$ and $\forall n = 1, \dots, N_f$, draw a sample from $[\nu^2_{i,n}|\text{rest}] \sim \text{InvGamma}(\alpha'_{i,n}, \beta'_{i,n})$ with

$$\begin{aligned} \alpha'_{i,n} &= \alpha_i + 1 \\ \beta'_{i,n} &= \beta_i + |c_{i,n} - \hat{c}^J_{i,n}|^2 \end{aligned} \tag{19}$$

7. $\forall i = 1, \dots, N_p$, draw a sample from

$$[\alpha_i|\text{rest}] \propto \exp\left(-N_f \log\Gamma(\alpha_i) + \left(\sum_{n=1}^{N_f} \log \frac{\beta_i}{\nu^2_{i,n}} - a_\alpha \right) \alpha_i \right) \tag{20}$$

This distribution is obtained by Bayes theorem as the product of an Inverse Gamma distribution with an exponential distribution and it is not easy to sample. A Metropolis-Hastings sampler might be needed. In practice, since α_i is high in the hierarchy (Fig. 1), deterministic moves are found to not compromise the convergence of the Markov Chain [38]. As a consequence, the mode of the distribution is chosen as sample at each iteration:

$$\begin{aligned} \alpha_i &= \underset{\alpha_i}{\text{argmax}} \left(-N_f \log\Gamma(\alpha_i) + \left(\sum_{n=1}^{N_f} \log \frac{\beta_i}{\nu^2_{i,n}} - a_\alpha \right) \alpha_i \right) \\ &= \psi^{-1}\left(\frac{1}{N_f} \left(\sum_{n=1}^{N_f} \log \frac{\beta_i}{\nu^2_{i,n}} - a_\alpha \right) \right) \end{aligned} \tag{21}$$

where $\psi^{-1}(\bullet)$ is the *inverse Digamma function.*

8. $\forall i = 1, \ldots, N_p$, draw a sample from $[\beta_i|\text{rest}] \sim \text{Gamma}(a'_\beta, b'_\beta)$ with

$$\begin{aligned} a'_\beta &= a_\beta + N \cdot \alpha_i \\ b'_\beta &= b_\beta + \sum_{n=1}^{N_f} \frac{1}{\nu_{i,n}^2} \end{aligned} \tag{22}$$

9. Go back to step 2 and iterate the process until a sufficiently large sample of parameters is collected after the convergence of the Markov chain.

In the Gibbs sampling, the most recent sampled values of the random variables are used to compute the parameters of the conditional posterior distributions at each step. Iterating the algorithm generates a Markov Chain whose long-run distribution is the joint posterior distribution of all the random variables. However, before a Markov Chain state can be considered from the long-run distribution, the chain has to "heat". These convergence iterations are called *burn-in* and they must not be taken into account when inferring the joint posterior distribution of the random variables.

Remarks on step 4. Without step 4, the algorithm would not converge towards one single solution, due to the indeterminacy on the gain (i.e. multiplying $\mathbf{\Lambda}$ and dividing $\mathbf{C}$ by the same number leads to the same result). This was already mentioned in [38], where it is added that a rigorous implementation would imply sampling each column of $\mathbf{\Lambda}$ from the set of norm 1 random vectors. Nevertheless, this approximation is satisfactory in practice. Moreover, since we consider complex variables the indeterminacy does not only concern the gain, but also the phase. Solving this indeterminacy can be achieved by shifting the phase of each column of $\mathbf{\Lambda}$ so that the phase of the first element is null. However if the reference phase is characterized by a large dispersion, imposing it as null might spread the dispersion to the other elements. For this reason, in the presented algorithm the phase is left free with only the influence of the prior law to limit its variation.

Remarks on step 5. It should be noted that steps 5 and 5 bis lead to the same result but they do not imply the same computational effort. The block strategy (step 5) is characterized by a joint sampling of the N_p sources for each $n \in \{1, \ldots, N_f\}$. Using a joint distribution constrains the sampling and it converges to the solution in less iterations than a full conditional approach. The drawback is that, due to the multi-dimensional nature of the problem, the block strategy consists of N_f problems which need the inversion of a $N_p \times N_p$ matrix. On the other hand, the one-by-one approach is a fully conditional approach: each source is estimated conditionally to all other sources and parameters. Such an approach can be easily vectorized and it does not imply any matrix inversion, but it generally takes more iterations to converge. Depending on the application, the one or the other method may be preferred: in the case of this work the one-by-one approach is applied.

2.5. Including structural input dispersion in the algorithm

As stated in the introduction, when dealing with real system FRFs, uncertainty is inevitable and, since the industrial mindset is now more than ever turned toward simulation, it is becoming central. For instance, usually the finite element model used to compute and assess the car performance is unique, although it should be representative of the production uncertainty and manufacturing variability (e.g. different trims, engines, gearboxes...). This issue has already been tackled experimentally by Hills *et al.* [53], who provided general statistical properties of acoustic FRFs by measuring several hundreds vehicles, and numerically by Durand *et al.* [39] who proposed to simulate the dispersion of the transfer functions using the non-parametric stochastic finite element method developed by Soize [40]. This latter approach, now implemented in commercial software [54], allows the computation of large numbers of samples of possible FRFs using a standard finite element model and only six dispersion hyper-parameters.

It is then evident that, more than just having one single FRF matrix sample, the tendency implies measuring (or simulating) several FRF matrix samples, in order to take into account the measurement error and structural uncertainty. Thus, in this paper the aim is to propose an algorithm capable of processing not only one FRF matrix, but a whole population of uncertain FRFs and providing a unified decomposition. The way the FRF matrix samples are obtained is not in the scope of this paper and is of no concern for the definition of the algorithm. They can come from end-of-line measurements as well as be computed with a stochastic

finite element method. The issue here is to be able to propagate the uncertainty represented by the processed samples to the terms of the decomposition model. Provided this aim is achieved, the resulting decomposition approach would still yield data reduction and high physical interpretability even when processing a whole FRF matrix population. Moreover, it would also allow the generation of FRF matrix samples representative of the measured (or simulated) population. Thus, the decomposition model could be used as a reduced surrogate model of the processed FRF matrix. The distribution of its samples would be representative of the way the processed FRF matrices have been obtained: for instance production uncertainty, manufacturing variability and measurement random error[3] when dealing with measured FRFs.

To do so and include a whole population of FRF matrices in the Bayesian ICA, two slight changes to the previous algorithm are necessary.

First of all, the measurement $\overline{\mathbf{Y}}$ is changed at each Gibbs sampling iteration. More specifically, at each iteration a measurement $\overline{\mathbf{Y}}^{(r)}$ is sampled uniformly out of the FRFs population and used in the Gibbs algorithm. After a high enough number of iterations, the whole measurement space will have been visited and the posterior distribution of the components will reflect the FRF population uncertainty.

Secondly, another thoughtful - but not compulsory - modification concerns the inference of the additive noise $\mathbf{E}$. It has been recognised that the structural uncertainty implies an uncertianty on FRFs that typically increases as the frequency increases [39]. In the Bayesian ICA model it is thus interesting to leave the variance of the noise free to change from one frequency to another to try and detect this increasing uncertainty. Under this hypothesis, the likelihood function in Eq. (8) becomes:

$$[\overline{\mathbf{y}}_v|\mathbf{\Lambda}, \mathbf{C}, \sigma^2_{y_1}, \ldots, \sigma^2_{y_{N_f}}] \sim \mathcal{N}_{\mathbb{C}} \left(\hat{\mathbf{y}}_v, \begin{bmatrix} \sigma^2_{y_1} \cdot \mathbb{I}_{N_e \cdot N_o} & & \\ & \ddots & \\ & & \sigma^2_{y_{N_f}} \cdot \mathbb{I}_{N_e \cdot N_o} \end{bmatrix} \right) \tag{23}$$

This implies some modifications on the posterior PDFs of the random variables, which are pointed out in Appendix C.

3. Application of the algorithm

3.1. An experimental application

The first application of the proposed algorithm concerns an FRF matrix measured on a production vehicle. The observation DOFs are $N_o = 4$ acoustic pressures at the ears of the driver and passengers, while the excitation DOFs are $N_e = 29$ loads and torques at the engine mounts and wheel centres. The FRF matrix has been constructed by using reciprocity [55] and $N_f = 451$ frequency points have been recorded in the [25 250] Hz frequency band with a resolution of 0.5 Hz. The measurement has taken place in a semi-anechoic room in which the whole vehicle has been placed. Inside the vehicle cabin, four acoustic sources excited the car in the targeted frequency range with a white noise. The response of the car has been recorded through triaxial accelerometers. Twenty samples of $H1$-type estimators of FRFs have been averaged to get a final estimation. The data are available and can be retrieved in the web repository [56]. For instance, Fig. 2 shows three of the most important FRFs, from the engine mounts towards the driver's ears.

Before applying the algorithm, one has to choose the number of extracted components N_p. The most straightforward way is to compute the eigenvalues d_l for $l \in [1, \ldots, N_e \cdot N_o]$ of the covariance matrix of $\overline{\mathbf{Y}}$. The relative cumulative sum of eigenvalues can then be written as:

$$g_j = \frac{\sum_{l=1}^{j} d_l}{\sum_{q=1}^{N_e \cdot N_o} d_q} \tag{24}$$

A threshold on g_j can be a good way to choose the number of components. Figure 3 shows that in this application $N_p = 19$ is large enough to describe more than the 97.5% of the information.

Given the number of components to be extracted, Bayesian ICA can be applied. Figure 4 compares the

[3]Note that the bias measurement error being present in every processed FRF matrix, it cannot be isolated in the error term in Eq. (7) and thus it inevitably introduces a bias in the decomposition terms too.

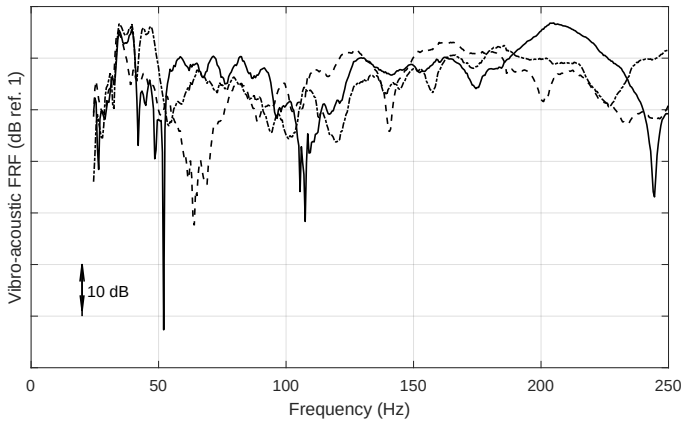


Figure 2: Example of FRFs measured experimentally: from the right engine mount along the Z axis (solid line), the left engine mount along the Z axis (dot-dashed line) and the bottom engine mount along the X axis (dashed line) toward the driver's ear.

