Engine noise separation through Gibbs sampling in a hierarchical Bayesian model

G. Brogna, J. Antoni, Quentin Leclere, O. Sauvage

Version of Record: https://www.sciencedirect.com/science/article/pii/S0888327019302225
Manuscript_9233bf89753f75c8b5034c8a89ec63b2

Engine noise separation through Gibbs sampling in a hierarchical Bayesian model

G. Brogna*[a,b], J. Antoni[a], Q. Leclere[a] and O. Sauvage[b]

[a]Univ Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, F-69621 Villeurbanne, France
[b]Groupe PSA, Scientific and Future Technologies Department/StelLab, Route de Gisy, 78943 Vélizy-Villacoublay cedex, France

March 24, 2019

Abstract

An algorithm based on a hierarchical Bayesian model is introduced to separate sources highly overlapping in time and frequency and observed through correlated references. The method is applied to internal combustion (IC) engine signals with the aim of separating the contributions due to different physical origins. The results are compared to the ones provided by classical Wiener filter. The Bayesian context allows correlated references to be taken into account with no consequences on the identifiability of the sources, thanks to the possibility of providing some regularizing prior information in the form of Bayesian prior laws. Moreover, the credibility interval on the estimated sources derives directly from the adopted sampling strategy. Finally, it is shown in a simple case that the proposed algorithm can be rewritten as a weighted sum of the classical and cyclic Wiener filters proposed by Pruvost in 2009 (Ref. [1]). As opposed to them, the present algorithm autonomously chooses one or the other depending on the characteristics of the analysed signals. Even if the development context is the separation of the sources in an IC engine, the presented method is general and can be applied to any source separation problem.
Index terms— Referenced source separation, Bayesian computation, Hierarchical modelling, Engine noise

*gianluigi.brogna@insa-lyon.fr

1 Introduction

Of major concern with NVH characteristics of car vehicles is the control and harmonization of levels of multiple noise sources. One source of primary interest is the engine noise, particularly in its diesel variant. Its sound quality remains one of the most discriminating characteristics in front of the gasoline engine. As a consequence, diesel engines have been the object of several studies aiming at analysing the reasons for this poor acoustic quality [2–5]. Several aspects have been studied, above all the combustion whose irregularities are impacting the perception of the diesel engine sound [6, 7]. Experimental results on the equipment isolated from the working engine are accurate, but not truly indicative of the behaviour of an engine in actual working conditions. For this reason, signal separation methods have become largely applied in the automotive industry: they allow the separation of the contribution of each equipment from the overall noise of a regular engine.
The separation is a difficult task since the noises are highly overlapping in both the time and the frequency domains, especially in the case of impulsive events. The coherence based filtering methods, such as the Wiener filter, present two major limitations:

- they have been developed considering the existence of one reference for each phenomenon whose noise has to be extracted. This means that the reference entirely observes the contribution of the source. Actually, such a unique reference is available only for very few noise sources. This yields to the need to use references that are only partly coherent to a single source, but which may be multiple;
- the reference also needs to be "pure", meaning it is not polluted by any additive noise (even uncoherent with other sources). If multiple references are considered for a single source, a subspace filtering method [8] can be applied in order to denoise them. However, the latter assumes that the Signal to Noise Ration (SNR) of all references is the same. In practice, if the references have different nature the hypothesis of equal SNR cannot be assumed.

Some methods for source extraction succeed in the separation even in operating conditions [1, 9, 10]. For instance, Antoni et al. presented a general separation method based on the Wiener filter in Ref. [11]. El Badaoui et al. [12] used the Wiener filtering method in the context of cyclo-stationary processes [13, 14]. Afterwards, Pruvost et al. developed an improved version of the filter in Ref. [1]. Finally, this approach is extended to the cyclo-non-stationary regime in [16]. These methods answer well to the case of a single but not pure reference. In order to make it pure, its random part - obtained after substraction of the

synchronous average - is isolated and used for the computation of the Wiener filter instead of the whole signal. As a consequence, another definition of the Wiener filter arises in the cyclo-stationary context. This new definition is proved to be better than the former [1] when the random part of the reference is more energetic than the deterministic part. This is not always the case in an operating engine.
A filter with a larger application field would be a linear combination of the classic Wiener filter and the one presented above, weighting them by the mean squared error of their estimation. However, the calculation of the mean squared errors assumes the knowledge of the noise characteristics, which cannot be estimated until after the estimation of the filters and contributions.
The method proposed in this work aims at building a hierarchical model in order to estimate the noise characteristics while estimating the separation filters. Moreover, it generalizes the extraction to the case of multiple polluted references with different SNRs. The Bayesian approach seems the most natural to this aim: a Bayesian hierarchical model of the system is proposed.
In the Bayesian context, the posterior joint probability density function is often fairly complex and a direct sampling of it is not possible. Simple Monte Carlo methods, such as the importance sampling, may fail or be slow when the posterior distribution is concentrated with respect to the prior. In this cases using Markov Chain Monte Carlo (MCMC) methods may be preferable. These methods involve the construction of an ergodic Markov chain whose stationary distribution corresponds to the target posterior probability density function. This way, after having attained the convergence of the chain, the samples from the Markov process are samples from the target distribution as well. Several applications of MCMC methods in different fields of science, such as biology [17], mechanics [18] and space research [19], have focused on their advantages.
The proposed approach will infer the unknown elements through the Gibbs sampler, a popular MCMC algorithm, typically well adapted to hierarchical Bayesian models. First of all, the Gibbs sampler samples from the joint posterior distribution using the conditional posterior distributions of the parameters: this makes possible the use of far simpler probability density functions with remarkable properties (such as Normal laws). Moreover, it keeps the specificities of a Bayesian model. For instance, as opposed to optimisation techniques, the Bayesian credible interval on the results can be easily obtained. The convergence of the application will be verified through parallel chains and the Gelman & Robin convergence test [20].
The first section of this work defines the problem of source separation and its solution using the Wiener filter [1]. The limits of this method come when the references are not pure: they are pointed out using

signals measured on a diesel engine. Therefore, a new model featuring latent variables is described in the same section. The hierarchical Bayesian approach to the problem and the application of the Gibbs sampler are the subject of the second section. It is shown in particular that the obtained filter is a linear combination of the Wiener filters cited above. The third section puts the method in practice in a numerical simulation with synthesized engine signals and its results are compared to the ones provided by a Wiener filter. The use of synthesized signals allows the comparison of the methods to known solutions. The application to real engine signals allows to actually show that the obtained filter can be interpreted as a combination of the Wiener filters: this is addressed in the fourth section. Finally, some perspectives on the method will be provided in the conclusion.
The engine application is just one among others, the presented method updates the existing ones and can be applied to a large scope of separation problems.

2 Problem statement

2.1 Signal modelling

Notational conventions: in this paper, a lower-case character stands for a scalar value, a lower-case bold character stands for a vector and an upper-case bold character stands for a matrix. The problem is first formulated in the time domain - with respect to the time variable t - and then solved in the frequency domain - with respect to the frequency variable f.
The problem consists in the separation of referenced sources. Be an overall measured signal named $d(t)$, composed of the K contributions that have to be separated, $x_k(t), k = 1, ..., K$, and an overall residual noise $n(t)$ independent of the $x_k(t)$'s:

$$d(t) = \sum_{k=1}^{K} x_k(t) + n(t). \tag{1}$$

For each contribution, one reference $r_k(t)$ is available and the relation between the reference and the contribution through a theoretical filter $\tilde{h}_k(t)$ can be written as follows:

$$x_k(t) = (r_k * \tilde{h}_k)(t) \tag{2}$$

where the operator $*$ stands for convolution. Figure 1 shows a schematic view of the model chosen for the overall measured signal $d(t)$.

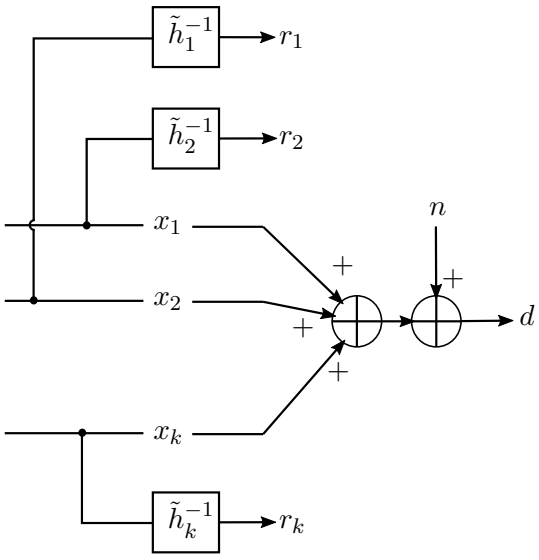

Figure 1: Explanatory diagram of the signal model as defined by equations (1) and (2).

2.1.1 Wiener filters and their limits

The Wiener filter method as explained for instance in Ref. [1] is able to solve such a problem and provide an estimate for the impulse response $\tilde{h}_k$, noted $h_k(t)$, if the references are pure and uncorrelated. A reference is pure if: 1) its signal contains information on the considered source alone and 2) if the information is enough to reconstruct the source perfectly through a linear filter, i.e. the coherence function between the reference and the source is everywhere equal to one. If these conditions are respected, then the method can be applied separately to each contribution. In practice in a Diesel engine this is rarely the case: all signals from the same engine are rather correlated, being all generated by the same rotating machine. The solution proposed in the context of cyclostationarity is to extract and consider just their random parts. For instance, figure 2 shows the cylinder pressure in blue: this reference is typically pure, but it is correlated to other engine signals. Its random part is obtained dividing the signal into cycles and computing the mean along the cycles: this yields an estimate of the first order cyclostationary (CS1) part of the signal, also called the synchronous average (SA). Then, removing the SA from each cycle of the signal leaves its random part (Fig. 3).

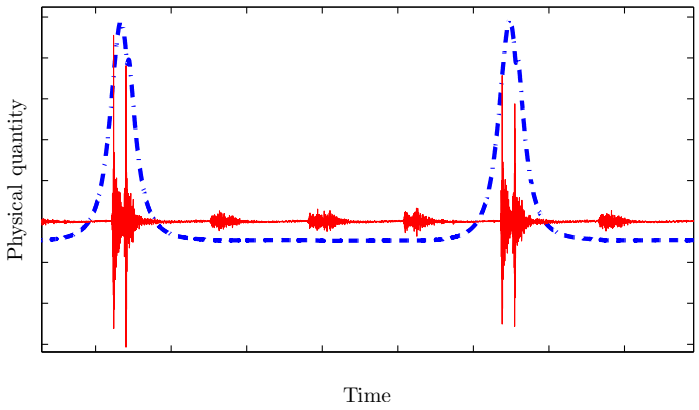

Figure 2: Signals provided by a cylinder pressure probe (solid red line) and by an accelerometer (dashed blue line) on the corresponding injector.

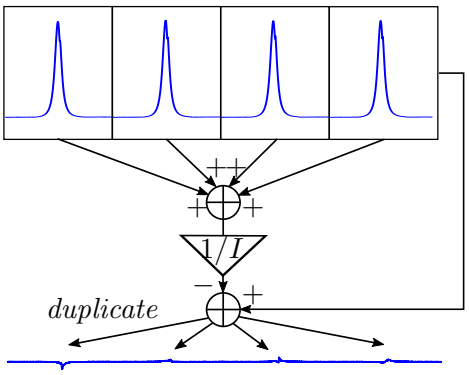

Figure 3: Explanatory diagram of the synchronous average on cyclostationary signals such as the cylinder pressure in a Diesel engine.

Thus, two possible Wiener filters can be defined: the one obtained from raw signals will be called $h_k(t)$ (Fig. 4a), the one obtained from the random part of the signals (removing the synchronous average) will be called $h_k^c(t)$ (Fig. 4b). The random parts being usually not correlated to the other signals by the mere fact of originating from the same machine, $h_k^c(t)$ boasts a smaller bias error. On the other hand, this manipulation decreases the SNR, increasing the random error of the estimator. Therefore, the two filters coexists, each one being better

than the other in some specific cases.

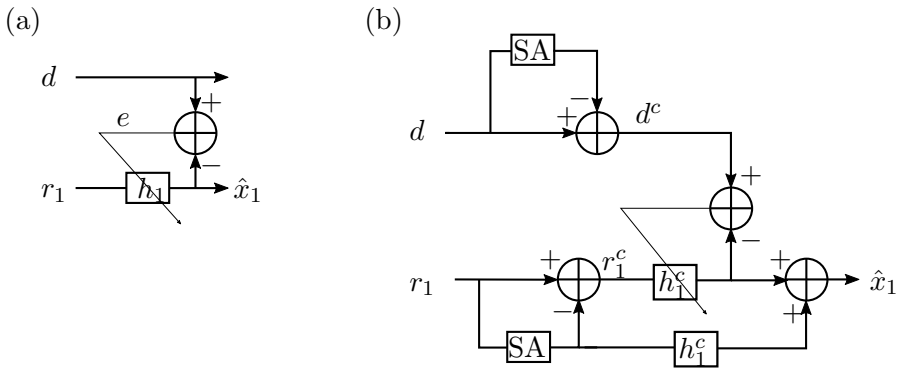

Figure 4: Explanatory diagram of the Wiener filter $h_k(t)$.

The advantages of $h_k^c(t)$ over $h_k(t)$ have been pointed out in Ref. [1, 21] through an application to the combustion noise. Actually, the combustion noise references (the cylinder pressure signals obtained from in-cylinder pressure probes) are well suited to this kind of application since their deterministic part is negligible from a frequency low enough (Fig. 5). This means that using the random part of the references does not decrease the SNR notably, but indeed decorrelates the signals leading to a better estimation of the filter.

