

# Gaussian Processes for Underdetermined Source Separation

Antoine Liutkus, Roland Badeau, *Senior Member, IEEE*, and G  el Richard, *Senior Member, IEEE*

**Abstract**—Gaussian process (GP) models are very popular for machine learning and regression and they are widely used to account for spatial or temporal relationships between multivariate random variables. In this paper, we propose a general formulation of underdetermined source separation as a problem involving GP regression. The advantage of the proposed unified view is first to describe the different underdetermined source separation problems as particular cases of a more general framework. Second, it provides a flexible means to include a variety of prior information concerning the sources such as smoothness, local stationarity or periodicity through the use of adequate covariance functions. Third, given the model, it provides an optimal solution in the minimum mean squared error (MMSE) sense to the source separation problem. In order to make the GP models tractable for very large signals, we introduce *framing* as a GP approximation and we show that computations for *regularly sampled* and *locally stationary* GPs can be done very efficiently in the frequency domain. These findings establish a deep connection between GP and nonnegative tensor factorizations (NTF) with the Itakura-Saito distance and lead to effective methods to learn GP hyperparameters for very large and regularly sampled signals.

**Index Terms**—Cokriging, Gaussian processes (GP), kriging, nonnegative matrix factorization (NMF), nonnegative tensor factorizations (NTF), probability theory, regression, source separation.

## I. INTRODUCTION

**G**AUSSIAN processes (GP) [28], [35], [36], [44] are commonly used to model functions whose mean and covariances are known. Given some learning points, they enable us to estimate the values taken by the function at any other points of interest. Their main advantages are to provide a simple and effective probabilistic framework for regression and classification as well as an effective means to optimize a model's parameters through maximization of the *marginal likelihood* of the observations. For these reasons, they are widely used in many areas to model dependencies between multivariate random variables

and their use can be traced back at least to works by Wiener in 1941 [43]. They have also been known in geostatistics under the name of *kriging* for almost 40 years [29]. A great surge of interest for GP models occurred when they were expressed as a general purpose framework for regression as well as for classification (see [35] for a review). Their relation to other methods commonly used in machine learning such as multilayer perceptrons, spline interpolation, or support vector machines are now well understood.

Source separation is another very intense field of research (see [10] for a review) where the objective is to recover several unknown signals called *sources* that were mixed together in observable *mixtures*. Source separation problems arise in many fields such as sound processing, telecommunications and image processing. They differ mainly in the relative number of mixtures per source signal and in the nature of the mixing process. The latter is generally modeled as *convolutive*, i.e., as a linear filtering of the sources into the mixtures. When the mixing filters reduce to a single amplification gain, the mixing is called *instantaneous*. When there are more mixtures than sources, the problem is called *overdetermined* and algorithms may rely on beamforming techniques to perform source separation. When there are fewer mixtures than sources, the problem is said to be *underdetermined* and is notably known to be very difficult. Indeed, in this case there are less observable signals than necessary to solve the underlying mixing equations. Many models were hence studied to address this problem and they all either restrict the set of possible source signals or assign prior probabilities to them in a Bayesian setting. Among the most popular approaches, we can mention independent component analysis [6] that focuses both on probabilistic independence between the source signals and on high order statistics. We can also cite nonnegative matrix factorization (NMF) source separation that models the sources as locally stationary with constant normalized power spectra and time-varying energy [16], [27].

In this study, we revisit underdetermined source separation (USS) as a problem involving GP regression. To our knowledge, no unified treatment of the different underdetermined linear source separation problems in terms of classical GP is available to date and we thus propose here an attempt at providing such a formulation whose advantages are numerous. First, it provides a unified framework for handling the different USS problems as particular cases, including convolutive or instantaneous mixing as well as single or multiple mixtures. Second, when prior information such as smoothness, local stationarity, or periodicity is available, it can be taken into account through appropriate *covariance functions*, thus providing a significant expressive power to the model. Third, it yields an optimal way

Manuscript received July 05, 2010; revised November 08, 2010 and February 15, 2011; accepted February 16, 2011. Date of publication February 24, 2011; date of current version June 15, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lieven DeLathauwer. This work was supported in part by the French National Research Agency (ANR) as a part of the DReaM project (ANR-09-CORD-006-03) and by the Quaero Programme, funded by OSEO, French State agency for innovation.

The authors are with Institut Telecom, Telecom ParisTech, CNRS LTCI, France.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2011.2119315

in the minimum mean squared error (MMSE) sense to proceed to the separation of the sources given the model.