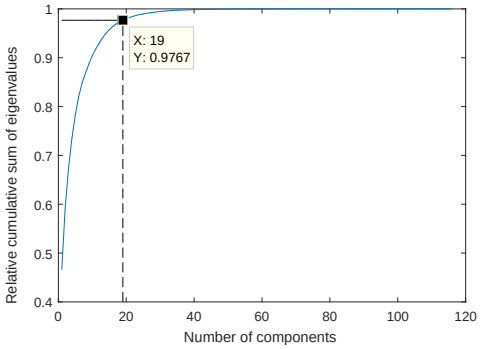


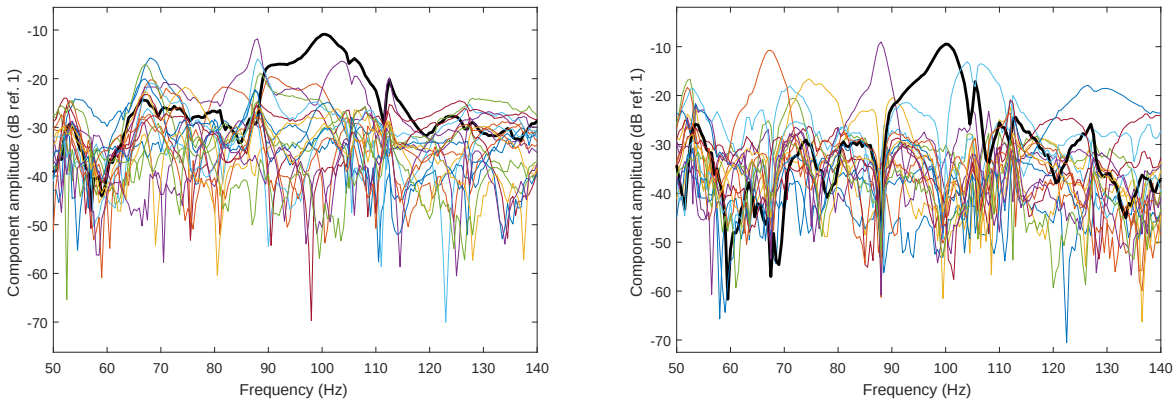
Figure 3: Relative cumulative sum of eigenvalues of $\overline{\mathbf{Y}}$. The number of components is chosen with a threshold of 97.5%.

components $\mathbf{C}$ obtained through PCA and ICA. As it can be seen, the ICA components are localized on narrow frequency bands, while the PCA ones are more spread. The localization feature is what justifies the use of ICA rather than PCA: localized components are easier to interpret and more useful, since generally for each frequency band just one independent component is dominant.

Going further in the exploitation, this means that the whole FRF matrix can be reduced to a simple pattern $\mathbf{S}_p$ when looking to a specific frequency range: the patterns offer an easy way of condensing the system behaviour. For instance, figure 5 shows the population obtained by Gibbs sampling for the dominant component around 100 Hz and figure 6 the mean of its corresponding spatial pattern. It can be seen that the pattern exhibits the dominant DOFs and phase relationships around 100 Hz. In the industrial context, this can be invaluable information, since it may allow understanding the origin of problems occurring in a specific frequency range, provided that the above mentioned patterns are analysed in an "automotive" way.

3.1.1. Enhancing the data reduction

This localized spectral behaviour calls for further data reduction. One can imagine to reduce the spectral support of an independent component in order to only keep the information around the peak. The Bayesian approach provides a most natural way of setting the threshold of this second stage reduction, by means of



(a) (b)

Figure 4: Components $c_p(\omega)$ for $p = 1, \ldots, 19$ obtained by PCA (a) and Bayesian hierarchical ICA (b). Highlighted the component dominating around 100 Hz, also analysed in Fig. 5.

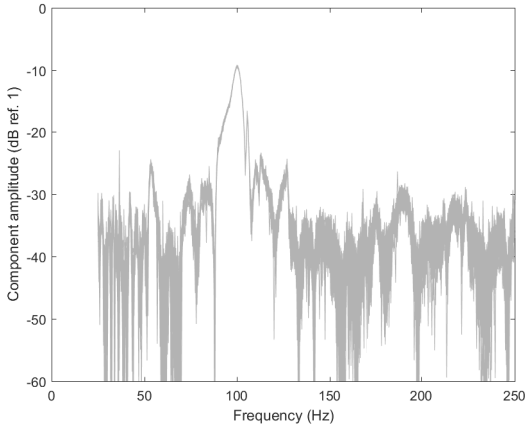


Figure 5: The independent component $c_{12}(\omega)$ concentrated around 100 Hz: its population obtained by Gibbs sampling (3000 samples).

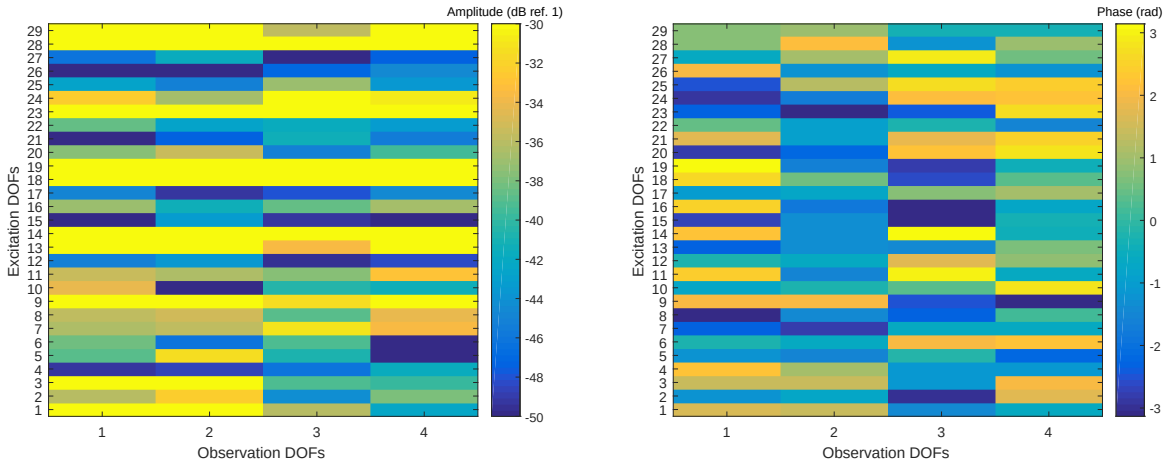
Bayesian hypothesis testing:

$$\forall i = 1, \ldots, N_p, \forall n = 1, \ldots, N_f \quad \begin{array}{l} H_0 : c_{i,n} = 0 \\ H_1 : c_{i,n} \neq 0 \end{array} \tag{25}$$

where H_0 is the null hypothesis and H_1 is the alternative hypothesis. In the Bayesian approach the posterior PDF of the component $c_{i,n}$ is known and the null hypothesis can be tested with a significance level α simply looking if the null value is included in the $(100 - \alpha)\%$ credible interval of the component. If this is the case, then the component is considered null and the support reduced. Figure 7 illustrates the procedure for the independent component around 68 Hz.

After the support reduction, point estimators of the decomposition factors are enough to obtain a reduced surrogate model of the FRFs. A first stage and second stage data reduction ratios can thus be defined respectively as:

$$\begin{aligned} \mathcal{R} &= \frac{N_e \cdot N_o \cdot N_f}{N_p(N_e \cdot N_o + N_f)} \\ \mathcal{R}' &= \frac{N_e \cdot N_o \cdot N_f}{N_e \cdot N_o \cdot N_p + \sum_{i=1}^{N_p} N_f^{(i)}} \end{aligned} \tag{26}$$



(a) (b)

Figure 6: The spatial "pattern" corresponding to component $c_{12}(\omega)$ of Fig. 5: magnitude (a) and phase (b). The "pattern" contains the relations between 4 acoustical output DOFs and 29 structural excitation DOFs.

where $N_f^{(i)}$ for $i \in [1, \ldots, N_p]$ is the size of the spectral support after the above reduction is performed for each independent component. In the present application $\mathcal{R} = 4.8$ and is due to the description through only $N_p = 19$ components. Then, the reduction of the spectral support applied to the measured FRF matrix with a significance level $\alpha = 1\%$ yields a second stage data reduction ratio $\mathcal{R}' = 7.1$.

3.1.2. FRF synthesis from the model

To verify the quality of the overall data reduction, a synthesis step is performed to compare the probabilistic model to the measurement. Figure 8 shows the mean reconstruction of two input frequency response functions. It is noticeable that the ICA yields the same error on all the FRFs and since the input data have not been standardized (i.e. reducing the variance to the unity), this error is negligible compared to the highest FRF values, but relatively significant for lower ones. This is purely a modelling choice that meets engineering needs, which mostly focus on the highest FRF values.

To go further in the comparison, a scalar indicator is necessary. It is reminded that the measurement $\overline{\mathbf{Y}}$ is considered as a sample of the random variable $\mathbf{Y}$. As a consequence, one trivial and effective indicator is the evaluation of the posterior marginal PDF of $\mathbf{Y}$ at the measurement $\overline{\mathbf{Y}}$. More interestingly, this indicator can be computed for each frequency and each observation-excitation DOF couple as:

$$\begin{aligned} &\forall j = 1, \ldots, N_e \cdot N_o, \quad \forall n = 1, \ldots, N_f, \\ &\mathcal{I}_{j,n} = \left(\int_{D_l} \int_{D_c} \int_{D_\sigma} [y_{j,n}|\mathbf{l}_j, \mathbf{c}_n, \sigma_y^2][\mathbf{l}_j, \mathbf{c}_n, \sigma_y^2|\overline{\mathbf{y}}_v] d\sigma_y^2 d\mathbf{c}_n d\mathbf{l}_j \right)_{y_{j,n}=\overline{y}_{j,n}} \end{aligned} \tag{27}$$

where the notations of the algorithm are used again.

Eq. (27) is the evaluation at the value $\overline{y}_{j,n}$ of the predictive posterior PDF of the (j,n) random variable out of $\mathbf{Y}$. It is written in terms of the posterior conditional PDF of $y_{j,n}$ multiplied by the joint posterior PDF of the model parameters, which are then marginalised out. In practice, the predictive posterior PDF of $y_{j,n}$ is completely characterized through the Bayesian algorithm, so that it is simply approximated by the samples issued from the Gibbs sampling as:

$$\forall r = 1, \ldots, N_{run} \quad y_{j,n}^{(r)} = \mathbf{l}_j^{(r)} \mathbf{c}_n^{(r)} + e_{j,n}^{(r)} \tag{28}$$

where N_{run} is the number of MCMC iterations, $\bullet^{(r)}$ stands for the r^{th} sample and $e_{j,n}^{(r)}$ is a sample from a complex circular Gaussian distribution of zero mean and variance $\sigma_y^{2(r)}$.