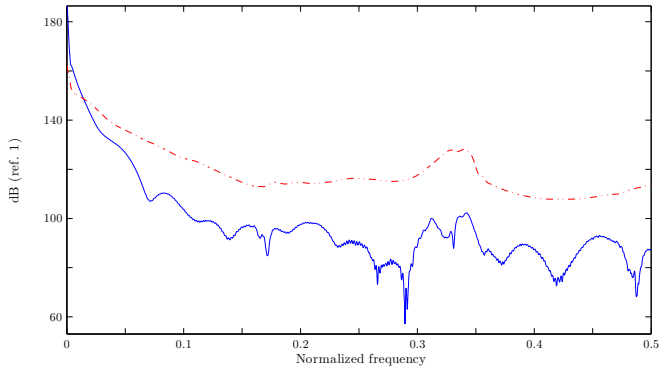

Figure 5: Spectra of the random (dashed red line) and deterministic part (solid blue line) of a cylinder pressure signal.

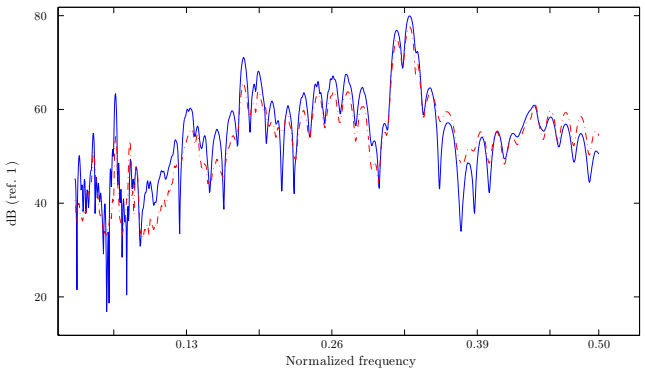

Figure 6: Spectra of the random (dashed red line) and deterministic part (solid blue line) of an accelerometer engine signal.

For all other contributions, the limits of the method are reached. For other kinds of sensors, such as accelerometers, the deterministic and random parts of the signal keep comparable values over all the frequency domain of interest (Fig. 6). As a consequence, removing the deterministic part of the reference decreases the SNR, not necessarily yielding to a gain in the application of this approach.
Moreover, the references must be pure. The random part of the cylinder pressure is a notably pure reference for the combustion noise. On the other hand, the references coming from classical sensors such as accelerometers and microphones are hardly pure. For instance, an accelerometer placed on one injector sees all the combustion events, not just the ones from the corresponding cylinder (see the accelerometer signal in Fig. 2). Usually a correct windowing strategy allows the interesting part of the signal to be isolated, for instance windowing the references around the combustion events of the targeted cylinder. Nevertheless, this approach may be limited when a less impulsive source is present (e.g. the high pressure pump in a diesel engine). One more general solution to this issue is to apply a waterfall approach (Fig. 7). In a multi-source problem, the waterfall approach is the following:

1. separate the source whose reference is the purest;
2. find its estimated contribution $\hat{x}_k(t)$ to the overall measured signal;
3. remove its contribution to the references of the remaining sources, in order to have cleaned references;
4. go back to the beginning and repeat using the following purest reference.

Figure 7 summarizes the procedure. The measured signal $d(t)$ is used with the purest reference $r_1(t)$ to compute a Wiener filter $h_1(t)$ (or $h_1^c(t)$). This filter yields the first source contribution $\hat{x}_1(t)$. After removing this contribution from $d(t)$, the remaining information is used together with the second reference $r_2(t)$ to compute the filter $h_2(t)$ (or $h_2^c(t)$), yielding then the second contribution $\hat{x}_2(t)$. The process is reiterated for each reference, from the purest to the noisiest.

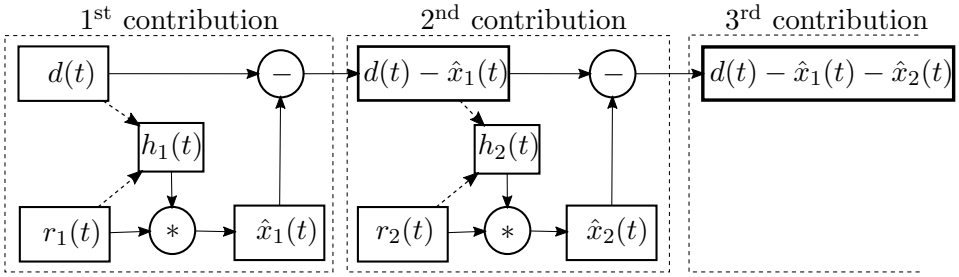

Figure 7: Explanatory diagram for the waterfall strategy.

The choice between $h_k(t)$ and $h_k^c(t)$ is of utmost importance for the waterfall approach and it leads to different estimations of the contributions. For instance, figure 8 shows the spectra of the injectors contribution $\hat{x}_2(t)$ to the noise emission of an engine $d(t)$, computed after a first waterfall step yielding the combustion contribution $\hat{x}_1(t)$. One curve is obtained using in the first step an $h_1(t)$ Wiener filter, whereas the other derives from the application of an $h_1^c(t)$ filter: they are sensibly different. Actually, in an engine application, the difference in meaning between the two filters has been pointed out by Leclère et al. [21]. If the first source considered is the combustion ($k = 1$), $h_1(t)$ extracts the combustion noise and the coherent mechanical noise, while $h_1^c(t)$ extracts the combustion noise alone. As a consequence:

- if $h_1^c(t)$ is used for the first step of the waterfall, then only the combustion contribution is removed from the remaining source references. The mechanical noise of other moving parts is left polluting them. Thus, the contribution extracted using these "partially" cleaned references originates not only from the corresponding source, but contains also some mechanical noise coherent for instance to the combustion itself. This is why in figure 8 the blue spectrum is more energetic than the red one;
- if $h_1(t)$ is used for the waterfall, then the combustion noise and the coherent mechanical noise are completely removed from the references. This yields an "over-cleaning" of the references. Indeed, any mechanical source in an engine is composed of a de-

terministic and a random part, so that its reference should be representative of both. The deterministic part, which repeats itself at each cycle, is strongly linked to the operation of the machine itself. This means that the deterministic parts of all sources (and references) are rather coherent simply from the fact of belonging to the same rotating machine. This coherent part is what has been called above "coherent mechanical noise". Therefore, removing it from a reference means not only removing the pollution from the other sources, but also removing information on the targeted source. For instance in engine applications, it is reasonable to assume that the considered source contribution has a periodic component to it since it follows the rotating process. However, the contribution spectrum obtained by $h_1(t)$ waterfall hardly shows this behaviour (Fig. 8): it seems that the harmonic part in the reference has been cleaned too in the waterfall process.

One of the limits of the Wiener filter method begins to stand out: in the mono-source as in the multi-source cases, two definitions for the Wiener filter coexist, both valid inside their field of application, but none generally accurate, posing a problem of choice.

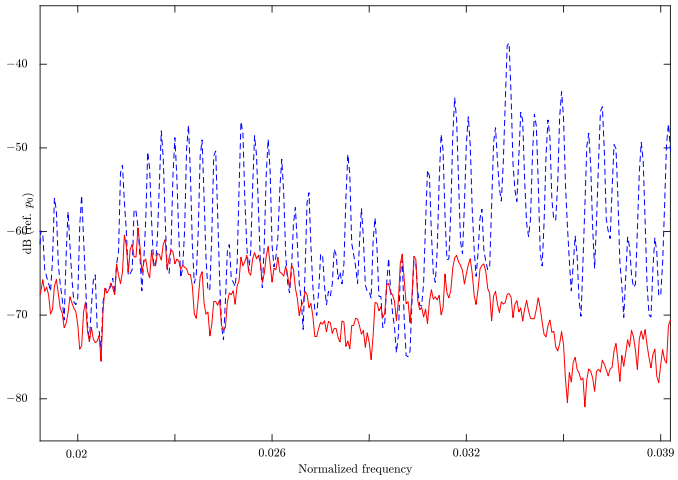

Figure 8: Spectrum of the injectors contribution to the engine noise. Solid red curve: result obtained using the waterfall strategy with an $h_1(t)$ filter. Dashed blue curve: result obtained using the waterfall strategy with an $h_1^c(t)$ filter.

Finally, note that what stated above can be applied to the case in which multiple references are available for each source. It can be said that a group of references is pure if it allows the reconstruction of the source contribution through a linear combination of linear filters. Therefore, it is not pure if:

1. the combination of the references does not entirely observe the source (for instance a frequency range is not available);
2. each reference is polluted by an additive noise.

Increasing the number of references for the same source helps to overcome the first limit, however it does not necessarily help to overcome the second. In order to do so, the references can be denoised through a decomposition on signal and noise subspaces [8]. Nevertheless, this method assumes a spatially white noise (homogeneous among the references), which is a hypothesis that does not stand if the references are of different natures.
After all that is said, the need for a method that computes a filter adapted to the characteristics of the considered signals and that manages to take into account several correlated references with heterogeneous SNRs is evident. This is the aim of the method proposed in this paper.

2.1.2 Consideration of correlated references

The engine signals will be treated in the framework of cyclostationarity and cycloergodicity. In a broad definition a cyclostationary signal is one that exhibits some hidden periodicity of its energy flow [14]. Each one-cycle-long portion of the signal can be considered as a realization of a stochastic process governed by constant characteristics (in our case: the Fourier coefficients corresponding to one frequency form a stationary and ergodic process along the cycles).
The Wold isomorphism [15] allows the model describing the mixing of the sources in a frequency band $[f_l, f_u]$ to be written as:

$$\mathbf{r}_i(f) = \mathbf{G}(f)\mathbf{z}_i(f) + \mathbf{n}_{r_i}(f) \tag{3}$$

$$d_i(f) = \mathbf{h}^t(f)\mathbf{z}_i(f) + d_0(f) + n_{d_i}(f) \tag{4}$$

where $\bullet^t$ stands for the transpose operator, $i \in \{1, ..., I_{obs}\}$ with I_{obs} the number of considered cycles and $\mathbf{r}_i(f) \in \mathbb{C}^{N_r}$ is the column vector of the N_r measured references at frequency f. As stated in Eq. (3), these correlated references are considered as a linear composition of $N_z \leqslant N_r$ uncorrelated latent variables stored in the vector $\mathbf{z}_i(f) \in \mathbb{C}^{N_z}$ and an additive noise $\mathbf{n}_{r_i}(f) \in \mathbb{C}^{N_r}$. The matrix $\mathbf{G}(f) \in \mathbb{C}^{N_r \times N_z}$ describes the relation between the latent references $\mathbf{z}_i$ and the measured ones $\mathbf{r}_i$ at a frequency f.

Equation (4) describes how the sources contribute to the global output signal $d_i(f) \in \mathbb{C}$. It is considered composed of a deterministic part independent from the sources $d_0(f) \in \mathbb{C}$, a series of contributions from the sources of interest and a zero mean additive noise $n_{d_i}(f) \in \mathbb{C}$. The column vector $\mathbf{h}(f) \in \mathbb{C}^{N_z}$ contains the transfer functions for each latent variable at a frequency f. The product between the transfer function and the latent variable yields to the contribution of the considered source to the global noise.
From here on, the dependence on the frequency will be omitted to simplify the notation.

2.2 Physical meaning of the latent variables

As the model has been presented in the previous section, the inputs of the method are the measured references $\mathbf{r}_i$ and the global response d_i. The outputs are all the other terms of the equations. In order to ensure the identifiability, the matrix $\mathbf{G}$ is assumed to have a triangular inferior shape. Actually, if this matrix were full, the problem stated by Eq. (3) would resume to an eigenvalue decomposition, without any direct way to know the physical meaning of the variables. On the other hand, the triangular inferior shape ensure the physical meaning of the latent references to be inferred from the measured ones. This implies that the measured references must be wisely chosen in order for the problem to be solved keeping a physical meaning.
Take for instance a cycle i, let r_1 to r_6 be the measured references, which the user considers well describing the sources he/she would like to extract from the overall engine noise. Let these sources be A, B and C, thus let $\mathbf{z} \in \mathbb{C}^3$ the column vector of the corresponding latent variables. The link between the sources and the latent variable is ensured by the measured references and the matrix $\mathbf{G} \in \mathbb{C}^{6\times 3}$. Equation (3) becomes:

$$\begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \end{pmatrix} = \begin{bmatrix} G_{11} & 0 & 0 \\ G_{21} & G_{22} & 0 \\ 0 & G_{32} & G_{33} \\ G_{41} & 0 & G_{43} \\ G_{51} & G_{52} & G_{53} \\ G_{61} & G_{62} & G_{63} \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{pmatrix} n_{r_1} \\ n_{r_2} \\ n_{r_3} \\ n_{r_4} \\ n_{r_5} \\ n_{r_6} \end{pmatrix}. \tag{5}$$

As it can be seen the first row implies that the physical meaning of the first latent variable is directly linked to the first measured reference alone. If the measured reference r_1 is chosen wisely as to be pure and contain information from the source A alone, it can be implied that the latent variable z_1 describes this source too. Having determined the first latent reference, the second row allows the determination of z_2 even if the measured reference r_2 contains information about both

source A and B. Finally, all other references are not required to be pure, since, having determined the first two latent references, z_3 as a latent pure reference on the source C can be determined as well.
It can be remarked that Eq. (5) describes the spatial decomposition [8] on the signal subspace, generated by the virtual references z_i, and the noise subspace, generated by the noises n_{r_i}. Nevertheless, the method proposed in the present paper does not need to assume the homogeneity of the noises, i.e. with identical variances. As a consequence, it can be applied to references having different natures.