In spite of all their interesting features, GP models come at a high  $\mathcal{O}(n^3)$  computational cost where  $n$  is the number of training points. For many applications such as audio signal processing where  $n \approx 10^7$  is common, this cost is prohibitive. Hence, the GP framework has to come along with effective methods to simplify the computations in order to be of practical use. Over the years, many approximation methods have been proposed [31], [34], [37], [38], [40] to address this issue and we show that the common practice of *framing* in audio signal processing can precisely be understood in terms of GP modeling as a particular choice for GP approximation. In particular, we give its connections with recently published partially independent conditional (PIC) approximation [37] and Compact Support (CS) covariance functions [31], [40]. For the special case of locally stationary and regularly sampled signals, we furthermore show that computations can be performed extremely efficiently in the frequency domain and we establish a novel connection between GP models and the emerging techniques of nonnegative tensor factorizations (NTF) [9] using the Itakura-Saito divergence.

The article is organized as follows. First, we present GP and particularly Gaussian process regression (GPR) in Section II. Then, we set out the various linear underdetermined source separation problems in terms of GPR in Section III. In order to make the GP models tractable for very large signals, we introduce *framing* as a GP approximation and we show that computations for *regularly sampled* and *locally stationary* GPs can be done very efficiently in the frequency domain in Section IV. Finally, we illustrate the performance of the methods on synthetic and real data in Section V and draw some conclusions in Section VI.

## II. GAUSSIAN PROCESSES

### A. Introduction

A Gaussian process [28], [35], [36], [44] is a possibly infinite set of scalar random variables  $\{f(x)\}_{x \in \mathcal{X}}$  indexed by an *input space*  $\mathcal{X}$ , typically  $\mathcal{X} = \mathbb{R}^D$ , and taking values on  $\mathbb{R}$ , such that for any *finite* set of inputs  $X = \{x_1 \dots x_n\} \in \mathcal{X}^n$ ,  $\mathbf{f} \triangleq [f(x_1) \dots f(x_n)]^\top$  is distributed with respect to a multivariate Gaussian distribution.<sup>1</sup> A GP is thus completely determined by a mean function  $m(x) = \mathbb{E}[f(x)]$  and a covariance function  $k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$ .

More fundamentally, a GP may be understood as a process whose mean  $m(x)$  and covariance  $k(x, x')$  between any two inputs are known. Given only this prior information, assigning a multivariate Gaussian distribution to  $\mathbf{f}$  given any finite set  $X$  of inputs from  $\mathcal{X}$  is a sensible choice, since it is the probability distribution that maximizes entropy when only the first two moments are known [23].

It has been shown that the class of valid covariance functions coincides with the class of positive definite functions [1]. Let  $X$  be a finite set of elements from  $\mathcal{X}$  that is possibly randomly drawn as in [19], the covariance matrix  $K_{f,XX}$  is defined as

$[K_{f,XX}]_{i,j} = k(x_i, x_j)$  and the probability of  $\mathbf{f}$  given  $X$  is then given by<sup>2</sup>

$$p(\mathbf{f}|X) = \frac{1}{(2\pi)^{\frac{n}{2}} |K_{f,XX}|^{\frac{1}{2}}} \times \exp\left(-\frac{1}{2}(\mathbf{f} - \mathbf{m})^\top K_{f,XX}^{-1}(\mathbf{f} - \mathbf{m})\right) \quad (1)$$

where  $\mathbf{m} \triangleq [m(x_1) \dots m(x_n)]^\top$ . This is usually written

$$f \sim \mathcal{GP}(m(x), k(x, x')).$$

Most studies in underdetermined source separation focus on the *single sensor* scenario  $\mathcal{X} = \mathbb{R}$ . Still, there is no difficulty involved in considering the general case  $\mathcal{X} = \mathbb{R}^D$  and we will see examples of GPs defined on a multidimensional input space in Sections III-C and IV. This framework thus easily allows modeling multivariate functions defined on arbitrary input spaces and many studies have used Gaussian processes for regression ( $f(x) \in \mathbb{R}$ ) as well as for classification ( $f(x) \in \mathbb{N}$ ). Their main advantages are to provide a probabilistic interpretation and a way to compute the variances of the estimates. From now on, we will focus on GPR, since our objective is to highlight the connections between GP and source separation, which is usually stated in terms of processes taking values in  $\mathbb{R}$ . For the sake of notational simplicity, we will assume *a priori* centered signals, i.e.,  $\forall x \in \mathcal{X}, m(x) = 0$ , as it is very common for audio signals. Still, there is no particular issue raised when considering arbitrary mean functions.