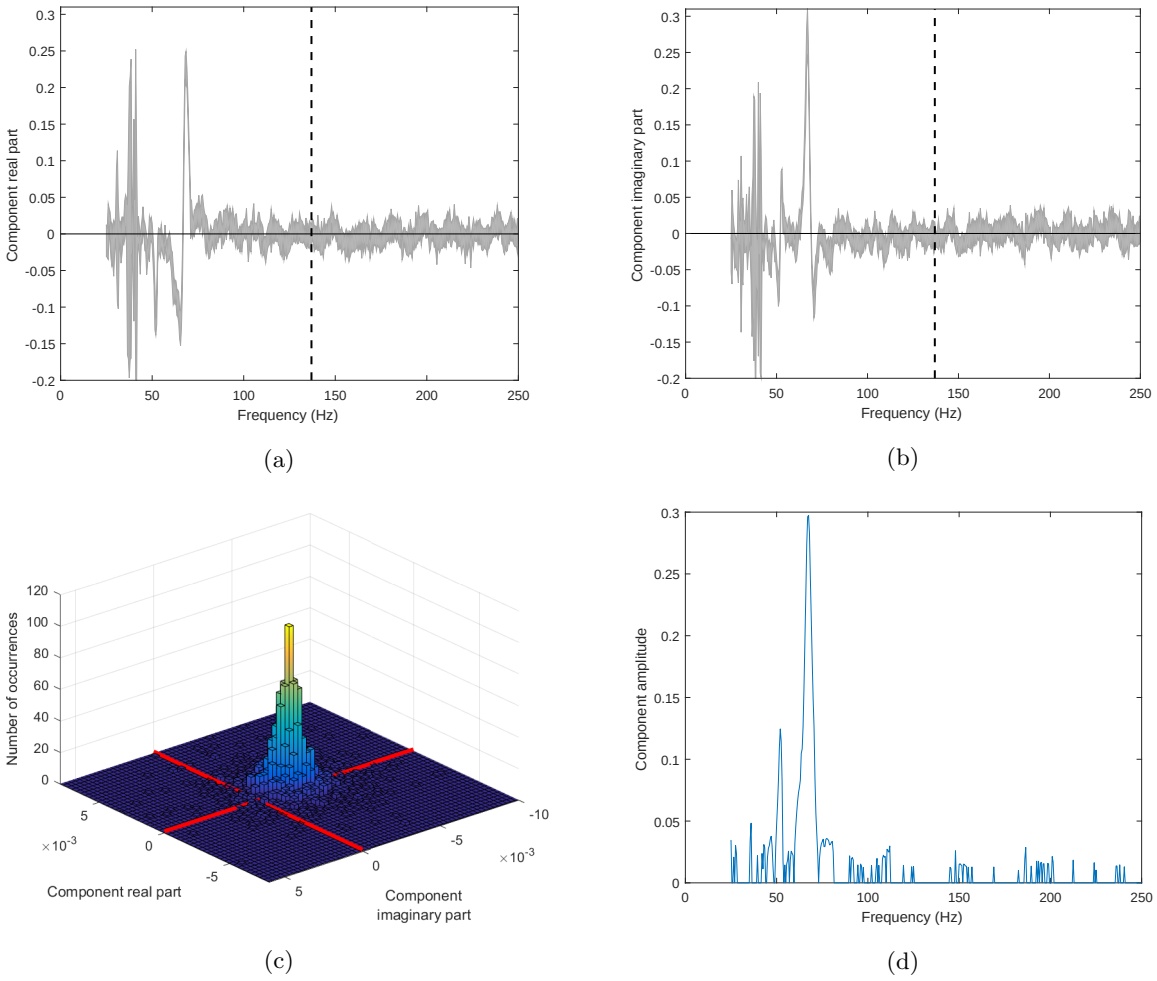


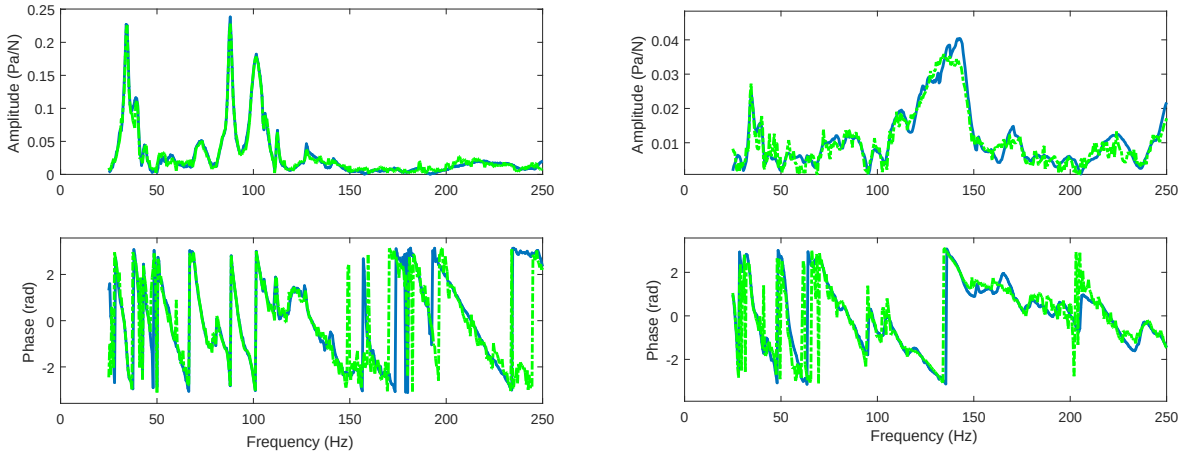
Figure 7: Application of the thresholding process: (a) and (b) show the real and imaginary parts of an independent component (3000 samples). For each frequency (e.g. the one at the dashed line - 137 Hz), the distribution over the Gibbs samples is considered (c). If the null value - at the red lines crossing - is inside the $(100-\alpha)\%$ credible region, the component value is considered null and the spectral support reduced. Figure (d) shows the component after the reduction with $\alpha = 1\%$.

Finally, in order to reduce the indicator support to [0 1], the predictive posterior PDF of $y_{j,n}$ is normalized by its maximum before being evaluated at $\overline{y}_{j,n}$. The obtained indicator is noted $\mathcal{I}^s_{j,n}$.

Figure 9 shows one particular FRF measurement with respect to its probabilistic model. The indicator $\mathcal{I}^s_{53,n}$ can be computed for each frequency $n \in [1, \ldots, N_f]$ and it is illustrated in figure 9(b). For the interpretation of $\mathcal{I}^s_{j,n}$, the measurement $\overline{y}_{j,n}$ has to be considered as one possible outcome of a random process:

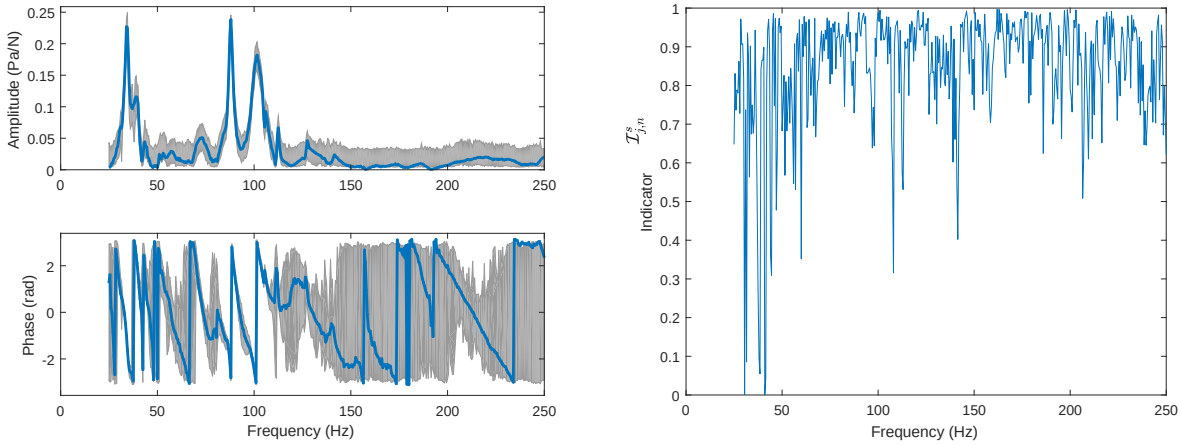
- If the indicator is close to 1, it means that the probabilistic model admits the measured value as the most probable one and only a random error is encountered when resynthesizing the measurement;
- If it is null, it means that the probabilistic model does not admit the measurement as a possible value and a bias error is present when reconstructing the input. This bias error may be small or big, no information about it is contained in the indicator;
- If it is somewhere between 1 and 0, the probabilistic model admits the measurement as a possible value, but not as the most probable one. This means that if the mean is chosen as point estimate for the measurement reconstruction, a relatively small bias error will pollute the result.

Ideally, $\mathcal{I}^s_{j,n}$ should very rarely be null and most often close to 1. Figure 9(b) illustrates well this case and



(a) (b)

Figure 8: Comparison between the mean FRF returned by the Bayesian hierarchical ICA model (green dotted line) and its measurement (blue line) for a high amplitude FRF (a) and a weak one (b).



(a) (b)

Figure 9: (a) comparison between one measured FRF (blue line - the same as in Fig. 8(a)) and the possible outcomes admitted by the Bayesian hierarchical ICA model (grey area - 90% credible interval). (b) frequency evolution of the indicator $\mathcal{I}^s_{53,:}$ for this FRF.

it testifies a good probabilistic model of the measured FRF.

3.2. A numerical application

For the FRFs population based application, several samples of the same FRFs matrix are processed. In particular, $N_{mc} = 20$ samples of the transfer matrix are obtained from a non-parametric stochastic finite element model of a production car. In this approach, the reduced mass, stiffness and damping matrices are computed from the finite element model of the vehicle, which is here an industrial model featuring around ten millions DOFs. These matrices are then used to obtain the searched FRF matrix. The computation is run several times (here 20 times) in a Monte Carlo framework and each time the modal matrices are modified according to the random matrix theory exposed by Soize [57]. This yields a population of frequency response functions each slightly different from the other. Their uncertainty is ruled by some dispersion hyper-parameters that are previously identified in order to fit the corresponding end-of-line measurements

[40]. In the present case, the coefficients had been previously identified[4] in order to simulate the uncertainty encountered in vibro-acoustic FRFs from end-of-line measurements on a population of vehicles belonging to the same model. The used protocol, now typical in Groupe PSA, is presented by Durand [39]. Note that the measured vehicles differed in engine type, gearbox and other customer options. Figure 10 shows the simulated uncertainty of the transfer function between one engine mount and a microphone inside the car. In accordance with the vehicle population used to identify the dispersion hyper-parameters, this simulation should cover production uncertainty, manufacturing variability and measurement errors.
In the present application, the observation DOFs are $N_o = 4$ acoustic pressures at the ears of the driver and passengers, while the excitations DOFs are $N_e = 24$ loads at different connections to the car body. The finite element computation is performed for $N_f = 226$ frequency points in the [25 250] Hz frequency band using the commercial solver Nastran.