3 Inference of parameters through Gibbs sampling

3.1 Principles

This section presents the way of solving the problem through the implementation of Gibbs sampling in a hierarchical Bayesian model. Let $\mathbf{x}$ be the vector storing the unknown parameters and Y the measurements, the application of the Bayes' rule allows obtaining the joint posterior probability density function (pdf) of $\mathbf{x}$ as:

$$[\mathbf{x}|Y] \propto [Y|\mathbf{x}] \cdot [\mathbf{x}]. \tag{6}$$

In Eq. (6) the following elements can be recognised:

- $[\mathbf{x}|Y]$: the posterior joint pdf of the unknown parameters, i.e. the joint pdf of the parameters given the measurements Y.
- $[Y|\mathbf{x}]$: the likelihood function, which stands for the probability of observing the data Y given a set of unknown parameters $\mathbf{x}$.
- $[\mathbf{x}]$: the prior pdf of the unknown parameters, which reflects what we know about the parameters before having collected the data Y.

The Gibbs sampling [22] is a procedure which, among the MCMC methods, allows Bayesian inference on the unknown parameters using conditional pdfs of one parameter given the others, instead of the joint pdf of all parameters. In this procedure, we will cycle through the parameters drawing each one in its conditional pdf given all other parameters at their most recent values. Each sample of a parameter is a step of a Markov chain whose long-run distribution represents the marginal pdf of the parameter. This method is particularly well suited in a hierarchical model [23].
For the application to the separation problem, each term in equations (3) and (4) is considered as a random variable. The hypothesis on the stationarity and ergodicity of the signals hopping from one cycle to the other yields to the real and imaginary parts of the random

variables to be independent, which facilitates considerably the calculations. The noises n_{d_i} and $\mathbf{n}_{r_i}$ are included in the measurements, which are then considered issued from circularly symmetric complex normal laws, given the other parameters. Thus, the likelihood functions are:

$$(3) \rightarrow [r_{ji}|\mathbf{g}_j, \mathbf{z}_i] \sim \mathcal{N}_{\mathbb{C}}(\mathbf{g}_j \mathbf{z}_i, \sigma^2_{r_j}), \forall i = 1, ..., I_{obs}, \forall j = 1, ..., N_r \quad (7)$$

$$(4) \rightarrow [d_i|\mathbf{h}, \mathbf{z}_i, d_0] \sim \mathcal{N}_{\mathbb{C}}(\mathbf{h}^t \mathbf{z}_i + d_0, \sigma^2_d), \forall i = 1, ..., I_{obs} \quad (8)$$

where $\mathbf{g}_j$ is the j^{th} row of the matrix $\mathbf{G}$ and σ^2_d, $\sigma^2_{r_j}$ are the variances of the zero-mean Gaussian noises in equations (3) and (4).
Each unknown parameter is considered following a probability density distribution whose parameters are considered as random variables as well. Figure 9 shows the associated hierarchical system. Hereafter, there are prior pdfs that have to be chosen for each random variable, where InvGamma(a, b) stands for an Inverse-Gamma law of shape and rate parameters respectively a and b:

$$\begin{gathered}
[d_0] \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2_{d_0}) \\
[\sigma^2_{d_0}] \sim \text{InvGamma}(a_0, b_0) \\
[\sigma^2_d] \sim \text{InvGamma}(a_d, b_d) \\
[\sigma^2_{r_j}] \sim \text{InvGamma}(a_r, b_r), \forall j = 1, ..., N_r \\
[\mathbf{z}_i] \sim \mathcal{N}_{\mathbb{C}}(0, \mathbb{I}), \forall i = 1, ..., I_{obs} \\
[\mathbf{h}^t] \sim \mathcal{N}_{\mathbb{C}}(0, \text{diag}(\sigma^2_{h_1}, ..., \sigma^2_{h_{N_z}})) \\
[\sigma^2_{h_k}] \sim \text{InvGamma}(a_{h_k}, b_{h_k}), \forall k = 1, ..., N_z \\
[\mathbf{g}_j] \sim \mathcal{N}_{\mathbb{C}}(0, \text{diag}(\sigma^2_{G_{j1}}, ..., \sigma^2_{G_{jN_z}})), \forall j = 1, ..., N_r \\
[\sigma^2_{G_{jk}}] \sim \text{InvGamma}(a_{G_{jk}}, b_{G_{jk}}), \forall j = 1, ..., N_r, \forall k = 1, ..., N_z.
\end{gathered} \quad (9)$$

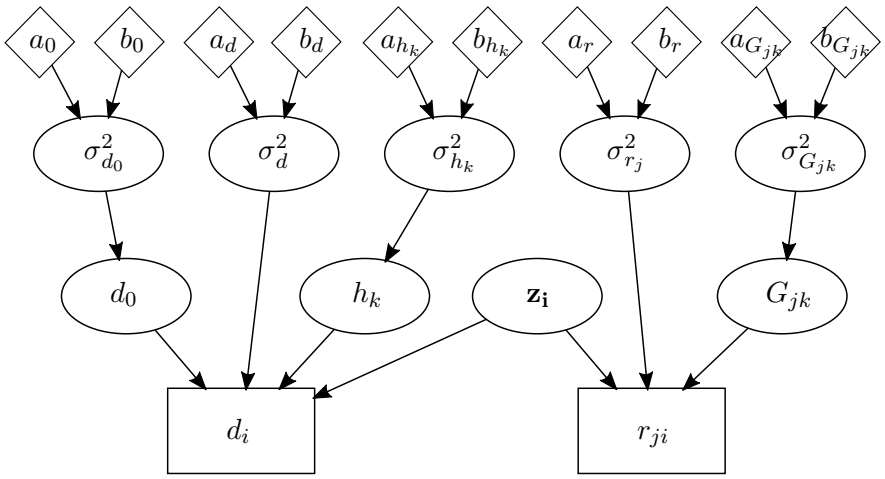

Figure 9: Hierarchical Bayesian model: data (boxes), parameters (ellipses), hyper-parameters (diamonds).

In accordance with the Bayesian method, the prior laws contain the subjective view or the actual knowledge of the user on the parameters. They are usually chosen quite flat in order to reflect the lack of knowledge on the parameters before any measurement. The hyper-parameters (a, b) are chosen by the user, but the higher they are in the hierarchical model, the less influence they have on the posterior pdf of the random variables [23]. In this case of application, the following information has been assumed known on the priors:

- the variances are supposed to follow Inverse Gamma laws since a variance can only have positive values;
- all the other parameters are supposed to follow circularly symmetric complex Normal laws;
- elements of the Gaussian random vector $\mathbf{h}$ are a priori uncorrelated;
- $\mathbf{h}$ is heteroskedastic, i.e. each element of $\mathbf{h}$ can have a different variance. Each variance follows an inverse gamma distribution;
- elements of $\mathbf{G}$ are a priori uncorrelated;
- $\mathbf{G}$ is heteroskedastic. Each variance follows an inverse gamma distribution;
- $\mathbf{z}_i$ follows a multivariate circularly symmetric complex normal distribution whose covariance matrix is the identity matrix $\mathbb{I}$. This means that the latent variables are supposed uncorrelated. An identity covariance matrix is imposed in order for the variances of $\mathbf{h}$ and $\mathbf{G}$ to be identifiable following Eqs. (3) and (4);

- in order to obtain an estimation of $\mathbf{z}_i$ for each cycle and rebuild the time domain signal, the i observations are considered independent realizations of a stochastic process. The Bayesian inference for $\mathbf{z}_i$ is thus realized for each cycle separately.

Finally it can be noted that Inverse Gamma distributions have been chosen for the variances and Gaussian ones for the means, since these two distributions are *conjugate priors* for the likelihood[1]. This is a sensible choice in order to greatly improve the performances of the following Gibbs sampler [23].

3.2 Proposed Gibbs sampler

In what follows the operator $\bullet^H$ is the Hermitian transpose and *rest* stands for all the random variables, except the random variables whose pdf is being expressed. The calculations are detailed in Appendix B. The application of the procedure follows 11 steps.

1. Initialize the values of d_0, $\sigma^2_{d_0}$, σ^2_d, $\sigma^2_{r_j}$, h_k, $\sigma^2_{h_k}$, G_{jk} and $\sigma^2_{G_{jk}}$.
2. $\forall i = 1, ..., I_{obs}$, draw a sample from $[\mathbf{z_i}|rest] \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}_i}, \boldsymbol{\Omega}_{\mathbf{z}_i})$ with

$$\boldsymbol{\mu}_{\mathbf{z}_i} = \boldsymbol{\Omega}_{\mathbf{z}_i}(\mathbf{h}^*(d_i - d_0)/\sigma^2_d + \mathbf{G}^H \begin{bmatrix} \sigma^{-2}_{r_1} & 0 & 0 \\ & \ddots & \\ 0 & 0 & \sigma^{-2}_{r_{N_r}} \end{bmatrix} \mathbf{r}_i)$$

$$\boldsymbol{\Omega}_{\mathbf{z}_i} = (\mathbf{h}^*\mathbf{h}^t/\sigma^2_d + \mathbf{G}^H \begin{bmatrix} \sigma^{-2}_{r_1} & 0 & 0 \\ & \ddots & \\ 0 & 0 & \sigma^{-2}_{r_{N_r}} \end{bmatrix} \mathbf{G} + \mathbb{I})^{-1}.$$

   where $\bullet^*$ stands for the complex conjugate and $\bullet^t$ for the matrix transpose.
3. Draw a sample from $[\sigma^2_d|rest] \sim \text{InvGamma}(I_{obs} + a_d, \sum_{i=1}^{I_{obs}} |d_i - \mathbf{h}^t\mathbf{z}_i - d_0|^2 + b_d)$.
4. $\forall j = 1, ..., N_r$, draw a sample from $[\sigma^2_{r_j}|rest] \sim \text{InvGamma}(I_{obs} + a_r, \sum_{i=1}^{I_{obs}} |r_{ji} - \mathbf{g}_j\mathbf{z}_i|^2 + b_r)$.
5. $\forall k = 1, ..., N_z$, draw a sample from $[\sigma^2_{h_k}|rest] \sim \text{InvGamma}(1 + a_{h_k}, |h_k|^2 + b_{h_k})$.
6. $\forall j = 1, ..., N_r$ and $\forall k = 1, ..., N_z$, draw a sample from $[\sigma^2_{G_{jk}}|rest] \sim \text{InvGamma}(1 + a_{G_{jk}}, |G_{jk}|^2 + b_{G_{jk}})$.

[1] In Bayesian probability, a prior is called a *conjugate prior* for a certain likelihood, if prior and posterior are in the same distribution family. For instance, for a Gaussian likelihood, an Inverse Gamma prior for the variance yields an Inverse Gamma posterior as well.

7. Draw a sample from $[\mathbf{h}^t|rest] \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{h}}, \boldsymbol{\Omega}_{\mathbf{h}})$ with

$$\boldsymbol{\mu}_{\mathbf{h}} = \sum_{i=1}^{I_{obs}} (d_i - d_0)\mathbf{z}_i^H \boldsymbol{\Omega}_{\mathbf{h}} / \sigma_d^2$$
$$\boldsymbol{\Omega}_{\mathbf{h}} = (\sum_{i=1}^{I_{obs}} \mathbf{z}_i \mathbf{z}_i^H / \sigma_d^2 + \begin{bmatrix} \sigma_{h_1}^{-2} & 0 & 0 \\ & \ddots & \\ 0 & & 0 \\ 0 & 0 & \sigma_{h_{N_z}}^{-2} \end{bmatrix})^{-1}.$$

8. $\forall j = 1, ..., N_r$, draw a sample from $[\mathbf{g}_j|rest] \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{g}_j}, \boldsymbol{\Omega}_{\mathbf{g}_j})$ with

$$\boldsymbol{\mu}_{\mathbf{g}_j} = \sum_{i=1}^{I_{obs}} r_{ji} \mathbf{z}_i^H \boldsymbol{\Omega}_{\mathbf{g}_j} / \sigma_{r_j}^2$$
$$\boldsymbol{\Omega}_{\mathbf{g}_j} = (\sum_{i=1}^{I_{obs}} \mathbf{z}_i \mathbf{z}_i^H / \sigma_{r_j}^2 + \begin{bmatrix} \sigma_{G_{j1}}^{-2} & 0 & 0 \\ & \ddots & \\ 0 & & 0 \\ 0 & 0 & \sigma_{G_{jN_z}}^{-2} \end{bmatrix})^{-1}.$$

9. Draw a sample from $[d_0|rest] \sim \mathcal{N}(\mu_{d_0}, \nu_{d_0})$ with

$$\mu_{d_0} = \nu_{d_0} \sum_{i=1}^{I_{obs}} (d_i - \mathbf{h}^t \mathbf{z}_i) / \sigma_d^2$$
$$\nu_{d_0} = (\sigma_{d_0}^{-2} + I_{obs} \sigma_d^{-2})^{-1}.$$

10. Draw a sample from $[\sigma_{d_0}^2|rest] \sim \text{InvGamma}(1 + a_{d_0}, |d_0|^2 + b_{d_0})$.
11. Go back to step 2 and repeat the process until a sufficiently large sample is collected after convergence.