### B. Gaussian Processes Regression

Suppose we observe  $y(x) = f(x) + \epsilon(x)$ , with  $f(x)$  being the signal of interest and  $\epsilon(x)$  being some additive signal—usually called *noise*—that is independent from  $f(x)$ , for a finite set  $X$  of input points from  $\mathcal{X}$ :  $X = \{x_1 \dots x_n\} \in \mathcal{X}^n$ . We want to estimate the values taken by  $f$  on a finite and possibly different set  $X^* = \{x_1^* \dots x_n^*\} \in \mathcal{X}^{n^*}$  of input points from  $\mathcal{X}$ . Let us furthermore assume that  $f \sim \mathcal{GP}(0, k_f(x, x'))$  and  $\epsilon \sim \mathcal{GP}(0, k_\epsilon(x, x'))$  where the covariance functions  $k_f$  and  $k_\epsilon$  are known. As  $f$  and  $\epsilon$  are supposed independent, we have

$$f + \epsilon \sim \mathcal{GP}(0, k_f(x, x') + k_\epsilon(x, x')). \quad (2)$$

Let  $K_{f,XX^*}$  be the covariance matrix defined by  $[K_{f,XX^*}]_{ij} = k_f(x_i, x_j^*)$ . We define  $K_{f,X^*X}$ ,  $K_{f,X^*X^*}$ ,  $K_{\epsilon,XX}$  in the same way. Let  $\mathbf{f} \triangleq [f(x_1) \dots f(x_n)]^\top$ ,  $\mathbf{f}^* \triangleq [f(x_1^*) \dots f(x_n^*)]^\top$  and similarly for  $\mathbf{y}$ . We have

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K_{f,XX} + K_{\epsilon,XX} & K_{f,XX^*} \\ K_{f,X^*X} & K_{f,X^*X^*} \end{bmatrix}\right).$$

Classical probability results then assert that the conditional distribution of  $\mathbf{f}^*$  given  $\mathbf{y}$  is (see [35])

$$\mathbf{f}^*|\mathbf{y} \sim \mathcal{N}(\bar{\mathbf{f}}^*, \text{cov}\mathbf{f}^*) \quad (3)$$

<sup>2</sup>Positive semidefinite covariance matrices are possible. In the case of singular  $K_{f,XX}$ , a characterization involving the characteristic function instead of (1) is required.

<sup>1</sup>The symbol  $\triangleq$  denotes a definition.

with<sup>3</sup>

$$\bar{\mathbf{f}}^* = K_{f,X^*X} [K_{f,XX} + K_{\epsilon,XX}]^{-1} \mathbf{y} \quad (4)$$

and

$$\text{covf}^* = K_{f,X^*X^*} - K_{f,X^*X} [K_{f,XX} + K_{\epsilon,XX}]^{-1} K_{f,XX^*}. \quad (5)$$

These expressions show that the maximum likelihood estimate  $\hat{\mathbf{f}}^*$  of  $\mathbf{f}^*|\mathbf{y}$  is found by setting  $\hat{\mathbf{f}}^* = \bar{\mathbf{f}}^*$ , which is also the minimum mean squared error (MMSE) estimate in the Gaussian case. This result will be fundamental when performing source separation using Gaussian processes. We can furthermore compute the covariance of the estimates.

### C. Covariance Functions

Many studies (see [1] for a review) concern valid covariance functions. They belong to the general family of *kernels* and must be definite positive. Some properties of interest have been demonstrated:

- Sums and products of valid covariance functions are valid covariance functions.
- When it is *stationary*, i.e., when it can be expressed as a function of  $\tau = x - x'$ , then the covariance function can be parameterized by its Fourier transform.

Two examples of covariance functions for  $\mathcal{X} = \mathbb{R}^D$  are as follows.

- The squared exponential (SE) covariance function defined by  $k_{\text{SE}}(x, x'|\sigma, M) = \sigma^2 \exp\left(-\frac{(x-x')^\top M(x-x')}{2}\right)$  with  $\sigma^2 > 0$  and  $M$  positive semidefinite. When  $D = 1$ , we have  $k_{\text{SE}}(x, x'|\sigma, \lambda) = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\lambda^2}\right)$ .  $\lambda$  is called a *characteristic length scale* in the sense that  $|x - x'| \gg \lambda$  is required for two points  $x$  and  $x'$  of the process to be independent.
- The less common periodic covariance function of period  $T$  given by  $k_{\text{periodic}}(x, x'|T, \lambda) = \sigma^2 \exp\left(-\frac{2 \sin^2 \frac{\pi(x-x')}{T}}{\lambda^2}\right)$ .