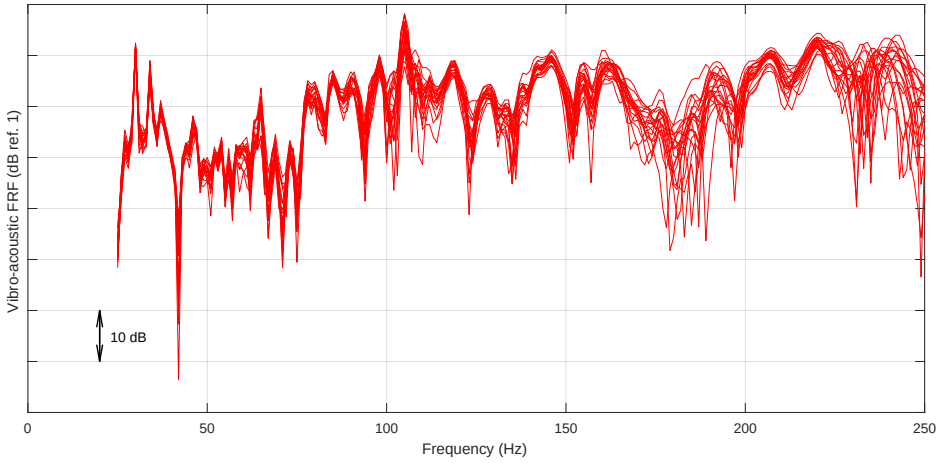


Figure 10: Illustration of structural uncertainties and measurement error on a vibro-acoustic FRF: 20 samples issued from a non-parametric stochastic finite element computation.

The number of components may be chosen applying the same method as before on the mean FRF matrix (obtained averaging the non-parametric samples). Here, using a threshold at 97.5% on the cumulative sum of eigenvalues, it yields $N_p = 23$. Then, the proposed algorithm can be applied, being aware that:

- At each Gibbs sampling iteration, the input $\overline{\mathbf{Y}}$ can be different: the FRF matrix to use is sampled uniformly from the FRF matrices of the non-parametric approach.
- The JADE estimations used for the prior laws of the components are computed on the mean FRF matrix. This helps to make the algorithm converge towards a stable mean value while propagating the uncertainty of the input to the extracted components.

Figure 11 shows the mean of each extracted component. As it can be seen, even if the input is uncertain the proposed algorithm still manages to separate components located in different frequency ranges. Looking more into details, figures 12(a) and 12(b) show the box-plots of the spatial patterns corresponding to the components $c_4(\omega)$ and $c_{23}(\omega)$, respectively dominant around 77 Hz and 248 Hz. Interestingly, the lower frequency pattern (Fig. 12(a)) appears to be less uncertain than the higher frequency one (Fig. 12(b)). This is consistent with the FRF uncertainty which increases with the frequency [39]. This trend is once more verified looking at the evolution of the inferred variance $\sigma^2_{y_n}$ along the frequency. Figure 13 shows the evolution along the frequency - i.e. for $n = 1, \ldots, N_f$ - of a point estimator (the mode) of the posterior PDF of the variance $\sigma^2_{y_n}$: as it can be noticed, the variance globally increases as the frequency does.
As for the experimental application, further data reduction can be achieved by Bayesian hypothesis testing by

[4]Not in the context of this study and not in the Bayesian framework.

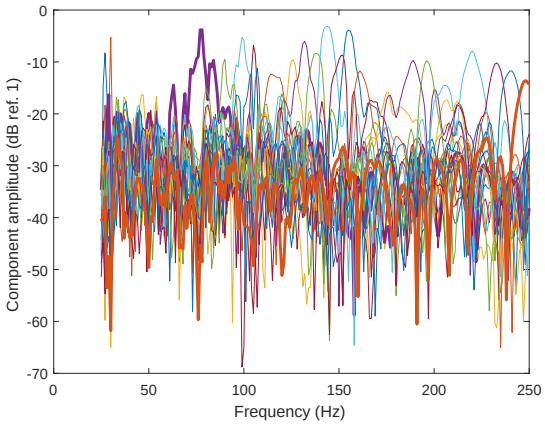
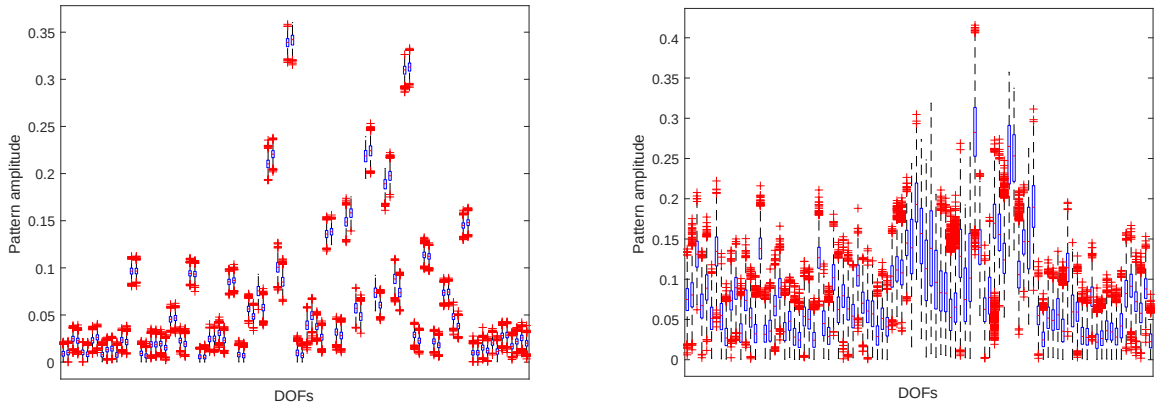


Figure 11: Mean components $c_p(\omega)$ with $p = 1, \ldots, 23$ obtained by Bayesian hierarchical ICA. Highlighted the components $c_4(\omega)$ and $c_{23}(\omega)$ used in Fig. 12.



(a) (b)

Figure 12: Box-plots of the spatial patterns corresponding to components (a) $c_4(\omega)$ - concentrated around 77 Hz - and (b) $c_{23}(\omega)$ - concentrated around 248 Hz.

reducing the spectral support when the extracted components are not statistically different from 0. Figure 14 shows the results of this processing approach for the component $c_4(\omega)$ whose spectral support can be reduced roughly to the half.

Finally, it is important to compare the reconstructed FRF matrix to the actual input data. Here it means comparing the posterior PDF of the model prediction as written in Eq. (28) to the PDF delivered by the non-parametric probabilistic finite element model. To this aim, the Jensen-Shannon divergence is used [58]. It is a symmetrized and smoothed version of the Kullback-Leibler divergence [59] and, given two probability density functions $[x]_P$ and $[x]_Q$, it reads:

$$\mathcal{D}_{JS}([x]_P||[x]_Q) = \frac{1}{2}\mathcal{D}_{KL}([x]_P||[x]_M) + \frac{1}{2}\mathcal{D}_{KL}([x]_Q||[x]_M) \tag{29}$$

where $[x]_M = 1/2([x]_P + [x]_Q)$ and $\mathcal{D}_{KL}(\bullet||\bullet)$ stands for the Kullback-Leibler divergence, which is defined for continuous PDFs $[x]_P = p(x)$ and $[x]_Q = q(x)$ of the random variable x as:

$$\mathcal{D}_{KL}([x]_P||[x]_Q) = \int_{-\infty}^{+\infty} p(x) \ln \frac{p(x)}{q(x)} dx \tag{30}$$

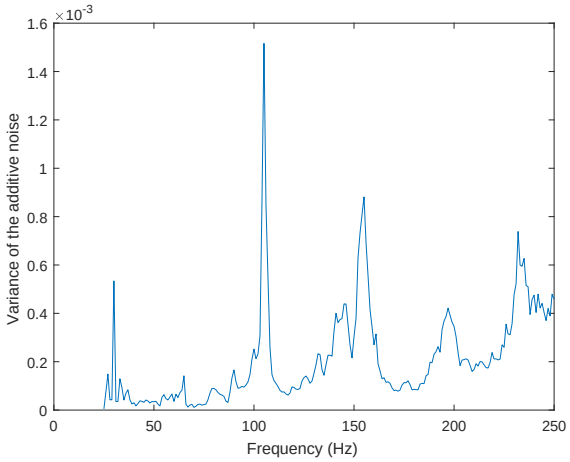
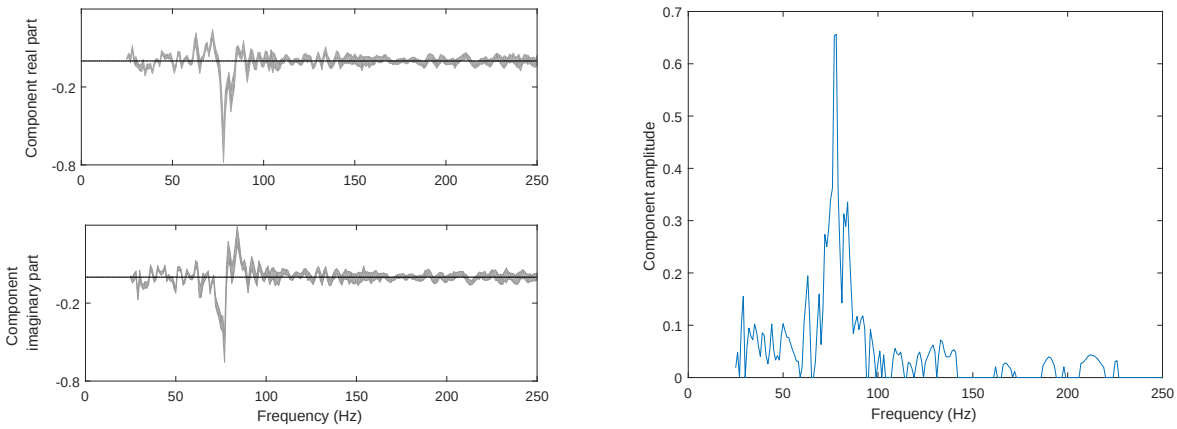


Figure 13: Spectral evolution of the mode of the posterior PDF of the additive noise variance $\sigma^2_{y_n}$.