The sequence of samples produced for one of the random variables forms a Markov chain, whose distribution after convergence is the posterior marginal distribution of the random variable. The first draws will not be from the target distribution: the Markov chain needs a burn-in phase after which its distribution is deemed converged. The length of the burn-in phase has to be chosen considering the convergence rate of the Markov chains.
In this application, the convergence has been inspected through the Gelman & Rubin method [20, 23]. Several chains are created starting from different initialization points. Then the inter-chain and the in-chain variances are compared and the information is condensed in a convergence indicator.
For any parameter, a random sample from its prior law can be chosen as initialization point: the Markov Chain Monte Carlo methods guarantee the convergence towards the posterior marginal distribution for any starting point given a large enough number of iterations. However, it could be useful to optimize the convergence in order to reduce the number of iterations and the calculation time. One way of doing this is to initialize the parameters with the results from the application of a

classical Wiener filter. As a consequence, the initialization points will already be near the main mode of the posterior marginal distributions, easing the convergence towards it. An example of the application of the Gelman & Rubin method starting from initialization points scattered around the Wiener filter results is produced in the following section.
It is worth mentioning that, as it can be seen in the algorithm steps, the more the observations the less the prior laws influence the target distributions.

3.3 Interpretation of the Gibbs filter as a linear combination of Wiener filters

As already stated in sub-section 2.1.1, one of the issues in using Wiener filters is the need to choose between $h_k(f)$ and $h_k^c(f)$ without knowing before hand which one estimates best the signal of interest.
In the case of one source and one reference, it can be easily proven that the mean of the filter μ_h after convergence (i.e. the expectation of the posterior law of the filter) is a linear combination of the Wiener filters $h(f)$ and $h^c(f)$. In what follows each term depends on the frequency, but for the sake of simplicity this dependence is left out. Just note that the analysis proposed can be done at each frequency. After development (Appendix C), the following expression can be obtained:

$$\mu_h = k_{h^c} h^c + k_h h = \alpha C h^c + (1-\alpha) h, \tag{10}$$

$$\alpha = \frac{d_0}{d_0 + \bar{x} + \epsilon\sigma_d/\sqrt{I_{obs}}}$$
$$C = \frac{\sum_{i=1}^{I_{obs}} |z_i^c|^2}{\sum_{i=1}^{I_{obs}} |z_i|^2} \quad . \tag{11}$$

In the former equations $\bar{x}$ is the deterministic part of the source contribution, while z_i^c stands for the random part of the reference. The term $\epsilon\sigma_d/\sqrt{I_{obs}}$, with ϵ a standard Gaussian random variable, stands for the measurement noise.
Figures 10 and 11 help the interpretation of the obtained coefficients:

- if the whole reference signal z_i is large with respect to its random part z_i^c, then C tends to zero and the filter obtained through the proposed method tends to h. This corresponds well to the Wiener filtering strategy of favouring the h filter when the random part of the references is too small: removing the synchronous average from the references would lead to a small SNR and a poor h^c filter estimation. It is interesting to point out that, even if C is null, the coefficient k_h is not necessarily equal to 1. This means that k_h can be considered as a corrective term, thus according to the remark of Leclère who noted that the filter h is always

bigger than h^c and it can thus be thought to over-estimate the contribution [21];

- if the reference signal z_i consists mainly in its random part z_i^c then C tends to 1 and the h^c filter is acceptable. Normally, in this case, if the reference has been wisely chosen, the aimed contribution too consists mainly of its random part, then α tends towards 1 and the h^c filter is preferred. This also is consistent with the strategy based on Wiener filter;
- looking more thoroughly at the coefficient α, for an easier understanding it can be rewritten as:

$$\alpha = \frac{1}{1 + \bar{x}/d_0 + \frac{\epsilon}{\sqrt{I_{obs}}}\frac{\sigma_d}{d_0}}. \tag{12}$$

  If the deterministic part of the signal noise (d_0) is large with respect to the variance of the noise σ_d and to the deterministic part of the contribution $\bar{x}$, then α tends to 1 and the h^c filter is preferred (Fig. 10). This means that if d_0 is important, then the filter should be computed from the random parts of the signals, which is in accord with the Wiener filter strategy. The same way, increasing the number of observations I_{obs} yields to an increase in the coefficient α tipping the scales in favour of h^c (Fig. 11);
- on the other hand, the coefficient α tends to zero if $\frac{\bar{x}}{d_0}$ and/or $\frac{\sigma_d}{d_0}$ are large, yielding more weight to the h filter (Fig. 10). This is the case when the deterministic part of the source contribution composes most of the measured signal d_i and it is thus bigger than the deterministic part of the noise (d_0). It holds too when the variance of the zero mean Gaussian additive noise is far bigger than its average d_0 (which represents the deterministic part of the noise), meaning that the noise is essentially random and not correlated with the reference signal and contribution. In both cases, there is no advantage in computing the h^c filter and the use of the h Wiener filter is preferred, which is in accord with what is pointed out in this analysis.

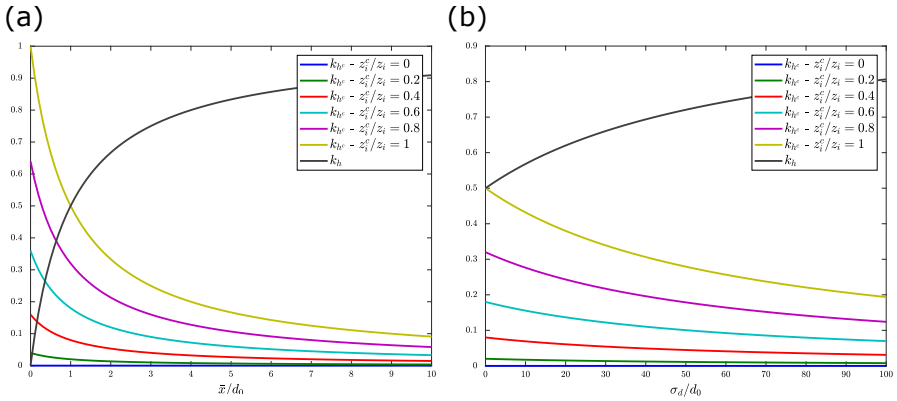

Figure 10: Sensitivity of k_h and k_{h^c} to $\frac{\bar{x}}{d_0}$ (a) and $\frac{\sigma_d}{d_0}$ (b) for several values of the reference random part (ratios $\frac{z_i^c}{z_i}$).

The former analysis shows that the proposed Gibbs filter can be considered as a linear combination of the two Wiener filters previously introduced in sub-section 2.1.1. Whereas the adoption of a Wiener filtering strategy yields to the choice of one filter or the other, the proposed method blends both as a function of the characteristics of the processed signals.

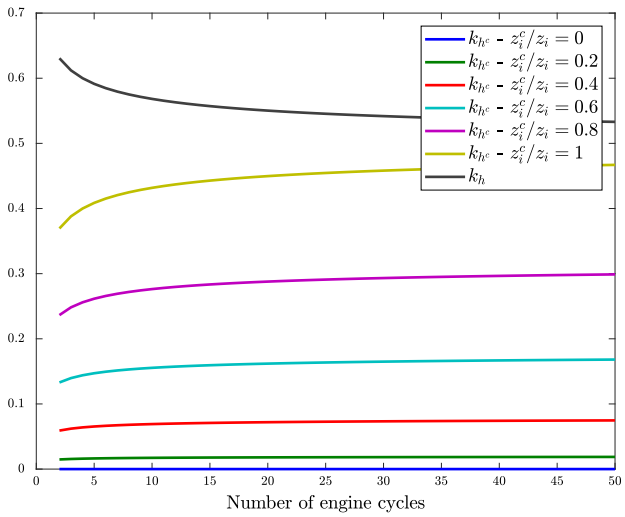

Figure 11: Sensitivity of k_h and k_{h^c} to the observation number I_{obs} for values of the reference random part (ratios $\frac{z_i^c}{z_i}$).

4 Experimental results on simulated signals

In this section, the algorithm explained in section 3.2 has been applied to signals simulating an engine outputs. The aim is to verify the algorithm and compare it to the Wiener filter separation. As a consequence, the application to a global noise whose components are known is compulsory. In order to do so, synthesized signals which simulate the references and the total noise are obtained through the application of Eqs. (3) and (4) to known latent variables $\mathbf{z}$, spectrofilters $\mathbf{h}$ and matrix $\mathbf{G}$. These, jointly with their respective variances, are shaped after the results of the algorithm being previously applied to actual engine signals. The data d_i and $\mathbf{r}_i$ are composed of the Fourier coefficients of respectively the global noise and the measured references corresponding to one single frequency and to the $i = 1, ..., I_{obs}$ observations. In what follows, $I_{obs} = 400$ engine cycles, the measured references are $N_r = 8$ and the latent ones are supposed to be $N_z = 3$. The measured

references have been chosen so that matrix $\mathbf{G}$ has the following form:

$$\mathbf{G} = \begin{bmatrix} G_{11} & 0 & 0 \\ 0 & G_{22} & 0 \\ G_{31} & G_{32} & 0 \\ G_{41} & G_{42} & 0 \\ G_{51} & G_{52} & 0 \\ G_{61} & 0 & G_{63} \\ G_{71} & 0 & G_{73} \\ G_{81} & 0 & G_{83} \end{bmatrix}. \tag{13}$$

Each measured reference, i.e. each term in the vector $\mathbf{r}_i$ has a dedicated variance $\sigma^2_{r_j}$ whose prior value is fixed considering the nature of the measured quantity (e.g. microphone or accelerometer).
The results presented in what follows concern the convergence of the algorithm and its comparison to the separation based on the Wiener filter method.
Figure 12 shows the convergence of σ^2_d over 10^4 iterations of the Markov chain. The convergence is rather fast and the zoom in Fig. 13 shows how the chains, whose initialization point is not the same, converge toward the same distribution.

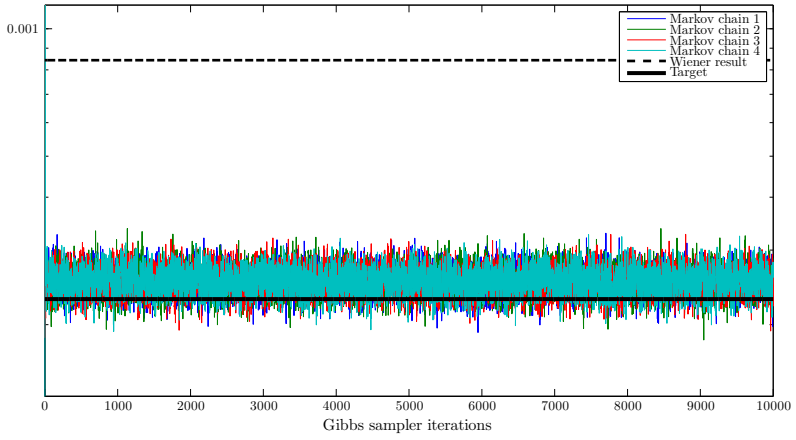

Figure 12: Convergence of a Markov chain toward the distribution describing σ^2_d.

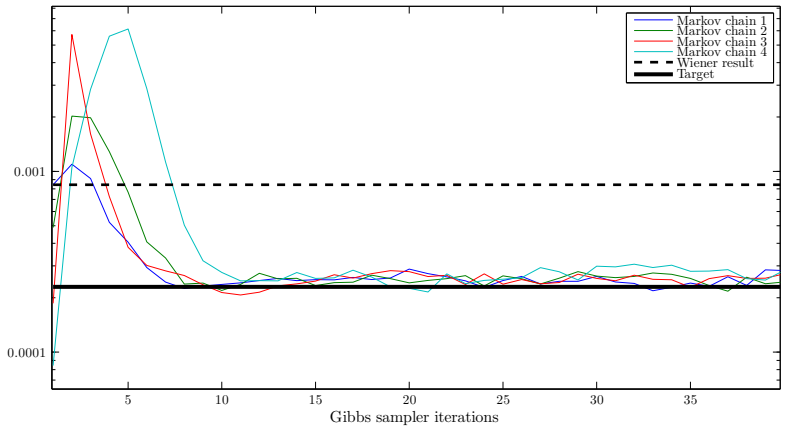

Figure 13: Convergence of a Markov chain toward the distribution describing σ_d^2 (zoom on the burn-in phase).

Given a parameter, the initialization points for the chains are scattered around the result returned by the classical Wiener filter: one initialization point corresponds to the "Wiener" solution, another to a tenth of its value and the remaining chains are initialized randomly drawing from a standard uniform distribution defined between 0 and twice the Wiener estimation of the parameter. The number of chains used is $N_c = 4$ and the Gelman & Rubin convergence indicator R is equal to 1.0001. Knowing that the asymptotic value of this indicator for the number of iterations which tends to infinite is 1 and knowing that values inferior to 1.1 can be considered acceptable, it can be concluded that the chains have well converged.
Figure 14 shows the convergence of the three contributions separated from the global engine noise. As it can be seen, the convergence is not toward a value but rather toward a distribution, with its mean and variance.