As can be seen, covariance functions are generally parameterized by a set of scalar values such as their Fourier transform when they are stationary, their characteristic length scales, their period, etc. These scalars are often called *hyperparameters* and are usually gathered in a *hyperparameter set*  $\Theta$ . Typical realizations of GPs with SE and periodic covariance functions are given for  $D = 1$  and  $D = 2$  in Fig. 1.

### D. Optimization of the Hyperparameters

In a Bayesian context, we may need to find the hyperparameters that maximize the marginal likelihood of the observations. In other words, we may need to find  $\Theta^*$  such that  $p(\mathbf{y}|X, \Theta^*)$  is maximum. Indeed, even if we may well guess the covariance functions that are adequate to the problem at hand, such as stationary covariance functions parameterized by their Fourier transform or SE covariance functions, it is likely that the hyperparameters that best explain the observations are not exactly known.

<sup>3</sup>In the case of singular covariance matrix  $K_{f,XX} + K_{\epsilon,XX}$ , numerical methods such as Moore-Penrose pseudoinversion may be used.

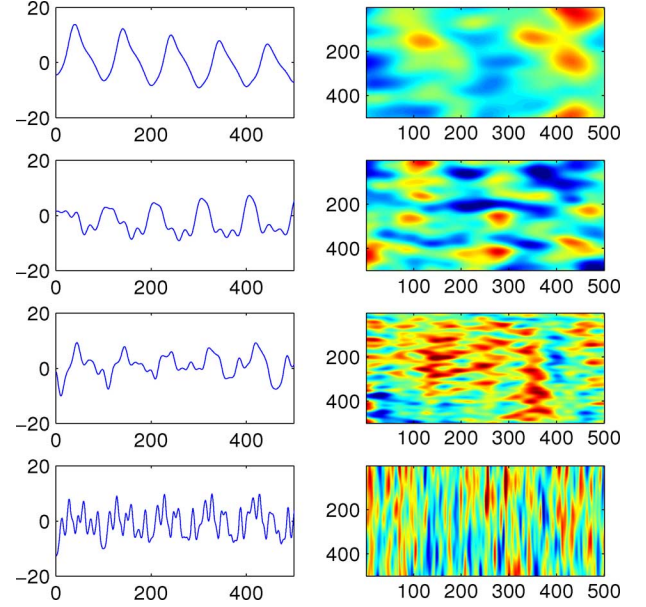


Fig. 1. Typical realizations of GP with SE and periodic covariance functions with different values of the hyperparameters. On the left column,  $D = 1$  and the processes are only parameterized by 3 scalars. On the right,  $D = 2$  and the processes are parameterized by 6 scalars.

To this purpose, we can compute the closed-form expression of the marginal log-likelihood of the observations,  $\log p(\mathbf{y}|X, \Theta)$  as (see [35])

$$\log p(\mathbf{y}|X, \Theta) = -\frac{n}{2} \log 2\pi - \frac{1}{2} \mathbf{y}^\top [K_{f,XX} + K_{\epsilon,XX}]^{-1} \mathbf{y} - \frac{1}{2} \log |K_{f,XX} + K_{\epsilon,XX}| \quad (6)$$

where each covariance matrix depends on  $\Theta$ . Using the opposite of (6) as a cost function, we can proceed to the optimization of the hyperparameters using classical optimization algorithms, in a principled probabilistic framework. Note that depending on the covariance function and the hyperparameter considered, the corresponding optimization problem may or may not be convex.

## III. GAUSSIAN PROCESSES FOR SOURCE SEPARATION

### A. Single Mixture With Instantaneous Mixing

The presentation of GPR given in Section II-B is actually slightly more general than what is usual in the literature. Indeed, it is often assumed that the covariance function  $k_\epsilon$  of the additive signal  $\epsilon$  is given by  $k_\epsilon(x, x') = \sigma^2 \delta_{xx'}$  where  $\delta_{xx'} = 1$  if and only if  $x = x'$  and zero otherwise. This assumption corresponds to additive independent and identically distributed (i.i.d.) white Gaussian noise of variance  $\sigma^2$ .

In our presentation, the additive signal  $\epsilon(x)$  is a GP itself and is potentially very complex. In any case, its covariance function is given by  $k_\epsilon$  and the only assumption made is its independence with the signal of interest  $f(x)$ . A particular example of a model where  $k_\epsilon$  non trivially depends on  $x$  and  $x'$  was for example studied in [20].