(a) (b)

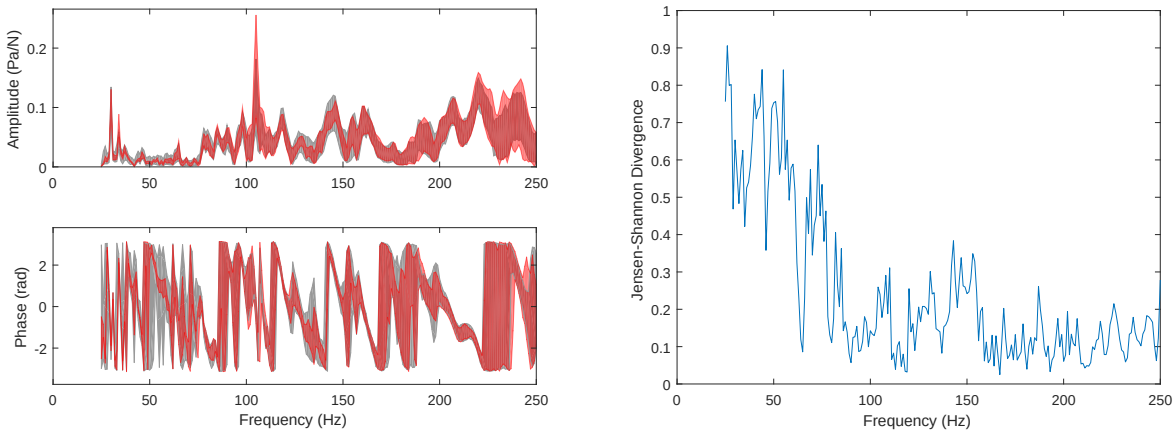
Figure 14: Thresholding process for the component $c_4(\omega)$: its distribution (a) and its amplitude after applying the threshold (b).

Like the Kullback-Leibler divergence, the Jensen-Shannon divergence gives some information on how a probability density functions diverges from another and, unlike the former, it has the advantage to be symmetric and to have support [0 1], with 0 for identical distributions and 1 for distributions that do not share the same support. Other values express grades of different behaviour between the two distributions: the closer to 0 the divergence, the more the two probability distributions behave similarly.

In this application, the PDF of the model prediction and the one issued from the non-parametric finite element computation are both approximated by kernel density estimation from their samples [60, 61]. Figure 15 shows the comparison between the reconstruction and the original data for one FRF and the corresponding Jensen-Shannon divergence. It can be noticed that the reconstruction support covers quite well the Monte Carlo support from the non-parametric probabilistic finite element model. This is also testified by relatively low values of the divergence particularly above 80 Hz.

4. Conclusions

In order to reduce and analyse large matrices of frequency response functions, this paper has introduced a decomposition as a sum of frequency independent matrices containing the magnitude and phase relations



(a) (b)

Figure 15: (a) comparison between the 90% credible interval of numerical FRFs (the same as in Fig. 10(a) - red area) and the one resulting from the Bayesian hierarchical model (grey area). (b) frequency evolution of the Jensen-Shannon divergence for this FRF.

between observation and excitation DOFs, each multiplied by a scalar sparse FRF. Each "pattern" matrix resumes the dynamic behaviour of the system over a frequency range: it may be interpreted as a "generalized" mode shape of the system, in the sense that it isolates a remarkable spatial pattern with a localized frequency content. Thus, the performance of the proposed approach is not related to the extraction of the system mode shapes, but rather to the identification of few "patterns" that alone are enough to accurately describe the measured FRFs. This yields a better insight on the system. The decomposition is achieved by means of a hierarchical Bayesian independent component analysis based on the Gibbs sampling algorithm.

The Bayesian context allows several samples of the FRF matrix to be taken into account with hardly any modification to the algorithm, so that structural uncertainties and random measurement errors in the processed samples can be easily propagated to the extracted components. Then, intrinsic to the Bayesian approach, point or interval estimators can be directly deduced from the joint posterior distribution of all the quantities involved in the FRF matrix decomposition. This allows an advanced second stage data reduction to be performed by pruning all insignificant coefficients. On the other hand, the Gibbs sampler used to explore the joint posterior distribution may need some specific knowledge on Markov Chain Monte Carlo algorithms in order to ensure its good exploitation. The computation time is also higher than for an ICA performed through classical batch algorithms. For these reasons, the proposed algorithm is specifically adapted when the uncertainty on the FRFs has to be conserved and so shared among the independent components.

Complex automotive vibro-acoustic transfer functions served in this work as input data for the proposed decomposition algorithm. However, its purpose is general and it can be used any time a physical insight on FRF matrices is sought and any time several measured or computed samples have to be analysed in a single unified approach.

Acknowledgments

This work was funded by Groupe PSA and performed within the framework of the OpenLab Vibro-Acoustic-Tribology@Lyon and the LabEx CeLyA of Université de Lyon, operated by the French National Research Agency (ANR-10-LABX-0060/ANR-11-IDEX-0007).

AppendixA. Usual probability density functions

AppendixA.1. Multivariate real normal distribution

The multivariate normal distribution of a N-dimensional random vector $\mathbf{x} \in \mathbb{R}^{N\times 1}$ is noted:

$$[\mathbf{x}] \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{A.1}$$

where $\boldsymbol{\mu}$ is the N-dimensional mean vector

$$\boldsymbol{\mu} = E[\mathbf{x}] \tag{A.2}$$

and $\boldsymbol{\Sigma}$ is the $N \times N$ covariance matrix

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^t] \tag{A.3}$$

with $E[\bullet]$ the expected value operator and $\bullet^t$ the transpose operator.
In the non-degenerate case (i.e. $\boldsymbol{\Sigma}$ is definite positive), the PDF of the multivariate real normal law is written:

$$[\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}] = \frac{1}{(2\pi)^{N/2}\det^{\frac{1}{2}}\boldsymbol{\Sigma}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{A.4}$$

where $\det \bullet$ stand for the determinant of a matrix.

AppendixA.2. Gamma distribution

The Gamma distribution can be parametrized using the shape parameter α and the rate parameter β. A random variable x which follows a Gamma distribution is noted:

$$[x] \sim \text{Gamma}(\alpha, \beta) \tag{A.5}$$

The corresponding PDF is written as:

$$[x|\alpha, \beta] = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \tag{A.6}$$

where $\Gamma(\bullet)$ stands for the Gamma function. Its support is $\mathbb{R}^+$.

AppendixA.3. Inverse-Gamma distribution

The Inverse-Gamma distribution can be parametrized using the shape parameter α and the rate parameter β. A random variable x which follows an Inverse-Gamma distribution is noted:

$$[x] \sim \text{InvGamma}(\alpha, \beta) \tag{A.7}$$

The corresponding PDF is written as:

$$[x|\alpha, \beta] = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}} \tag{A.8}$$

Its support is $\mathbb{R}^+$. If $[x] \sim \text{Gamma}(\alpha, \beta)$ then $[1/x] \sim \text{InvGamma}(\alpha, \frac{1}{\beta})$.

AppendixA.4. Exponential distribution

The exponential distribution is parametrized using a rate parameter λ. A random variable x which follows an Exponential distribution is noted:

$$[x] \sim \text{Exp}(\lambda) \tag{A.9}$$

The corresponding PDF is written as:

$$[x|\lambda] = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{A.10}$$

Its support is $\mathbb{R}^+$. If $[x] \sim \text{Exp}(\lambda)$ then $[x] \sim \text{Gamma}(1, \lambda)$.

AppendixA.5. Multivariate complex normal distribution

If the complex normal random vector $\mathbf{x}$ is improper, than the PDF is written:

$$[\mathbf{x}|\underline{\boldsymbol{\mu}}_x, \underline{\boldsymbol{\Sigma}}_{xx}] = \frac{1}{\pi^n \det^{\frac{1}{2}} \underline{\boldsymbol{\Sigma}}_{xx}} \exp\left\{-\frac{1}{2}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}_x)^H \underline{\boldsymbol{\Sigma}}_{xx}^{-1}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}_x)\right\} \tag{A.11}$$

where the underlined elements are the *augmented* versions the random variable ($\underline{\mathbf{x}}$), the mean ($\underline{\boldsymbol{\mu}}$) and the covariance ($\underline{\boldsymbol{\Sigma}}_{xx}$).

If the complex normal random vector $\mathbf{x}$ is proper, than the PDF is written:

$$[\mathbf{x}|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}] = \frac{1}{\pi^n \det \boldsymbol{\Sigma}_{xx}} \exp\left\{-(\mathbf{x} - \boldsymbol{\mu}_x)^H \boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)\right\} \tag{A.12}$$

where $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ are respectively the mean and the covariance of $\mathbf{x}$.

AppendixB. Posterior computation for the algorithm

In what follows, the details about the computation of the Bayesian posterior distributions for the algorithm are proposed. Note that all Gaussian distributions are circularly symmetric complex and they are simply noted $\mathcal{N}_y$ with y the random variable that follows the distribution.

The basis of the posterior computation is the Bayes theorem, which writes:

$$[\boldsymbol{\theta}|X] \propto [X|\boldsymbol{\theta}] \cdot [\boldsymbol{\theta}] \tag{B.1}$$

$[X|\boldsymbol{\theta}]$ stands for the likelihood function, while $[\boldsymbol{\theta}]$ stands for the prior PDF of the parameter which has to be inferred. In the presented application, the likelihood of the measure $\mathbf{Y}$ is:

$$[\overline{\mathbf{y}}_v|\boldsymbol{\Lambda}, \mathbf{C}, \sigma_y^2] \sim \mathcal{N}_{\overline{\mathbf{y}}_v}(\hat{\mathbf{y}}_v, \sigma_y^2 \cdot \mathbb{I}_{N_e \cdot N_o \cdot N_f}) \tag{B.2}$$

The posteriors are obtained as follows:

- Posterior law of σ_y^2:

$$\begin{aligned} \text{Likelihood}: &\quad [\overline{\mathbf{y}}_v|\boldsymbol{\Lambda}, \mathbf{C}, \sigma_y^2] \sim \mathcal{N}_{\overline{\mathbf{y}}_v}(\hat{\mathbf{y}}_v, \sigma_y^2 \cdot \mathbb{I}_{N_e \cdot N_o \cdot N_f}) \\ \text{Prior}: &\quad [\sigma_y^2] \sim \text{InvGamma}(\alpha_y, \beta_y) \end{aligned} \tag{B.3}$$

$$\begin{aligned} [\sigma_y^2|\text{rest}] &\propto [\overline{\mathbf{y}}_v|\boldsymbol{\Lambda}, \mathbf{C}, \sigma_y^2][\sigma_y^2] \\ &\propto \mathcal{N}_{\hat{\mathbf{y}}_v}(\hat{\mathbf{y}}_v, \sigma_y^2 \cdot \mathbb{I}_{N_e \cdot N_o \cdot N_f}) \cdot \text{InvGamma}_{\sigma_y^2}(\alpha_y, \beta_y) \\ &\propto \frac{1}{\pi^{N_e \cdot N_o \cdot N_f}(\sigma_y^2)^{N_e \cdot N_o \cdot N_f}} \exp\left(-\frac{1}{\sigma_y^2}(\overline{\mathbf{y}}_v - \hat{\mathbf{y}}_v)^H(\overline{\mathbf{y}}_v - \hat{\mathbf{y}}_v)\right) \frac{\beta_y^{\alpha_y}}{\Gamma(\alpha_y)}(\sigma_y^2)^{-\alpha_y - 1} e^{-\frac{\beta_y}{\sigma_y^2}} \\ &\propto (\sigma_y^2)^{-\alpha_y - N_e \cdot N_o \cdot N_f - 1} \exp\left(-\frac{1}{\sigma_y^2}\left((\overline{\mathbf{y}}_v - \hat{\mathbf{y}}_v)^H(\overline{\mathbf{y}}_v - \hat{\mathbf{y}}_v) + \beta_y\right)\right) \end{aligned} \tag{B.4}$$

  This corresponds to an Inverse-Gamma distribution with parameters $\alpha_y' = \alpha_y + N_e \cdot N_o \cdot N_f$ and $\beta_y' = (\overline{\mathbf{y}}_v - \hat{\mathbf{y}}_v)^H(\overline{\mathbf{y}}_v - \hat{\mathbf{y}}_v) + \beta_y$.