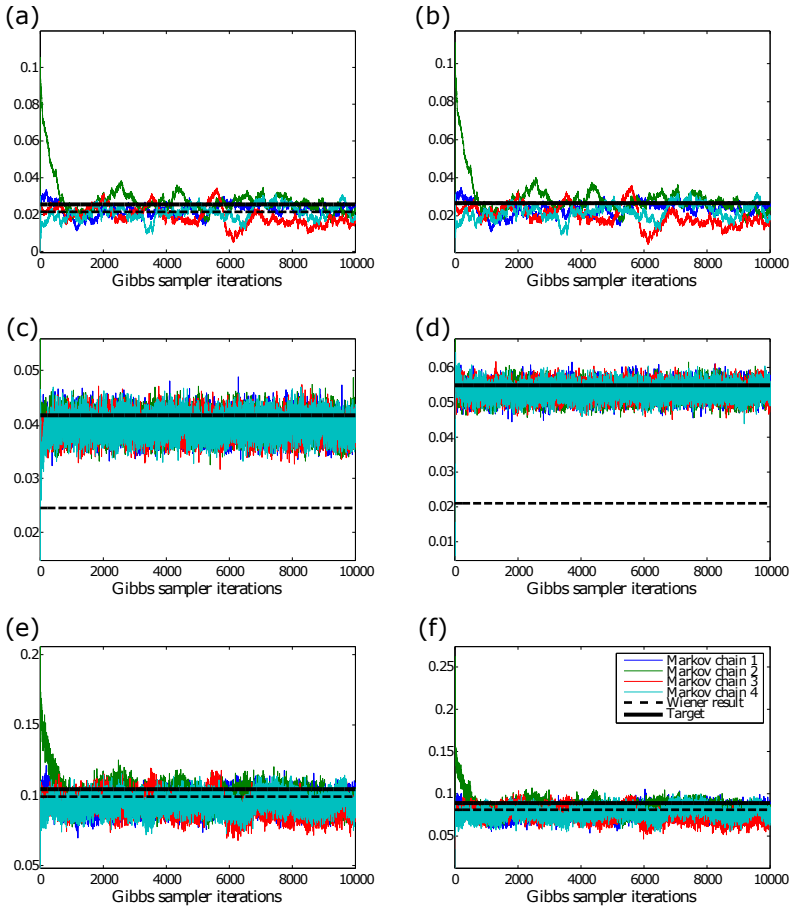

Figure 14: Convergence of a Markov chain toward the distribution of the contributions $\hat{x}_{ki} = h_k z_{ki}, \forall k = 1, ..., N_z$ and $i = \{200, 400\}$. The convergence of the first, second and third source are shown respectively in (a-b), (c-d), (e-f). This for their values at the engine cycles $i = 200$ (a-c-e) and $i = 400$ (b-d-f). In each figure, the bold solid line is the target result, the bold dashed line is the result obtained by Wiener filter h^c and the thin coloured lines are four Markov chains.

Figure 15 makes the comparison between the relative error of the Wiener filter, based on a mean-square optimization, and the one of the present method. For each observation i, the mean of the Markov chains from the Gibbs sampling and the value result of the Wiener filtering are normalized by the theoretical contribution corresponding to the measure d_i. It is interesting to remark that the algorithm does

not exactly converge toward the theoretical value. This is expected, since the measure d_i is itself just a realization of a stochastic process, so there is no reason for it to correspond to the mean of the Markov chain, whose distribution describes entirely the stochastic process. However, the 95% credible interval allows to verify that this measure is usually included in it.
In order to compare the results of the proposed method to the ones presented in [1], an indicator e_k is introduced. For a contribution, it is obtained as the sum along the engine cycles of the absolute value of the differences between the known contribution and its estimation:

$$e_k = \frac{\sum_{i=1}^{I_{obs}} |x_{ki} - h_k z_{ki}|}{\sum_{i=1}^{I_{obs}} |x_{ki}|}, \forall k = 1, ..., N_z, \tag{14}$$

where x_{ki} is the known k^{th} contribution for the engine cycle i. It is important to note that this indicator might be frequency dependent according to the frequency range studied. In this section, one single Fourier coefficient has been synthesized (i.e. one single frequency analysed), so that e_k is a scalar value.
Table 1 shows the result. As it can be seen, the error of the proposed method is less than the one obtained through the application of the Wiener filter h^c. This is expected since the presence of correlated references implies an additional waterfall step in the Wiener filter strategy: this step, as explained in paragraph 2.1.1, may remove from the references information useful for the correct identification of a source. On the other hand, the proposed approach considers the possible correlation between the measured references directly in the formulation and extracts from them some uncorrelated latent references in a non-supervised way. These latent references are then used for the source separation.
Moreover, as Fig. 15 shows, the result from the present method can be always followed by the 95% credible interval which conveys some information on its dispersion.

source contribution	Wiener	Gibbs
1	0.300	0.227
2	0.624	0.077
3	0.106	0.082

Table 1: Comparison between Wiener and Gibbs errors.

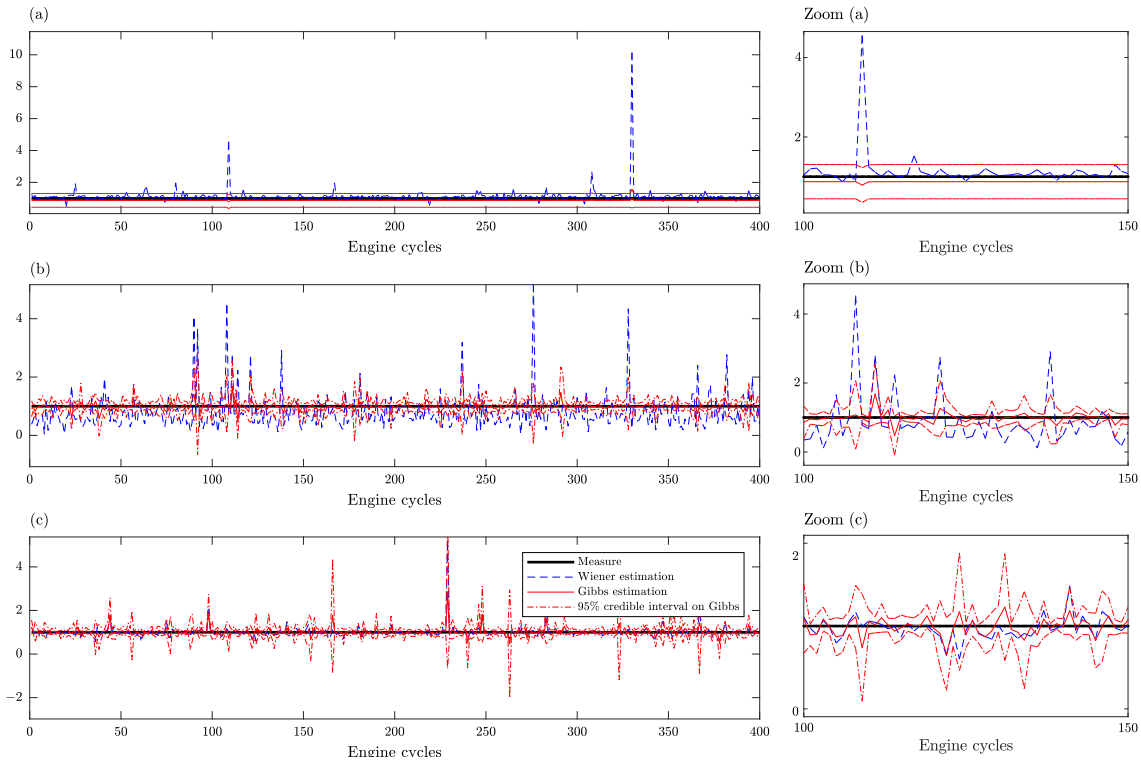

Figure 15: Estimation error of the algorithms $\forall i = 1, ..., I_{obs}$. The first source is in (a), the second in (b) and the third in (c). The bold black line is the reference (the measure); in dashed blue is the error of the Wiener filter h^c; in solid red is the error of the proposed algorithm, whose result is followed also by the credibility interval (dashed red). Zoom between the 100^{th} and 150^{th} engine cycles in the figures at the right.

5 Experimental results on measured signals

In this section the proposed algorithm is applied to measured signals issued from a diesel engine in order to extract the contributions to the sound pressure of the combustion, the injection and high pressure pump. The presented results concern the contributions of one of the cylinders of the engine, being the signals windowed around the Top

Dead Center (TDC)[2] of the considered cylinder. The number of observed engine cycles is $I_{obs} = 201$ and, after having applied the Fourier Transform, the evolution of the Fourier coefficients for one frequency is observed along the cycles. Therefore:

- d_i is the evolution along engine cycles of one Fourier coefficient of the overall engine noise measured by a microphone;
- among the measured references $\mathbf{r}_i$, the first one ($i = 1$) derives from the cylinder pressure, the second one ($i = 2$) from an electric information on the injectors and the others ($i = 3, \ldots, 8$) from accelerometers placed on the engine (Fig. 16). In particular, the combustion is considered as the dominant source and as such it is "visible" in all the accelerometer signals. Moreover, each accelerometer is sensitive to the mechanical part on which it is placed. As a consequence, the accelerometers placed on the injectors ($i = 3, \ldots, 5$) are sensitive to the combustion and the injection, whereas the ones placed on the high pressure pump ($i = 6, \ldots, 8$) are sensitive to the combustion and the pump movement. Thus, the assumed topology of the matrix $\mathbf{G}$ is:

$$\mathbf{G} = \begin{bmatrix} G_{11} & 0 & 0 \\ 0 & G_{22} & 0 \\ G_{31} & G_{32} & 0 \\ G_{41} & G_{42} & 0 \\ G_{51} & G_{52} & 0 \\ G_{61} & 0 & G_{63} \\ G_{71} & 0 & G_{73} \\ G_{81} & 0 & G_{83} \end{bmatrix}; \tag{15}$$

- such a topology yields to $\mathbf{z}_i$ ($i = 1, 2, 3$) components being virtual pure references for the combustion, the injection and the high pressure pump, respectively.

[2] In an internal combustion engine, the TDC denotes the position of a piston in which it is furthest from the crankshaft.

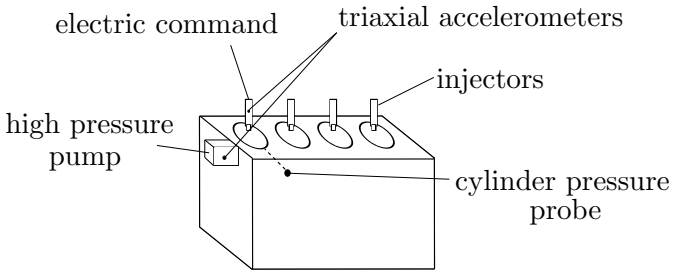

Figure 16: Sketch of the measurement protocol for one cylinder.

As it has been done for the synthesized signals, each measured reference has been assigned a different variance $\sigma^2_{r_i}$ whose prior value is chosen taking into consideration the physical nature of the signal (heteroskedastic mode).
The convergence is again rather fast. It is checked through the Gelman & Rubin method using four parallel Markov chains. For each random variable, the initial values of the chains are randomly scattered around the results from a classical Wiener filter separation realized beforehand. The convergence indicator after 10^4 iterations is $R = 1.0001$, therefore the chains can be considered to have converged.

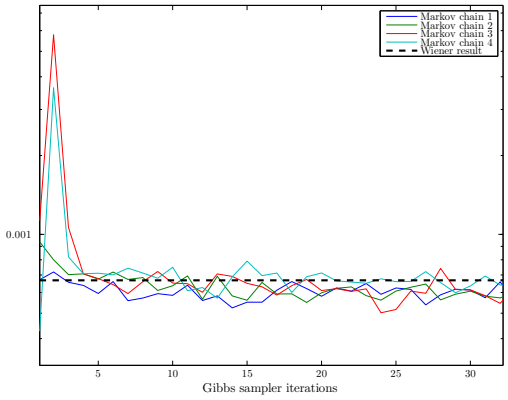

Figure 17: Convergence of a Markov chain toward the distribution describing σ^2_d (zoom on the burn-in phase).

As opposed to the application on synthesized signals (section 4), in this case the separation results cannot be compared to a reference solution, but only compared to the Wiener filtering results. What is

most interesting to analyse is the combustion contribution. Its reference being very pure, the Wiener filtering is expected to yield a good estimation of the contribution. The Gibbs extracted contribution is seen to converge towards it (Fig. 18(a)). What is more, the purity of the reference is clearly visible in the quality of the estimation, as hinted by the narrow 95% credible interval. As opposed to the cylinder pressure, the other references are more polluted and this is testified by the larger credibility intervals of the estimations (Fig. 18(b-c)). The proposed algorithm has been thought for just such cases: combining more polluted references to find a virtually pure reference and proposing an estimator jointly with its credible interval giving some information on the accuracy of the task achieved.

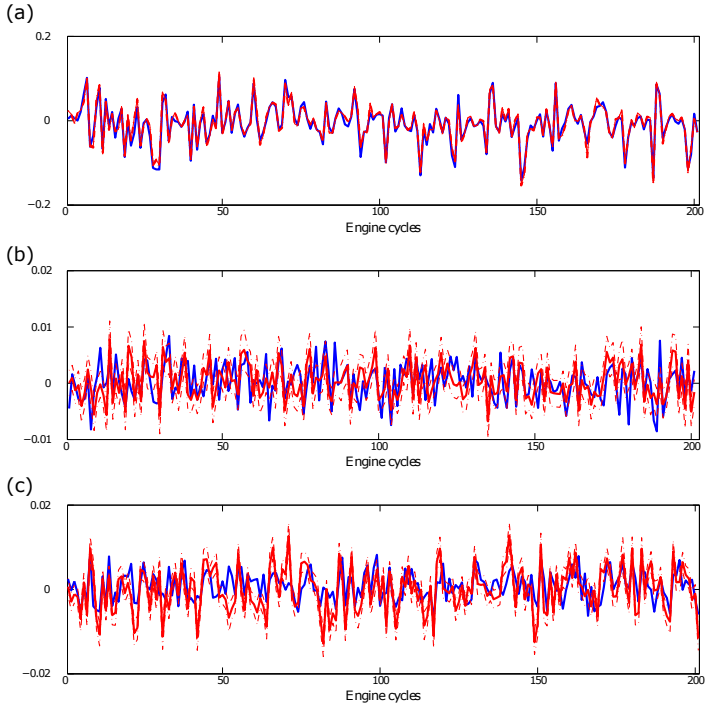

Figure 18: Comparison between the Wiener estimation based on h^c (blue lines) and the Bayesian estimation (red solid lines) of the contributions $h_k z_{ki}, \forall k = 1, ..., N_z$. (a), (b) and (c) show respectively the first, second and third contributions. The Bayesian result is followed by the 95% credibility interval (red dashed lines).