- Posterior law of $\nu_{i,n}^2$, $\forall i = 1, \ldots, N_p$ and $\forall n = 1, \ldots, N_f$:

$$\begin{aligned} \text{Likelihood}: &\quad [c_{i,n}|\hat{c}_{i,n}^J, \nu_{i,n}^2] \sim \mathcal{N}_{c_{i,n}}(\hat{c}_{i,n}^J, \nu_{i,n}^2) \\ \text{Prior}: &\quad [\nu_{i,n}^2] \sim \text{InvGamma}(\alpha_i, \beta_i) \end{aligned} \tag{B.5}$$

595

$$
\begin{aligned}
[\nu^2_{i,n}|\text{rest}] &\propto [c_{i,n}|\hat{c}^J_{i,n},\nu^2_{i,n}][\nu^2_{i,n}] \\
&\propto \mathcal{N}_{c_{i,n}}(\hat{c}^J_{i,n},\nu^2_{i,n})\cdot \text{InvGamma}_{\nu^2_{i,n}}(\alpha_i,\beta_i) \\
&\propto \frac{1}{(2\pi\nu^2_{i,n})^{1/2}}\exp\left(-\frac{1}{\nu^2_{i,n}}(c_{i,n}-\hat{c}^J_{i,n})^H(c_{i,n}-\hat{c}^J_{i,n})\right)\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}(\nu^2_{i,n})^{-\alpha_i-1}e^{-\frac{\beta_i}{\nu^2_{i,n}}} \\
&\propto (\nu^2_{i,n})^{-\alpha_i-1-1}\exp\left(-\frac{1}{\nu^2_{i,n}}\left((c_{i,n}-\hat{c}^J_{i,n})^H(c_{i,n}-\hat{c}^J_{i,n})+\beta_i\right)\right)
\end{aligned}
\tag{B.6}
$$

This corresponds to an Inverse-Gamma distribution with parameters $\alpha'_i = \alpha_i + 1$ and $\beta'_i = (c_{i,n} - \hat{c}^J_{i,n})^H(c_{i,n} - \hat{c}^J_{i,n}) + \beta_i$.

- Posterior law of β_i, $\forall i = 1,\dots,N_p$:

$$
\begin{aligned}
\text{Likelihood}: &\quad [\nu^2_{i,:}|\alpha_i,\beta_i] \sim \prod_{n=1}^{N_f}\text{InvGamma}_{\nu^2_{i,n}}(\alpha_i,\beta_i) \\
\text{Prior}: &\quad [\beta_i] \sim \text{Gamma}(a_\beta,b_\beta)
\end{aligned}
\tag{B.7}
$$

$$
\begin{aligned}
[\beta_i|\text{rest}] &\propto [\nu^2_{i,:}|\alpha_i,\beta_i][\beta_i] \\
&\propto \prod_{n=1}^{N_f}\text{InvGamma}_{\nu^2_{i,n}}(\alpha_i,\beta_i)\cdot\text{Gamma}(a_\beta,b_\beta) \\
&\propto \prod_{n=1}^{N_f}\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}(\nu^2_{i,n})^{-\alpha_i-1}e^{-\frac{\beta_i}{\nu^2_{i,n}}}\frac{b_\beta^{a_\beta}}{\Gamma(a_\beta)}\beta_i^{a_\beta-1}e^{-b_\beta\beta_i} \\
&\propto \beta_i^{a_\beta-1+N_f\alpha_i}e^{-\beta_i(b_\beta+\sum_{n=1}^{N_f}\frac{1}{\nu^2_{i,n}})}
\end{aligned}
\tag{B.8}
$$

600 This corresponds to a Gamma distribution with parameters $a'_\beta = a_\beta + N_f\alpha_i$ and $b'_\beta = b_\beta + \sum_{n=1}^{N_f}\frac{1}{\nu^2_{i,n}}$.

- Posterior law of α_i, $\forall i = 1,\dots,N_p$:

$$
\begin{aligned}
\text{Likelihood}: &\quad [\nu^2_{i,:}|\alpha_i,\beta_i] \sim \prod_{n=1}^{N_f}\text{InvGamma}_{\nu^2_{i,n}}(\alpha_i,\beta_i) \\
\text{Prior}: &\quad [\alpha_i] \sim \text{Exp}(a_\alpha)
\end{aligned}
\tag{B.9}
$$

$$
\begin{aligned}
[\alpha_i|\text{rest}] &\propto [\nu^2_{i,:}|\alpha_i,\beta_i][\alpha_i] \\
&\propto \prod_{n=1}^{N_f}\text{InvGamma}_{\nu^2_{i,n}}(\alpha_i,\beta_i)\cdot\text{Exp}(a_\alpha) \\
&\propto \prod_{n=1}^{N_f}\frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)}(\nu^2_{i,n})^{-\alpha_i-1}e^{-\frac{\beta_i}{\nu^2_{i,n}}}a_\alpha e^{-a_\alpha\alpha_i} \\
&\propto e^{-N_f\log\Gamma(\alpha_i)+\left(\sum_{n=1}^{N_f}\log\frac{\beta_i}{\nu^2_{i,n}}-a_\alpha\right)\alpha_i}
\end{aligned}
\tag{B.10}
$$

This form cannot be recognised as a usual distribution. However, it can be sampled using sampling algorithms such as Metropolis-Hasting or slice sampling.

- Posterior law of $\mathbf{l}_j^t$, $\forall j = 1, \ldots, N_e \cdot N_o$:

$$\begin{aligned} \text{Likelihood}: \quad & [\overline{\mathbf{y}}_j^t | \mathbf{l}_j^t, \mathbf{C}, \sigma_y^2] \sim \mathcal{N}_{\overline{\mathbf{y}}_j^t}((\mathbf{l}_j\mathbf{C})^t, \sigma_y^2 \cdot \mathbb{I}_{N_f}) \\ \text{Prior}: \quad & [\mathbf{l}_j] \propto 1 \end{aligned} \tag{B.11}$$

$$\begin{aligned} [\mathbf{l}_j^t|\text{rest}] &\propto [\overline{\mathbf{y}}_j^t | \mathbf{l}_j^t, \mathbf{C}, \sigma_y^2] \\ &\propto \mathcal{N}_{\overline{\mathbf{y}}_j^t}((\mathbf{l}_j\mathbf{C})^t, \sigma_y^2 \cdot \mathbb{I}_{N_f}) \\ &\propto \frac{1}{(\pi\sigma_y^2)^{N_f}} \exp\left(-\frac{1}{\sigma_y^2}(\overline{\mathbf{y}}_j^t - (\mathbf{l}_j\mathbf{C})^t)^H(\overline{\mathbf{y}}_j^t - (\mathbf{l}_j\mathbf{C})^t)\right) \\ &\propto \exp\left(\frac{1}{\sigma_y^2}(-\mathbf{l}_j^*\mathbf{C}^*\mathbf{C}^t\mathbf{l}_j^t + \overline{\mathbf{y}}_j^*\mathbf{C}^t\mathbf{l}_j^t + \mathbf{l}_j^*\mathbf{C}^*\overline{\mathbf{y}}_j^t)\right) \end{aligned} \tag{B.12}$$

A Multivariate Gaussian distribution for a random variable $\mathbf{l}^t$ with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$ would lead to:

$$[\mathbf{l}^t|\mathbf{m}, \mathbf{S}] \propto \exp\left(-\mathbf{l}^*\mathbf{S}^{-1}\mathbf{l}^t + \mathbf{l}^*\mathbf{S}^{-1}\mathbf{m} + \mathbf{m}^H\mathbf{S}^{-1}\mathbf{l}^t\right) \tag{B.13}$$

Equations (B.12) and (B.13) have the same form and by identification it can be recognised that the posterior distribution of $\mathbf{l}_j^t$ is a Multivariate Gaussian distribution with parameters:

$$\begin{aligned} \boldsymbol{\mu}_{l_j} &= \frac{1}{\sigma_y^2}\boldsymbol{\Sigma}_{l_j}\mathbf{C}^*\overline{\mathbf{y}}_j^t \\ \boldsymbol{\Sigma}_{l_j}^{-1} &= \frac{1}{\sigma_y^2}(\mathbf{C}\mathbf{C}^H)^* \end{aligned} \tag{B.14}$$

- Using the "block" approach, the posterior law of $\mathbf{c}_n$, $\forall n = 1, \ldots, N_f$ is:

$$\begin{aligned} \text{Likelihood}: \quad & [\overline{\mathbf{y}}_n | \boldsymbol{\Lambda}, \mathbf{c}_n, \sigma_y^2] \sim \mathcal{N}_{\overline{\mathbf{y}}_n}(\boldsymbol{\Lambda}\mathbf{c}_n, \sigma_y^2 \cdot \mathbb{I}_{N_e \cdot N_o}) \\ \text{Prior}: \quad & [\mathbf{c}_n | \hat{\mathbf{c}}_n^J, \boldsymbol{\nu}_n^2] \sim \mathcal{N}_{\mathbf{c}_n}(\hat{\mathbf{c}}_n^J, \text{diag}(\boldsymbol{\nu}_n^2)) \end{aligned} \tag{B.15}$$

$$\begin{aligned} [\mathbf{c}_n|\text{rest}] &\propto [\overline{\mathbf{y}}_n | \boldsymbol{\Lambda}, \mathbf{c}_n, \sigma_y^2] \cdot [\mathbf{c}_n | \hat{\mathbf{c}}_n^J, \boldsymbol{\nu}_n^2] \\ &\propto \mathcal{N}_{\overline{\mathbf{y}}_n}(\boldsymbol{\Lambda}\mathbf{c}_n, \sigma_y^2 \cdot \mathbb{I}_{N_e \cdot N_o}) \cdot \mathcal{N}_{\mathbf{c}_n}(\hat{\mathbf{c}}_n^J, \text{diag}(\boldsymbol{\nu}_n^2)) \\ &\propto \frac{1}{(\pi\sigma_y^2)^{N_e \cdot N_o}} \exp\left(-\frac{1}{\sigma_y^2}(\overline{\mathbf{y}}_n - \boldsymbol{\Lambda}\mathbf{c}_n)^H(\overline{\mathbf{y}}_n - \boldsymbol{\Lambda}\mathbf{c}_n)\right) \cdot \\ &\frac{1}{\pi^{N_e \cdot N_o} \prod_{j=1}^{N_e \cdot N_o} \boldsymbol{\nu}_n^2} \exp\left(-(\mathbf{c}_n - \hat{\mathbf{c}}_n^J)^H \text{diag}(\frac{1}{\boldsymbol{\nu}_n^2})(\mathbf{c}_n - \hat{\mathbf{c}}_n^J)\right) \\ &\propto \exp\left(-\mathbf{c}_n^H(\frac{\boldsymbol{\Lambda}^H\boldsymbol{\Lambda}}{\sigma_y^2} + \text{diag}(\frac{1}{\boldsymbol{\nu}_n^2}))\mathbf{c}_n + \mathbf{c}_n^H(\frac{\boldsymbol{\Lambda}^H\overline{\mathbf{y}}_n}{\sigma_y^2} + \text{diag}(\frac{1}{\boldsymbol{\nu}_n^2})\hat{\mathbf{c}}_n^J)\right) \cdot \\ &\exp\left((\frac{\overline{\mathbf{y}}_n^H\boldsymbol{\Lambda}}{\sigma_y^2} + \hat{\mathbf{c}}_n^{J^H}\text{diag}(\frac{1}{\boldsymbol{\nu}_n^2}))\mathbf{c}_n\right) \end{aligned} \tag{B.16}$$