From what has been said in the previous paragraph it follows that the Gibbs combustion extraction yields similar results to the Wiener

extraction. After having applied the same algorithm to the whole frequency range of the measured signals, it is interesting to note whether the Gibbs filter has converged toward h or toward h^c (see discussion in subsection 3.3). It can be remarked that the Gibbs result converges generally toward the filter h^c (Fig. 19). According to Leclère [21] and as explained in section 2.1.1, indeed the h^c Wiener filter boasts a smaller bias error than the h one, but it is really more convenient only if the reference random part is important enough to yield an acceptable SNR and thus a small random error. In the present application it seems to be the case from $f \simeq 0.046$ onward. On the other hand, in the low frequency range the random part of the reference is too weak and the estimation error of the h^c filter is too important for it to be acceptable. It can be seen that, in such cases the Gibbs estimator converges toward the Wiener filter h (Fig. 20). The credibility intervals prove that the two filters are significantly different.

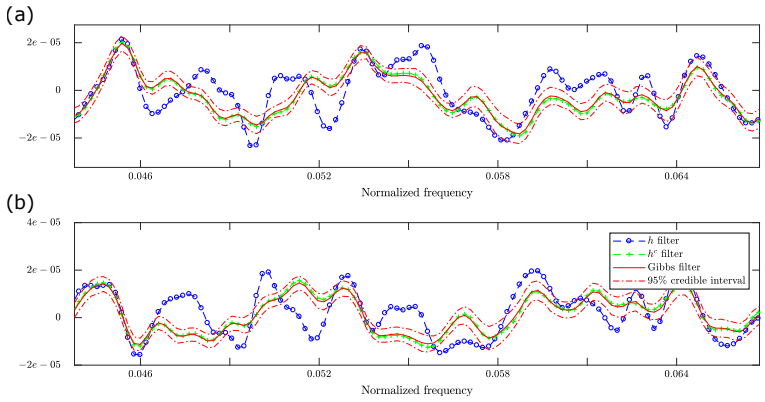

Figure 19: Convergence of the Gibbs combustion filter (red solid line) towards the h^c Wiener filter (green dashed line - plus markers) in a high frequency interval, for the real part (a) and the imaginary part (b). It is shown that the h Wiener filter (blue dashed line - round markers) is out of the 95% credibility interval of the Gibbs filter (red dashed lines). The frequency scale is normalized by the sampling frequency.

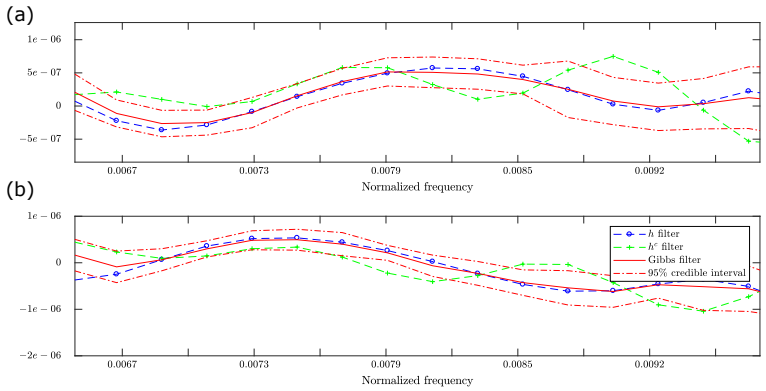

Figure 20: Convergence of the Gibbs combustion filter (red solid line) towards h Wiener filter (blue dashed line - round markers) in a low frequency interval, for the real part (a) and the imaginary part (b). It is shown that the h^c Wiener filter (green dashed line - plus markers) is generally out of the 95% credibility interval of the Gibbs filter (red dashed lines). The frequency scale is normalized by the sampling frequency.

6 Conclusion

A hierarchical Bayesian separation method has been proposed and validated on synthesized and measured signals. The main advantages of this approach lay in (1) achieving the separation task when the available references are correlated and noisy, (2) using a Bayesian framework which yields to the consideration of the prior knowledge, (3) accompanying the results with credible intervals which testify for the quality of the results, (4) allowing a non-supervised choice of the filter as a function of the signals characteristics.

Using this approach, some points need particular care. First of all, the more engine cycles are observed, the better the results will be, since the influence of the prior laws will be reduced. However, this yields to long signal acquisitions during which the cyclostationarity and cycloergodicity hypotheses have to be verified all along. In the engine case, this means that one stabilized operating condition is analysed in a run of the algorithm.

Moreover, the a priori given shape of the matrix $\mathbf{G}$ imposes constraints on the measured references. Here also, expertise and prior knowledge are compulsory in order to safely assume the source contribution in each reference.

Finally, depending on the analysed signals, the posterior probability distributions could be multi-modal. In this case, more advanced sam-

pling techniques could be used in order to improve the chain mixing properties [17].
The proposed approach proved itself worthy of interest for the source separation in the particularly difficult case of IC engines, which are characterized by multiple sources overlapped in the time and the frequency domains. Nevertheless, the approach is general and it can be used in all application where correlated references are available.

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).

A Usual probability density functions

Multivariate real normal distribution

The multivariate normal distribution of a N-dimensional random vector $\mathbf{x} \in \mathbb{R}^{N\times 1}$ of mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is noted:

$$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{A.1}$$

In the non-degenerate case (i.e. $\boldsymbol{\Sigma}$ is definite positive), the PDF of the multivariate real normal law is written:

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

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

Inverse-Gamma distribution

The Inverse-Gamma distribution can be parametrized using the shape parameter α and the rate parameter β. The Inverse-Gamma distribution of a random variable x is noted:

$$\text{InvGamma}(\alpha, \beta) \tag{A.3}$$

The corresponding PDF is written as:

$$[x|\alpha, \beta] = \text{InvGamma}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}} \tag{A.4}$$

Multivariate complex normal distribution

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

$$[\mathbf{x}|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}] = \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) = \frac{1}{\pi^n \det \boldsymbol{\Sigma}_{xx}} \exp\left\{-(\mathbf{x}-\boldsymbol{\mu}_x)^H \boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)\right\} \tag{A.5}$$

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

B Posterior laws calculation

In what follows, the details about the computation of the Bayesian posterior distributions are proposed.
The basis of the posterior computation is the Bayes theorem as presented in (7) and (8). The posteriors are obtained as follows:

1. Posterior law of $\mathbf{z}_i$, $\forall i = 1, \ldots, I_{obs}$ is:

$$\begin{aligned}
\text{Likelihood of } \mathbf{r}_i \quad & [\mathbf{r}_i|\mathbf{G}, \mathbf{z}_i, \sigma^2_{r_1} \ldots \sigma^2_{r_{N_r}}] \sim \mathcal{N}_{\mathbf{r}_i}(\mathbf{G}\mathbf{z}_i, \text{diag}(\sigma^2_{r_1}, \ldots, \sigma^2_{r_{N_r}})) \\
\text{Likelihood of } d_i \quad & [d_i|\mathbf{h}, \mathbf{z}_i, d_0, \sigma_d^2] \sim \mathcal{N}_{d_i}(\mathbf{h}^t\mathbf{z}_i + d_0, \sigma_d^2) \\
\text{Prior :} \quad & [\mathbf{z}_i] \sim \mathcal{N}_{\mathbf{z}_i}(\mathbf{0}, \mathbb{I})
\end{aligned} \tag{B.1}$$

$$\begin{aligned}
[\mathbf{z}_i|\text{rest}] &\propto [\mathbf{r}_i|\mathbf{G}, \mathbf{z}_i, \sigma^2_{r_1} \ldots \sigma^2_{r_{N_r}}][d_i|\mathbf{h}, \mathbf{z}_i, d_0, \sigma_d^2][\mathbf{z}_i] \\
&\propto \mathcal{N}_{\mathbf{r}_i}(\mathbf{G}\mathbf{z}_i, \text{diag}(\sigma^2_{r_1} \ldots \sigma^2_{r_{N_r}})) \cdot \mathcal{N}_{d_i}(\mathbf{h}^t\mathbf{z}_i + d_0, \sigma_d^2) \cdot \mathcal{N}_{\mathbf{z}_i}(\mathbf{0}, \mathbb{I}) \\
&\propto \exp\left(-(\mathbf{r}_i - \mathbf{G}\mathbf{z}_i)^H \text{diag}(\frac{1}{\sigma_{r_{1,\ldots,N_r}}})(\mathbf{r}_i - \mathbf{G}\mathbf{z}_i)\right) \cdot \\
&\exp\left(-\frac{1}{\sigma_d^2}((d_i - d_0) - \mathbf{h}^t\mathbf{z}_i)^H((d_i - d_0) - \mathbf{h}^t\mathbf{z}_i)\right) \cdot \exp\left(-\mathbf{z}_i^H\mathbf{z}_i\right) \\
&\propto \exp\left(-\mathbf{z}_i^H\left(\frac{\mathbf{h}^*\mathbf{h}^t}{\sigma_d^2} + \mathbf{G}^H\text{diag}(\frac{1}{\sigma_{r_{1,\ldots,N_r}}})\mathbf{G} + \mathbb{I}\right)\mathbf{z}_i\right) \cdot \\
&\exp\left(\mathbf{z}_i^H(\frac{\mathbf{h}^*}{\sigma_d^2}(d_i - d_0) + \mathbf{G}^H\text{diag}(\frac{1}{\sigma_{r_{1,\ldots,N_r}}})\mathbf{r}_i) + ((d_i - d_0)\frac{\mathbf{h}^t}{\sigma_d^2} + \mathbf{r}_i\text{diag}(\frac{1}{\sigma_{r_{1,\ldots,N_r}}})\mathbf{G})\right)
\end{aligned} \tag{B.2}$$

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

$$[\mathbf{z}|\mathbf{m}, \mathbf{S}] \propto \exp\left(-\mathbf{z}^H\mathbf{S}^{-1}\mathbf{z} + \mathbf{z}^H\mathbf{S}^{-1}\mathbf{m} + \mathbf{m}^H\mathbf{S}^{-1}\mathbf{z}\right) \tag{B.3}$$

Equations (B.2) and (B.3) have the same form and by identification it can be recognised that the posterior distribution of $\mathbf{z}_n$ is

a Multivariate Gaussian distribution with parameters:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{z}_i} &= \boldsymbol{\Omega}_{\mathbf{z}_i}(\mathbf{h}^*(d_i - d_0)/\sigma_d^2 + \mathbf{G}^H \begin{bmatrix} \sigma_{r_1}^{-2} & 0 & 0 \\ & \ddots & \\ 0 & 0 & \sigma_{r_{N_r}}^{-2} \end{bmatrix} \mathbf{r}_i) \\ \boldsymbol{\Omega}_{\mathbf{z}_i} &= (\mathbf{h}^*\mathbf{h}^t/\sigma_d^2 + \mathbf{G}^H \begin{bmatrix} \sigma_{r_1}^{-2} & 0 & 0 \\ & \ddots & \\ 0 & 0 & \sigma_{r_{N_r}}^{-2} \end{bmatrix} \mathbf{G} + \mathbb{I})^{-1}.\end{aligned} \tag{B.4}$$

2. Posterior law of σ_d^2:

$$\begin{aligned}\text{Likelihood}: \quad & [d_i|\mathbf{h}, \mathbf{z}_i, d_0, \sigma_d^2] \sim \mathcal{N}_{d_i}(\mathbf{h}^t\mathbf{z}_i + d_0, \sigma_d^2) \\ \text{Prior}: \quad & [\sigma_d^2] \sim \text{InvGamma}(a_d, b_d)\end{aligned} \tag{B.5}$$

$$\begin{aligned}[\sigma_d^2|\text{rest}] &\propto \prod_{i=1}^{I_{obs}} [d_i|\mathbf{h}, \mathbf{z}_i, d_0, \sigma_d^2][\sigma_d^2] \\ &\propto \prod_{i=1}^{I_{obs}} \mathcal{N}_{d_i}(\mathbf{h}^t\mathbf{z}_i + d_0, \sigma_d^2) \cdot \text{InvGamma}(a_d, b_d) \\ &\propto \frac{1}{(\sigma_d^2)^{I_obs}} \exp\left(-\frac{1}{\sigma_d^2}\sum_{i=1}^{I_{obs}} |d_i - d_0 - \mathbf{h}^t\mathbf{z}_i|^2\right) \frac{1}{(\sigma_d^2)^{a_d+1}} \exp\left(-\frac{b_d}{\sigma_d^2}\right) \\ &\propto \frac{1}{(\sigma_d^2)^{I_{obs}+a_d+1}} \exp\left(-\frac{1}{\sigma_d^2}(\sum_{i=1}^{I_{obs}} |d_i - d_0 - \mathbf{h}^t\mathbf{z}_i|^2 + b_d)\right)\end{aligned} \tag{B.6}$$

This corresponds to an Inverse-Gamma distribution with parameters $a'_d = I_{obs} + a_d$ and $b'_d = \sum_{i=1}^{I_{obs}} |d_i - d_0 - \mathbf{h}^t\mathbf{z}_i|^2 + b_d$.