A Multivariate Gaussian distribution for a random variable $\mathbf{c}$ with mean $\mathbf{m}$ and covariance matrix $\mathbf{S}$ would lead to:

$$[\mathbf{c}|\mathbf{m}, \mathbf{S}] \propto \exp\left(-\mathbf{c}^H\mathbf{S}^{-1}\mathbf{c} + \mathbf{c}^H\mathbf{S}^{-1}\mathbf{m} + \mathbf{m}^H\mathbf{S}^{-1}\mathbf{c}\right) \tag{B.17}$$

Equations (B.16) and (B.17) have the same form and by identification it can be recognised that the posterior distribution of $\mathbf{c}_n$ is a Multivariate Gaussian distribution with parameters:

$$\begin{aligned} \boldsymbol{\mu}_{c_n} &= \boldsymbol{\Sigma}_{c_n}\left(\frac{1}{\sigma_y^2}\boldsymbol{\Lambda}^H\overline{\mathbf{y}}_n + \text{diag}\left(\frac{1}{\boldsymbol{\nu}_n^2}\right)\hat{\mathbf{c}}_n^J\right) \\ \boldsymbol{\Sigma}_{c_n}^{-1} &= \frac{1}{\sigma_y^2}\boldsymbol{\Lambda}^H\boldsymbol{\Lambda} + \text{diag}\left(\frac{1}{\boldsymbol{\nu}_n^2}\right) \end{aligned} \tag{B.18}$$

- Using the "one-by-one" approach, the decomposition model of matrix $\mathbf{Y}$ has to be rewritten for all $n = 1, \ldots, N_f$ as:

$$\begin{aligned} \mathbf{y}_n &= c_{i,n}\boldsymbol{\lambda}_i + \sum_{k\neq i} c_{k,n}\boldsymbol{\lambda}_k + \mathbf{e}_n \\ \frac{\boldsymbol{\lambda}_i^H \overline{\mathbf{y}}_n}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i} &= c_{i,n} + \sum_{k\neq i} c_{k,n}\frac{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_k}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i}c_{k,n} + \frac{\boldsymbol{\lambda}_i^H \mathbf{e}_n}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i} \\ \frac{\boldsymbol{\lambda}_i^H \overline{\mathbf{y}}_n}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i} - \sum_{k\neq i} c_{k,n}\frac{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_k}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i}c_{k,n} &= c_{i,n} + \frac{\boldsymbol{\lambda}_i^H \mathbf{e}_n}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i} \end{aligned} \tag{B.19}$$

The left hand term is the random variable $y_{i|-i,n}$ (using the notation in *Algorithm 3*) and conditionally to the other parameters it follows a Gaussian distribution of mean $c_{i,n}$ and variance $\sigma_i^2 = \frac{\sigma_y^2}{\|\boldsymbol{\lambda}_i\|_2^2}$. As a consequence, the posterior law of $c_{i,n}$, $\forall n = 1, \ldots, N_f$ and $\forall i = 1, \ldots, N_p$ is:

$$\begin{aligned} \text{Likelihood}: \quad & [\overline{y}_{i|-i,n}|c_{i,n}, \sigma_i^2] \sim \mathcal{N}_{\overline{y}_{i|-i,n}}(c_{i,n}, \sigma_i^2) \\ \text{Prior}: \quad & [c_{i,n}|\hat{c}_{i,n}^J, \nu_{i,n}^2] \sim \mathcal{N}_{c_{i,n}}(\hat{c}_{i,n}^J, \nu_{i,n}^2) \end{aligned} \tag{B.20}$$

$$\begin{aligned} [c_{i,n}|\text{rest}] &\propto [\overline{y}_{i|-i,n}|c_{i,n}, \sigma_i^2] \cdot [c_{i,n}|\hat{c}_{i,n}^J, \nu_{i,n}^2] \\ &\propto \mathcal{N}_{\overline{y}_{i|-i,n}}(c_{i,n}, \sigma_i^2) \cdot \mathcal{N}_{c_{i,n}}(\hat{c}_{i,n}^J, \nu_{i,n}^2) \\ &\propto \frac{1}{\pi\sigma_i^2}\exp\left(-\frac{1}{\sigma_i^2}(\overline{y}_{i|-i,n} - c_{i,n})^*(\overline{y}_{i|-i,n} - c_{i,n})\right) \cdot \\ &\frac{1}{\pi\nu_{i,n}^2}\exp\left(-\frac{1}{\nu_{i,n}^2}(c_{i,n} - \hat{c}_{i,n}^J)^*(c_{i,n} - \hat{c}_{i,n}^J)\right) \\ &\propto \exp\left(-(\frac{1}{\sigma_i^2} + \frac{1}{\nu_{i,n}^2})c_{i,n}^* c_{i,n} + c_{i,n}^*(\frac{\overline{y}_{i|-i,n}}{\sigma_i^2} + \frac{\hat{c}_{i,n}^J}{\nu_{i,n}^2}) + (\frac{\overline{y}_{i|-i,n}^*}{\sigma_i^2} + \frac{\hat{c}_{i,n}^{J^*}}{\nu_{i,n}^2})c_{i,n}\right) \end{aligned} \tag{B.21}$$

A Gaussian distribution for a random variable c with mean m and covariance s^2 would lead to:

$$[c|m, s^2] \propto \exp\left(-\frac{1}{s^2}c^* c + \frac{1}{s^2}c^* m + \frac{1}{s^2}m^* c\right) \tag{B.22}$$

Equations (B.21) and (B.22) have the same form and by identification it can be recognised that the posterior distribution of $c_{i,n}$ is a Gaussian distribution with parameters:

$$\begin{aligned} \mu_{c_{i,n}} &= \sigma_{c_{i,n}}^2\left(\frac{\overline{y}_{i|-i,n}}{\sigma_i^2} + \frac{\hat{c}_{i,n}^J}{\nu_{i,n}^2}\right) \\ \sigma_{c_{i,n}}^2 &= \left(\frac{1}{\sigma_i^2} + \frac{1}{\nu_{i,n}^2}\right)^{-1} \end{aligned} \tag{B.23}$$

AppendixC. Bayesian ICA algorithm for including input dispersion

Using the likelihood in Eq. (23) in the algorithm leads to the following modifications on the posterior PDFs:

1. $\forall n = 1, \ldots, N_f$, draw a sample from $[\sigma_{y_n}^2|\text{rest}] \sim \text{InvGamma}(\alpha'_y, \beta'_y)$ with

$$\begin{aligned} \alpha'_y &= \alpha_y + N_e \cdot N_o \\ \beta'_y &= \beta_y + \sum_{j=1}^{N_e \cdot N_o} |\overline{y}_{j,n} - \mathbf{l}_j \mathbf{c}_n|^2 \end{aligned} \tag{C.1}$$

2. $\forall j = 1, \ldots, N_e \cdot N_o$, draw a sample from $[\mathbf{l}_j^t|\text{rest}] \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_{l_j}, \boldsymbol{\Sigma}_{l_j})$ with

$$\begin{aligned} \boldsymbol{\mu}_{l_j} &= \boldsymbol{\Sigma}_{l_j} \mathbf{C}^* \boldsymbol{\Sigma}_y^{-1} \overline{\mathbf{y}}_j^t \\ \boldsymbol{\Sigma}_{l_j} &= ((\mathbf{C}\boldsymbol{\Sigma}_y^{-1}\mathbf{C}^H)^*)^{-1} \end{aligned} \tag{C.2}$$

with $\boldsymbol{\Sigma}_y = diag(\sigma_{y_1}^2, \ldots, \sigma_{y_{N_f}}^2)$.

3. $\forall n = 1, \ldots, N_f$, draw a sample of a column $\mathbf{c}_n$ out of $\mathbf{C}$ from $[\mathbf{c}_n|\text{rest}] \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_{c_n}, \boldsymbol{\Sigma}_{c_n})$ with

$$\begin{aligned} \boldsymbol{\mu}_{c_n} &= \boldsymbol{\Sigma}_{c_n} \left(\frac{1}{\sigma_{y_n}^2} \boldsymbol{\Lambda}^H \overline{\mathbf{y}}_n + \text{diag}\left(\frac{1}{\boldsymbol{\nu}_n^2} \right) \hat{\mathbf{c}}_n^J \right) \\ \boldsymbol{\Sigma}_{c_n} &= \left(\frac{1}{\sigma_{y_n}^2} \boldsymbol{\Lambda}^H \boldsymbol{\Lambda} + \text{diag}\left(\frac{1}{\boldsymbol{\nu}_n^2} \right) \right)^{-1} \end{aligned} \tag{C.3}$$

3 bis. $\forall i = 1, \ldots, N_p$ and $\forall n = 1, \ldots, N_f$, draw a sample from $[c_{i,n}|\text{rest}] \sim \mathcal{N}_{\mathbb{C}}(\mu_{c_{i,n}}, \sigma_{c_{i,n}}^2)$ with

$$\begin{aligned} \mu_{c_{i,n}} &= \sigma_{c_{i,n}}^2 \left(\frac{\overline{y}_{i|-i,n}}{\sigma_{i,n}^2} + \frac{\hat{c}_{i,n}^J}{\nu_{i,n}^2} \right) \\ \sigma_{c_{i,n}}^2 &= \left(\frac{1}{\sigma_{i,n}^2} + \frac{1}{\nu_{i,n}^2} \right)^{-1} \\ \sigma_{i,n}^2 &= \frac{\sigma_{y_n}^2}{\|\boldsymbol{\lambda}_i\|_2^2} \\ \overline{y}_{i|-i,n} &= \frac{\boldsymbol{\lambda}_i^H \overline{\mathbf{y}}_n}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i} - \sum_{k \neq i} \frac{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_k}{\boldsymbol{\lambda}_i^H \boldsymbol{\lambda}_i} c_{k,n} \end{aligned} \tag{C.4}$$

These results are obtained carrying out the same kind of computation presented in Appendix B, which is here omitted for simplicity.
The other posterior PDFs do not change.

References

[1] R. Genesio, M. Milanese, A note on the derivation and use of reduced-order models, IEEE Transactions on Automatic Control 21 (1) (1976) 118–122.

[2] M. Decoster, A. Van Cauwenberghe, A comparative study of different reduction methods, journal A 17 (3) (1976) 125–134.

[3] Y. Shamash, Stable reduced-order models using padé-type approximations, IEEE transactions on Automatic Control 19 (5) (1974) 615–616.

[4] Y. Bistritz, G. Langholz, Model reduction by chebyshev polynomial techniques, IEEE Transactions on Automatic Control 24 (5) (1979) 741–747.

[5] H. Hotelling, Analysis of a complex of statistical variables into principal components., Journal of educational psychology 24 (6) (1933) 417.

[6] F. Chinesta, P. Ladeveze, E. Cueto, A short review on model order reduction based on proper generalized decomposition, Archives of Computational Methods in Engineering 18 (4) (2011) 395.

[7] W. Halvorsen, P. BARNEY, D. Brown, Developing impedance-type models of structural/acoustic systems, Sound and Vibration (1991) 18–26.

[8] K. Dippery, A. Phillips, R. Allemang, Condensation of the spatial domain in modal parameter estimation, in: Proceedings of the 12th International Modal Analysis, Vol. 2251, 1994, p. 818.

[9] J. R. d. F. Arruda, S. A. Vianna do Rio, B. Santos, L. A. Silva, A space-frequency data compression method for spatially dense laser doppler vibrometer measurements, Shock and Vibration 3 (2) (1996) 127–133.

[10] Y. Weiss, H. S. Chang, W. T. Freeman, Learning compressed sensing, in: Snowbird Learning Workshop, Allerton, CA, Citeseer, 2007.

[11] K. Fukunaga, W. L. Koontz, Representation of random processes using the finite karhunen-loeve expansion, Information and Control 16 (1) (1970) 85–101.

[12] Y. Gwon, H. Kung, D. Vlah, Compressive sensing with optimal sparsifying basis and applications in spectrum sensing, in: Global Communications Conference (GLOBECOM), 2012 IEEE, IEEE, 2012, pp. 5386–5391.

[13] N. Saito, B. M. Larson, B. Bénichou, Sparsity vs. statistical independence from a best-basis viewpoint, in: Wavelet Applications in Signal and Image Processing VIII, Vol. 4119, International Society for Optics and Photonics, 2000, pp. 474–487.

[14] N. Saito, The generalized spike process, sparsity, and statistical independence, Modern signal processing 46 (2004) 317–340.

[15] N. Saito, B. Bénichou, The spike process: a simple test case for independent or sparse component analysis, in: Proc. 3rd International Conference on Independent Component Analysis and Signal Separation, T.-W. Lee, T.-P. Jung, S. Makeig, and TJ Sejnowski, eds., IEEE, 2001, pp. 698–703.

[16] J. Eriksson, V. Koivunen, Complex-valued ica using second order statistics, in: Machine Learning for Signal Processing, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop, IEEE, 2004, pp. 183–192.

[17] J.-F. Cardoso, T. Adali, The maximum likelihood approach to complex ica, in: Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, Vol. 5, IEEE, 2006, pp. V–V.

[18] T. Adali, H. Li, M. Novey, J.-F. Cardoso, Complex ica using nonlinear functions, IEEE Transactions on Signal Processing 56 (9) (2008) 4536–4544.

[19] J. Eriksson, V. Koivunen, Complex random vectors and ica models: Identifiability, uniqueness, and separability, IEEE Transactions on Information theory 52 (3) (2006) 1017–1029.

[20] J. Antoni, Blind separation of vibration components: Principles and demonstrations, Mechanical Systems and Signal Processing 19 (6) (2005) 1166–1180.

[21] J. Antoni, S. Chauhan, A study and extension of second-order blind source separation to operational modal analysis, Journal of Sound and Vibration 332 (4) (2013) 1079–1106.

[22] I. G. Araújo, J. A. G. Sánchez, P. Andersen, Modal parameter identification based on combining transmissibility functions and blind source separation techniques, Mechanical Systems and Signal Processing 105 (2018) 276–293.

[23] A. Hyvärinen, J. Karhunen, E. Oja, Independent component analysis, Vol. 46, John Wiley & Sons, 2004.

[24] M. S. Kompella, R. J. Bernhard, Measurement of the statistical variation of structural-acoustic characteristics of automotive vehicles, Tech. rep., SAE Technical Paper (1993).

[25] A. Pereira, J. Antoni, Q. Leclere, Empirical bayesian regularization of the inverse acoustic problem, Applied Acoustics 97 (2015) 11–29.

[26] E. Zhang, J. Antoni, P. Feissel, Bayesian force reconstruction with an uncertain model, Journal of Sound and Vibration 331 (4) (2012) 798–814.

[27] C. Faure, F. Ablitzer, J. Antoni, C. Pézerat, Empirical and fully bayesian approaches for the identification of vibration sources from transverse displacement measurements, Mechanical Systems and Signal Processing 94 (2017) 180–201.

[28] E. Zhang, P. Feissel, J. Antoni, A comprehensive bayesian approach for model updating and quantification of modeling errors, Probabilistic Engineering Mechanics 26 (4) (2011) 550–560.

[29] T. Marwala, I. Boulkaibet, S. Adhikari, Probabilistic finite element model updating using bayesian statistics: applications to aeronautical and mechanical engineering, John Wiley & Sons, 2016.

[30] E. N. Chatzi, C. Papadimitriou, Identification methods for structural health monitoring, Vol. 567, Springer, 2016.

[31] V. Yaghoubi, S. Marelli, B. Sudret, T. Abrahamsson, Sparse polynomial chaos expansions of frequency response functions using stochastic frequency transformation, Probabilistic engineering mechanics 48 (2017) 39–58.

[32] R. Ghanem, P. D. Spanos, Polynomial chaos in stochastic finite elements, Journal of Applied Mechanics 57 (1) (1990) 197–202.

[33] A. Rutherford, D. Inman, G. Park, F. Hemez, Use of response surface metamodels for identification of stiffness and damping coefficients in a simple dynamic system, Shock and Vibration 12 (5) (2005) 317–331.

[34] T. E. Fricker, J. E. Oakley, N. D. Sims, K. Worden, Probabilistic uncertainty analysis of an frf of a structure using a gaussian process emulator, Mechanical Systems and Signal Processing 25 (8) (2011) 2962–2975.

[35] F. DiazDelaO, S. Adhikari, E. S. Flores, M. Friswell, Stochastic structural dynamic analysis using bayesian emulators, Computers & Structures 120 (2013) 24–32.

[36] K. H. Knuth, A bayesian approach to source separation, arXiv preprint physics/0205032.

[37] A. Mohammad-Djafari, A bayesian approach to source separation, in: AIP Conference proceedings, Vol. 567, AIP, 2001, pp. 221–244.

[38] C. Févotte, Bayesian audio source separation, Blind Speech Separation (2007) 305–335.

[39] J.-F. Durand, C. Soize, L. Gagliardini, Structural-acoustic modeling of automotive vehicles in presence of uncertainties and experimental identification and validation, The Journal of the Acoustical Society of America 124 (3) (2008) 1513–1525.

[40] C. Soize, E. Capiez-Lernout, J.-F. Durand, C. Fernandez, L. Gagliardini, Probabilistic model identification of uncertainties in computational models for dynamical systems and experimental validation, Computer Methods in Applied Mechanics and Engineering 198 (1) (2008) 150–163.

[41] K. Pearson, Liii. on lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2 (11) (1901) 559–572.

[42] L. Fontanella, L. Ippoliti, Karhunen–loéve expansion of temporal and spatio-temporal processes, in: Handbook of Statistics, Vol. 30, Elsevier, 2012, pp. 497–520.

[43] R. Dony, et al., Karhunen-loeve transform, The transform and data compression handbook 1 (2001) 1–34.

[44] J.-F. Cardoso, A. Souloumiac, Blind beamforming for non-gaussian signals, in: IEE proceedings F (radar and signal processing), Vol. 140, IET, 1993, pp. 362–370.

[45] T. Bayes, M. Price, An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs, Philosophical Transactions (1683-1775) (1763) 370–418.

[46] W. Gilks, P. Wild, Adaptive rejection sampling for gibbs sampling, Applied Statistics (1992) 337–348.

[47] N. Metropolis, A. W. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, Equation of state calculations by fast computing machines, The journal of chemical physics 21 (6) (1953) 1087–1092.

[48] W. Hastings, Monte carlo sampling methods using markov chains and their applications, Biometrika 57 (1) (1970) 97–109.

[49] S. Geman, D. Geman, Gibbs distributions, and the bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (1984) 721–741.

[50] W. Bolstad, Understanding computational Bayesian statistics, Vol. 644, John Wiley & Sons, 2010.

[51] H. Jeffreys, An invariant form for the prior probability in estimation problems, Proc. R. Soc. Lond. A 186 (1007) (1946) 453–461.

[52] T. Eltoft, T. Kim, T.-W. Lee, Multivariate scale mixture of gaussians modeling, in: International Conference on Independent Component Analysis and Signal Separation, Springer, 2006, pp. 799–806.

[53] E. Hills, B. Mace, N. Ferguson, Acoustic response variability in automotive vehicles, Journal of Sound and Vibration 321 (1-2) (2009) 286–304.

[54] M. Nastran, Dynamic analysis user's guide, MSC Software (2010).

[55] T. Ten Wolde, J. W. Verheij, H. Steenhoek, Reciprocity method for the measurement of mechano-acoustical transfer functions, Journal of Sound and Vibration 42 (1) (1975) 49–55.

[56] G. Brogna, Measured frf matrix - supplementary material, `https://github.com/ntlva/JSV2019_Brogna_supplementary_material` (2019).

[57] C. Soize, Random matrix theory for modeling uncertainties in computational mechanics, Computer methods in applied mechanics and engineering 194 (12-16) (2005) 1333–1366.

[58] J. Lin, Divergence measures based on the shannon entropy, IEEE Transactions on Information theory 37 (1) (1991) 145–151.

[59] S. Kullback, Information theory and statistics, Courier Corporation, 1997.

[60] E. Parzen, On estimation of a probability density function and mode, The annals of mathematical statistics 33 (3) (1962) 1065–1076.

[61] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, The Annals of Mathematical Statistics (1956) 832–837.