3. Posterior law of $\sigma_{r_j}^2$, $\forall j = 1, \ldots, N_r$:

$$\begin{aligned}\text{Likelihood}: \quad & [r_{ji}|\mathbf{g}_j, \mathbf{z}_i, \sigma_{r_j}^2] \sim \mathcal{N}_{r_{ji}}(\mathbf{g}_j\mathbf{z}_i, \sigma_{r_j}^2) \\ \text{Prior}: \quad & [\sigma_{r_j}^2] \sim \text{InvGamma}(a_r, b_r)\end{aligned} \tag{B.7}$$

$$\begin{aligned}[\sigma_{r_j}^2|\text{rest}] &\propto \prod_{i=1}^{I_{obs}} [r_{ji}|\mathbf{g}_j, \mathbf{z}_i, \sigma_{r_j}^2][\sigma_{r_j}^2] \\ &\propto \prod_{i=1}^{I_{obs}} \mathcal{N}_{r_{ji}}(\mathbf{g}_j\mathbf{z}_i, \sigma_{r_j}^2) \cdot \text{InvGamma}(a_r, b_r) \\ &\propto \frac{1}{(\sigma_{r_j}^2)^{I_obs}} \exp\left(-\frac{1}{\sigma_{r_j}^2}\sum_{i=1}^{I_{obs}} |r_{ji} - \mathbf{g}_j\mathbf{z}_i|^2\right) \frac{1}{(\sigma_{r_j}^2)^{a_r+1}} \exp\left(-\frac{b_r}{\sigma_{r_j}^2}\right) \\ &\propto \frac{1}{(\sigma_{r_j}^2)^{I_{obs}+a_r+1}} \exp\left(-\frac{1}{\sigma_{r_j}^2}(\sum_{i=1}^{I_{obs}} |r_{ji} - \mathbf{g}_j\mathbf{z}_i|^2 + b_r)\right)\end{aligned} \tag{B.8}$$

This corresponds to an Inverse-Gamma distribution with parameters $a'_r = I_{obs} + a_r$ and $b'_r = \sum_{i=1}^{I_{obs}} |r_{ji} - \mathbf{g}_j \mathbf{z}_i|^2 + b_r$.

4. Posterior law of $\sigma^2_{h_k}$, $\forall k = 1, \ldots, N_z$:

$$\begin{aligned} \text{Likelihood}: \quad & [h_k|\sigma^2_{h_k}] \sim \mathcal{N}_{h_k}(0, \sigma^2_{h_k}) \\ \text{Prior}: \quad & [\sigma^2_{h_k}] \sim \text{InvGamma}(a_{h_k}, b_{h_k}) \end{aligned} \tag{B.9}$$

$$\begin{aligned} [\sigma^2_{h_k}|\text{rest}] &\propto [h_k|\sigma^2_{h_k}][\sigma^2_{h_k}] \\ &\propto \mathcal{N}_{h_k}(0, \sigma^2_{h_k}) \cdot \text{InvGamma}(a_{h_k}, b_{h_k}) \\ &\propto \frac{1}{\sigma^2_{h_k}} \exp\left(-\frac{1}{\sigma^2_{h_k}}|h_k|^2\right) \frac{1}{(\sigma^2_{h_k})^{a_{h_k}+1}} \exp\left(-\frac{b_{h_k}}{\sigma^2_{h_k}}\right) \\ &\propto \frac{1}{(\sigma^2_{h_k})^{1+a_{h_k}+1}} \exp\left(-\frac{1}{\sigma^2_{h_k}}(|h_k|^2 + b_{h_k})\right) \end{aligned} \tag{B.10}$$

This corresponds to an Inverse-Gamma distribution with parameters $a'_{h_k} = a_{h_k} + 1$ and $b'_{h_k} = |h_k|^2 + b_{h_k}$.

5. Posterior law of $\sigma^2_{G_{jk}}$, $\forall j = 1, \ldots, N_r \quad \forall k = 1, \ldots, N_z$:

$$\begin{aligned} \text{Likelihood}: \quad & [G_{jk}|\sigma^2_{G_{jk}}] \sim \mathcal{N}_{G_{jk}}(0, \sigma^2_{G_{jk}}) \\ \text{Prior}: \quad & [\sigma^2_{G_{jk}}] \sim \text{InvGamma}(a_{G_{jk}}, b_{G_{jk}}) \end{aligned} \tag{B.11}$$

$$\begin{aligned} [\sigma^2_{G_{jk}}|\text{rest}] &\propto [G_{jk}|\sigma^2_{G_{jk}}][\sigma^2_{G_{jk}}] \\ &\propto \mathcal{N}_{G_{jk}}(0, \sigma^2_{G_{jk}}) \cdot \text{InvGamma}(a_{G_{jk}}, b_{G_{jk}}) \\ &\propto \frac{1}{\sigma^2_{G_{jk}}} \exp\left(-\frac{1}{\sigma^2_{G_{jk}}}|G_{jk}|^2\right) \frac{1}{(\sigma^2_{G_{jk}})^{a_{G_{jk}}+1}} \exp\left(-\frac{b_{G_{jk}}}{\sigma^2_{G_{jk}}}\right) \\ &\propto \frac{1}{(\sigma^2_{G_{jk}})^{1+a_{G_{jk}}+1}} \exp\left(-\frac{1}{\sigma^2_{G_{jk}}}(|G_{jk}|^2 + b_{G_{jk}})\right) \end{aligned} \tag{B.12}$$

This corresponds to an Inverse-Gamma distribution with parameters $a'_{G_{jk}} = a_{G_{jk}} + 1$ and $b'_{G_{jk}} = |G_{jk}|^2 + b_{G_{jk}}$.

6. Posterior law of $\mathbf{h}^t$ is:

$$\begin{aligned} \text{Likelihood}: \quad & [d_i|\mathbf{h}, \mathbf{z}_i, d_0, \sigma^2_d] \sim \mathcal{N}_{d_i}(\mathbf{h}^t \mathbf{z}_i + d_0, \sigma^2_d) \\ \text{Prior}: \quad & [\mathbf{h}^t] \sim \mathcal{N}_{\mathbf{h}^t}(\mathbf{0}, \text{diag}(\sigma^2_{h_1}, \ldots, \sigma^2_{h_{N_z}})) \end{aligned} \tag{B.13}$$

$$
\begin{aligned}
[\mathbf{h}^t|\text{rest}] &\propto \prod_{i=1}^{I_{obs}} [d_i|\mathbf{h}, \mathbf{z}_i, d_0, \sigma_d^2][\mathbf{h}^t] \\
&\propto \prod_{i=1}^{I_{obs}} \mathcal{N}_{d_i}(\mathbf{h}^t\mathbf{z}_i + d_0, \sigma_d^2) \cdot \mathcal{N}_{\mathbf{h}^t}(\mathbf{0}, \text{diag}(\sigma_{h_1}^2, \ldots, \sigma_{h_{N_z}}^2)) \\
&\propto \exp\left(-\frac{1}{\sigma_d^2}\sum_{i=1}^{I_{obs}}((d_i - d_0) - \mathbf{h}^t\mathbf{z}_i)((d_i - d_0) - \mathbf{h}^t\mathbf{z}_i)^H\right) \cdot \\
&\exp\left(-\mathbf{h}^t \text{diag}(\frac{1}{\sigma_{h_1}^2}, \ldots, \frac{1}{\sigma_{h_{N_z}}^2})(\mathbf{h}^t)^H\right) \\
&\propto \exp\left(-\mathbf{h}^t\left(\sum_{i=1}^{I_{obs}} \frac{\mathbf{z}_i\mathbf{z}_i^H}{\sigma_d^2} + \text{diag}(\frac{1}{\sigma_{h_1}^2}, \ldots, \frac{1}{\sigma_{h_{N_z}}^2})\right)(\mathbf{h}^t)^H\right) \cdot \\
&\exp\left(\frac{1}{\sigma_d^2}\sum_{i=1}^{I_{obs}}(d_i - d_0)\mathbf{z}_i^H(\mathbf{h}^t)^H + \frac{\mathbf{h}^t}{\sigma_d^2}\sum_{i=1}^{I_{obs}}\mathbf{z}_i(d_i - d_0)^*\right)
\end{aligned}
\tag{B.14}
$$

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

$$
[\mathbf{h}^t|\mathbf{m}, \mathbf{S}] \propto \exp\left(-\mathbf{h}^t\mathbf{S}^{-1}(\mathbf{h}^t)^H + \mathbf{h}^t\mathbf{S}^{-1}\mathbf{m}^H + \mathbf{m}\mathbf{S}^{-1}(\mathbf{h}^t)^H\right) \tag{B.15}
$$

Equations (B.14) and (B.15) have the same form and by identification it can be recognised that the posterior distribution of $\mathbf{h}^t$ is a Multivariate Gaussian distribution with parameters:

$$
\begin{aligned}
\boldsymbol{\mu}_{\mathbf{h}} &= \sum_{i=1}^{I_{obs}}(d_i - d_0)\mathbf{z}_i^H\boldsymbol{\Omega}_{\mathbf{h}}/\sigma_d^2 \\
\boldsymbol{\Omega}_{\mathbf{h}} &= (\sum_{i=1}^{I_{obs}}\mathbf{z}_i\mathbf{z}_i^H/\sigma_d^2 + \begin{bmatrix} \sigma_{h_1}^{-2} & 0 & 0 \\ & \ddots & \\ 0 & 0 & \sigma_{h_{N_z}}^{-2} \end{bmatrix})^{-1}.
\end{aligned}
\tag{B.16}
$$

7. Posterior law of $\mathbf{g}_j$, $\forall j = 1, \ldots, N_r$ is:

$$
\begin{aligned}
\text{Likelihood}: &\quad [r_{ji}|\mathbf{g}_j, \mathbf{z}_i, \sigma_{r_j}^2] \sim \mathcal{N}_{r_{ji}}(\mathbf{g}_j\mathbf{z}_i, \sigma_{r_j}^2) \\
\text{Prior}: &\quad [\mathbf{g}_j] \sim \mathcal{N}_{\mathbf{g}_j}(\mathbf{0}, \text{diag}(\sigma_{G_{j1}}^2, \ldots, \sigma_{G_{jN_z}}^2))
\end{aligned}
\tag{B.17}
$$

$$
\begin{aligned}
[\mathbf{g}_j|\text{rest}] &\propto \prod_{i=1}^{I_{obs}} [r_{ji}|\mathbf{g}_j, \mathbf{z}_i, \sigma^2_{r_j}][\mathbf{g}_j] \\
&\propto \prod_{i=1}^{I_{obs}} \mathcal{N}_{r_{ji}}(\mathbf{g}_j\mathbf{z}_i, \sigma^2_{r_j}) \cdot \mathcal{N}_{\mathbf{g}_j}(\mathbf{0}, \text{diag}(\sigma^2_{G_{j1}}, \ldots, \sigma^2_{G_{jN_z}})) \\
&\propto \exp\left(-\frac{1}{\sigma^2_{r_j}} \sum_{i=1}^{I_{obs}} (r_j i - \mathbf{g}_j\mathbf{z}_i)(r_j i - \mathbf{g}_j\mathbf{z}_i)^H\right) \cdot \exp\left(-\mathbf{g}_j \text{diag}(\frac{1}{\sigma^2_{G_{j1}}}, \ldots, \frac{1}{\sigma^2_{G_{jN_z}}})\mathbf{g}_j^H\right) \\
&\propto \exp\left(-\mathbf{g}_j(\sum_{i=1}^{I_{obs}} \frac{\mathbf{z}_i\mathbf{z}_i^H}{\sigma^2_{r_j}} + \text{diag}(\frac{1}{\sigma^2_{G_{j1}}}, \ldots, \frac{1}{\sigma^2_{G_{jN_z}}}))\mathbf{g}_j^H + \frac{1}{\sigma^2_{r_j}} \sum_{i=1}^{I_{obs}} r_{ji}\mathbf{z}_i^H\mathbf{g}_j^H + \frac{\mathbf{g}_j}{\sigma^2_{r_j}} \sum_{i=1}^{I_{obs}} \mathbf{z}_i r^*_{ji}\right)
\end{aligned}
\tag{B.18}
$$

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

$$
[\mathbf{g}|\mathbf{m}, \mathbf{S}] \propto \exp\left(-\mathbf{g}\mathbf{S}^{-1}\mathbf{g}^H + \mathbf{g}\mathbf{S}^{-1}\mathbf{m}^H + \mathbf{m}\mathbf{S}^{-1}\mathbf{g}^H\right) \tag{B.19}
$$

Equations (B.18) and (B.19) have the same form and by identification it can be recognised that the posterior distribution of $\mathbf{g}_j$ is a Multivariate Gaussian distribution with parameters:

$$
\begin{aligned}
\boldsymbol{\mu}_{\mathbf{g}_j} &= \sum_{i=1}^{I_{obs}} r_{ji}\mathbf{z}_i^H \boldsymbol{\Omega}_{\mathbf{g}_j}/\sigma^2_{r_j} \\
\boldsymbol{\Omega}_{\mathbf{g}_j} &= (\sum_{i=1}^{I_{obs}} \mathbf{z}_i\mathbf{z}_i^H/\sigma^2_{r_j} + \begin{bmatrix} \sigma^{-2}_{G_{j1}} & 0 & 0 \\ & \ddots & \\ 0 & 0 & \sigma^{-2}_{G_{jN_z}} \end{bmatrix})^{-1}.
\end{aligned}
\tag{B.20}
$$

8. Posterior law of d_0 is:

$$
\begin{aligned}
\text{Likelihood}: &\quad [d_i|\mathbf{h}, \mathbf{z}_i, d_0, \sigma^2_d] \sim \mathcal{N}_{d_i}(\mathbf{h}^t\mathbf{z}_i + d_0, \sigma^2_d) \\
\text{Prior}: &\quad [d_0] \sim \mathcal{N}_{d_0}(0, \sigma_{d_0})
\end{aligned}
\tag{B.21}
$$

$$
\begin{aligned}
[d_0|\text{rest}] &\propto \prod_{i=1}^{I_{obs}} [d_i|\mathbf{h}, \mathbf{z}_i, d_0, \sigma^2_d][d_0] \\
&\propto \prod_{i=1}^{I_{obs}} \mathcal{N}_{d_i}(\mathbf{h}^t\mathbf{z}_i + d_0, \sigma^2_d) \cdot \mathcal{N}_{d_0}(0, \sigma_{d_0}) \\
&\propto \exp\left(-\frac{1}{\sigma^2_d} \sum_{i=1}^{I_{obs}} ((d_i - \mathbf{h}^t\mathbf{z}_i) - d_0)^H((d_i - \mathbf{h}^t\mathbf{z}_i) - d_0)\right) \cdot \exp\left(-\frac{|d_0|^2}{\sigma^2_{d_0}}\right) \\
&\propto \exp\left(-\left(\frac{I_{obs}}{\sigma^2_d} + \frac{1}{\sigma^2_{d_0}}\right)|d_0|^2 + \frac{d_0}{\sigma^2_d} \sum_{i=1}^{I_{obs}} (d_i - \mathbf{h}^t\mathbf{z}_i)^* + \frac{d_0^*}{\sigma^2_d} \sum_{i=1}^{I_{obs}} (d_i - \mathbf{h}^t\mathbf{z}_i)\right)
\end{aligned}
\tag{B.22}
$$

A Gaussian distribution for a random variable d_0 with mean m and variance s^2 would lead to:

$$[d_0|m, s^2] \propto \exp\left(-\frac{|d_0|^2}{s^2} + \frac{d_0^* m}{s^2} + + \frac{m^* d_0}{s^2}\right) \tag{B.23}$$

Equations (B.22) and (B.23) have the same form and by identification it can be recognised that the posterior distribution of d_0 is a Gaussian distribution with parameters:

$$\begin{aligned} \mu_{d_0} &= \nu_{d_0} \sum_{i=1}^{I_{obs}} (d_i - \mathbf{h}^t \mathbf{z}_i)/\sigma_d^2 \\ \nu_{d_0} &= (\sigma_{d_0}^{-2} + I_{obs}\sigma_d^{-2})^{-1}. \end{aligned} \tag{B.24}$$

9. Posterior law of $\sigma_{d_0}^2$:

$$\begin{aligned} \text{Likelihood}: \quad & [d_0|\sigma_{d_0}^2] \sim \mathcal{N}_{d_0}(0, \sigma_{d_0}^2) \\ \text{Prior}: \quad & [\sigma_{d_0}^2] \sim \text{InvGamma}(a_{d_0}, b_{d_0}) \end{aligned} \tag{B.25}$$

$$\begin{aligned} [\sigma_{d_0}^2|\text{rest}] &\propto [d_0|\sigma_{d_0}^2][\sigma_{d_0}^2] \\ &\propto \mathcal{N}_{d_0}(0, \sigma_{d_0}^2) \cdot \text{InvGamma}(a_{d_0}, b_{d_0}) \\ &\propto \frac{1}{\sigma_{d_0}^2} \exp\left(-\frac{1}{\sigma_{d_0}^2}|d_0|^2\right) \frac{1}{(\sigma_{d_0}^2)^{a_{d_0}+1}} \exp\left(-\frac{b_{d_0}}{\sigma_{d_0}^2}\right) \\ &\propto \frac{1}{(\sigma_{d_0}^2)^{1+a_{d_0}+1}} \exp\left(-\frac{1}{\sigma_{d_0}^2}(|d_0|^2 + b_{d_0})\right) \end{aligned} \tag{B.26}$$

This corresponds to an Inverse-Gamma distribution with parameters $a'_{d_0} = a_{d_0} + 1$ and $b'_{d_0} = |d_0|^2 + b_{d_0}$.

C Expressing the Gibbs filter as a combination of Wiener filters

For the sake of clarity the development corresponds to the case where N_z is equal to one. Moreover, in order to compare the Bayesian filter to the Wiener filters, the variances σ_{h_k} are reduced to zero to simulate the deterministic result characteristic of the Wiener approach. Under these hypothesis, the variance of the Gibbs estimate for the filter becomes

$$\omega_h = \frac{\sigma_d^2}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.1}$$

and the expected value of the filter can thus be written as:

$$\mu_h = \frac{\sum_{i=1}^{I_{obs}} (d_i - d_0) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.2}$$

where $\bullet^*$ stands for the complex conjugate. Some manipulation is necessary to make the Wiener filters appear. To this aim, their formulation is reminded here after:

$$h = \frac{\sum_{i=1}^{I_{obs}} d_i z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.3}$$

$$h^c = \frac{\sum_{i=1}^{I_{obs}} (d_i - \overline{d})(z_i - \overline{z})^*}{\sum_{i=1}^{I_{obs}} |z_i - \overline{z}|^2} = \frac{\sum_{i=1}^{I_{obs}} (d_i - \overline{d}) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i - \overline{z}|^2} \tag{C.4}$$

where the synchronous average (deterministic part) of the terms is noted $\overline{\bullet}$.
Equation (C.2) can then be manipulated as follows:

$$\mu_h = \frac{\sum_{i=1}^{I_{obs}} (d_i - d_0) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.5}$$

$$= \frac{\sum_{i=1}^{I_{obs}} ((d_i - \overline{d}) - d_0 + \overline{d}) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.6}$$

$$= \frac{\sum_{i=1}^{I_{obs}} ((\alpha + (1-\alpha))(d_i - \overline{d}) - d_0 + \overline{d}) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.7}$$

$$= \alpha \frac{\sum_{i=1}^{I_{obs}} (d_i - \overline{d}) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} + \frac{\sum_{i=1}^{I_{obs}} ((1-\alpha)(d_i - \overline{d}) - d_0 + \overline{d}) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.8}$$

$$= \alpha \frac{\sum_{i=1}^{I_{obs}} (d_i - \overline{d}) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} + (1-\alpha) \frac{\sum_{i=1}^{I_{obs}} d_i z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} + \frac{\sum_{i=1}^{I_{obs}} (\alpha \overline{d} - d_0) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.9}$$

$$= \alpha \frac{\sum_{i=1}^{I_{obs}} |z_i - \overline{z}|^2}{\sum_{i=1}^{I_{obs}} |z_i|^2} \frac{\sum_{i=1}^{I_{obs}} (d_i - \overline{d}) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i - \overline{z}|^2} + (1-\alpha) \frac{\sum_{i=1}^{I_{obs}} d_i z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} + \frac{\sum_{i=1}^{I_{obs}} (\alpha \overline{d} - d_0) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.10}$$

Noting $C = \frac{\sum_{i=1}^{I_{obs}} |z_i - \overline{z}|^2}{\sum_{i=1}^{I_{obs}} |z_i|^2}$, the Wiener filter h^c can be recognised in the first term, whereas the h filter in the second. Therefore, equation (C.2) can be rewritten as:

$$\mu_h = \alpha C h^c + (1-\alpha) h + \frac{\sum_{i=1}^{I_{obs}} (\alpha \overline{d} - d_0) z_i^*}{\sum_{i=1}^{I_{obs}} |z_i|^2} \tag{C.11}$$

Finally, the arbitrary value α can be chosen in order to eliminate the third term:

$$\begin{aligned}\alpha &= \frac{d_0}{\overline{d}} \\ &= \frac{d_0}{\frac{\sum_{i=1}^{I_{obs}} d_i}{I_{obs}}} \\ &= \frac{d_0}{\frac{\sum_{i=1}^{I_{obs}} (d_0+x_i+n_i)}{I_{obs}}} \\ &= \frac{d_0}{d_0 + \overline{x} + \epsilon\sigma_d/\sqrt{I_{obs}}}\end{aligned} \tag{C.12}$$

where x_i is the contribution of the source, $\overline{x}$ its deterministic part and the mean of the random noise $\sum_{i=1}^{I_{obs}} n_i$ follows a complex Gaussian distribution defined by $\epsilon\sigma_d/\sqrt{I_{obs}}$ with $\epsilon \sim \mathcal{N}_{\mathbb{C}}(0,1)$ being a standard circular complex Gaussian random variable.
This yields equations (10) and (11).

References

[1] L. Pruvost, Q. Leclère, E. Parizet, *Diesel engine combustion and mechanical noise separation using an improved spectrofilter*, Mechanical Systems and Signal Processing, Elsevier (2009), pp. 2072-2087.

[2] A. Hastings, *Sound quality of diesel engines*, Purdue Universities dissertations, Purdue University (2004).

[3] U. Groemping, R. Heinrichs, *Customer driven diesel vehicle sound quality*, Czech Acoustical Society, in *Proceedings of the 33rd Inter-Noise, Prague, Czech Republic, 2004 August 22-25*, Prague (2004).

[4] N. Alt, H. Sonntag, S. Heuer, R. Thiele, *Diesel engine cold start noise improvement*, S.A.E. Technical paper series (2005).

[5] C. Renard, L. Polac, J. Pascal, S. Sahraoui, *Extraction of vibration sources in diesel engines*, in *Proceedings of the 11th International Congress on Sound and Vibration, St. Petersbourg, Russia*, St. Petersbourg (2004).

[6] G. Lowet, P. V. Ponseele, S. Pauwels, T. V. Wayanberge, P. Sas, *Development of a metric to quantify diesel engines irregularities*, in *Proceedings of the 23rd International Seminar on Modal Analysis*, K.U. Leuven (1998), Belgium.

[7] O. Sauvage, A. Laurac, M. Bezat, V. Roussaire, P. Guillemain, *Diesel knock noise from combustion phenomenon to perceived signals*, in Société Française d'Acoustique, editor, *Acoustics 2012 proceedings, Nantes, France, 2012 April*, Nantes (2012).

[8] L. L. Sharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Prentice Hall (1991).

[9] L. Xianhua, R. B. Randall, J. Antoni, *Blind separation of internal combustion engine vibration signal by a deflation method*, Mechanical Systems and Signal Processing, Vol. 22, Elsevier (2008), pp. 1082-1091.

[10] I. Hirano, M. Kondo, Y. Uraki, Y. Asahara, *Using multiple regression analysis to estimate the contribution of engine-radiated noise components*, JSAE Review, Vol. 20, Society of Automotive Engineers of Japan and Elsevier (1999), pp. 363-368.

[11] J. Antoni, R. Boustany, F. Gautier, S. Wang, *Novel source separation method for measuring combustion noise and structural loss.*, in *6th European Conference on Noise Control (EURONOISE), Tampere, Finland, 2006 30 May-1 June*, Tampere (2006).

[12] M. El Badaoui, J. Danière, F. Guillet, C. Servière *Separation of combustion noise and piston-slap in diesel engine, part I: separation of combustion noise and piston-slap in diesel engine by cyclic Wiener filtering*, Mechanical Systems and Signal Processing, Vol. 19, No. 6, Elsevier (2005), pp. 1209-1217.

[13] J. Antoni, *Apports de l'échantillonnage angulaire et de la cyclostationnarité au diagnostic par analyse vibratoire des moteurs thérmiques*, PhD thesis, Institut National Polytechnique de Grenoble (2000).

[14] J. Antoni, *Cyclostationarity by examples*, Mechanical Systems and Signal Processing, Vol. 23, No. 4, Elsevier (2008), pp. 987-1036.

[15] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui *Cyclostationary modelling of rotating machine vibration signals*, Mechanical systems and signal processing, Vol. 18, No. 6, Elsevier (2004), pp.1285-1314.

[16] J. Antoni, N. Ducleaux, G. Nghiem, S. Wang, *Separation of combustion noise in IC engines under cyclo-non-stationary regime*, Mechanical Systems and Signal Processing, Vol. 1, No. 38, Elsevier (2013), pp. 223-236.

[17] Y. Ahmadian, J. W. Pillow, L. Paninski, *Efficient Markov Chain Monte Carlo Methods for Decoding Neural Spike Trains*, Neural Computation, Vol. 23, MIT Press Cambridge (2011), pp. 46-96.

[18] P. L. Green, *Bayesian system identification of dynamical systems using large sets of training data: A MCMC solution*, Probabilistic Engineering Mechanics, Vol. 42, Elsevier (2015), pp. 54-63.

[19] J. Huang, W. Hu, *MCMC-Particle-based group tracking of space objects within Bayesian framework*, Advances in Space Research, Vol. 53, Elsevier (2014), pp. 280-294.

[20] A. Gelman, D. B. Rubin, *Inference from iterative simulation using multiple sequences*, Statistical Science, Vol. 7, No. 4, Institute of Mathematical Statistics (1992), pp. 457-511.

[21] Q. Leclère, C. Sandier, O. Sauvage, *Source separation in diesel engines using Wiener filtering: physical interpretations*, in *Automotive NVH Comfort, Le Mans, France, 2014 October*, Le Mans (2014).

[22] S. Geman, D. Geman, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis ans Machine Intelligence, Vol. 6, IEEE (1998), pp. 721-741.

[23] W. M. Bolstad, *Understanding computational Bayesian statistics*, John Wiley & Sons (2010).