The results obtained can very well be generalized to the situation where  $y$  is the sum of  $M$  independent latent Gaussian processes

$$\forall x \in \mathcal{X}, y(x) = \sum_{m=1}^M f_m(x)$$

with

$$f_m \sim \mathcal{GP}(0, k_m(x, x')).$$

In this case, if our objective is to extract the signal corresponding to the source  $m_0$ , we only need to replace  $k_f$  with  $k_{m_0}$  and  $k_\epsilon$  with  $\sum_{m \neq m_0} k_m$  in Section II-B. Note that inversion of  $K_{f,XX} + K_{\epsilon,XX}$  is needed only once for the extraction of all sources. Similarly, we can also jointly optimize the hyperparameters of all covariance functions using exactly the same framework as in Section II-D. We now consider the case of convolutive mixtures of independent GPs.

### B. Single Mixture With Convolutive Mixing

An important fact, which has already been noticed in some studies such as [2] and [4], is that the convolution of a GP, as a linear combination of Gaussian random variables, remains a GP. Indeed, let us consider some GP  $f_{0,m} \sim \mathcal{GP}(0, k_{0,m}(x, x'))$  and let us define

$$f_m(x) = \int_{\mathcal{X}} a_m(x-z) f_{0,m}(z) dz \triangleq (a_m * f_{0,m})(x)$$

where  $a_m : \mathcal{X} \rightarrow \mathbb{R}$  is a stable *mixing filter* from  $f_{0,m}$  to  $f_m$ . If the mean function of  $f_{0,m}$  is identically 0, the mean function of  $f_m$  is easily seen to also be identically 0. The covariance function of  $f_m$  can be computed as  $k_m(x, x') = \mathbb{E}[f_m(x)f_m(x')]$ , that is

$$k_m(x, x') = \int_{\mathcal{X}} \int_{\mathcal{X}} a_m(x-z) a_m(x'-z') k_{0,m}(z, z') dz dz'$$

which is the convolution of  $k_{0,m}$  by  $a_m \times a_m \triangleq (x, x') \in \mathcal{X}^2 \mapsto a_m(x)a_m(x')$

$$k_m(x, x') = ((a_m \times a_m) * k_{0,m})(x, x'). \quad (7)$$

Moreover, if several convolved GPs  $\{f_m = (a_m * f_{0,m})\}_{m=1 \dots M}$  are summed up in a mixture, it can readily be shown that the  $f_m$  are independent if the  $f_{0,m}$  are independent. We thus get back to the instantaneous mixing model using modified covariance functions (7).

### C. Multiple Output GP

We have for now only considered GPs whose outputs lie in  $\mathbb{R}$ . A sizable body of literature focuses on possible extensions of this framework to cases where the processes of interest are multiple-valued, i.e., whose outputs lie in  $\mathbb{R}^C$  for  $C \in \mathbb{N}^*$ . In geostatistics for example, important applications comprise the modeling of co-occurrences of minerals or pollutants in a spatial field. First attempts in this direction [24] include the so-called

*linear model of coregionalization*, that considers each output as a linear combination of some latent processes. The name of *cokriging* has often been used for such systems in the field of geostatistics. If the latent processes are assumed to be GPs, the outputs are also GPs.

In the machine learning community, multiple-output GPs have been introduced [5] and popularized under the name of *dependent GPs*. Several extensions of such models have been proposed subsequently [2]–[4], [30] and we focus here on the model presented in [2] which is very close to the usual convolutive mixing model commonly used in multichannel source separation, e.g., in [32].

Let  $\{y_c(x)\}_{c=1 \dots C}$  be the  $C$  output signals called the *mixtures*. The *convolutive GP* model consists in assuming that each observable signal  $y_c$  is the sum of convolved versions of  $M$  latent GPs of interest  $\{f_{0,m} \sim \mathcal{GP}(0, k_{0,m}(x, x'))\}_{m=1 \dots M}$  that we will call *sources*, plus one specific additional term  $\epsilon_c \sim \mathcal{GP}(0, k_{\epsilon c}(x, x'))$  that is often referred to as *additive noise*. We thus have

$$y_c(x) = \sum_{m=1}^M (a_{cm} * f_{0,m})(x) + \epsilon_c(x). \quad (8)$$

Instead of making a fundamental distinction between  $c$  and  $x$ , the GP framework allows us to consider that  $\{y_c(x)\}_{(c,x) \in \{1 \dots C\} \times \mathcal{X}}$  is a single signal  $\{y(x')\}_{x' \in \{1 \dots C\} \times \mathcal{X}}$  indexed on an extended input space  $\{1 \dots C\} \times \mathcal{X}$ . If we assume that the different underlying sources  $\{f_{0,m}\}_{m=1 \dots M}$  are independent, which is frequent in source separation and that the different  $\{\epsilon_c\}_{c=1 \dots C}$  are also independent, we can express the covariance function  $k((c, x), (c', x'))$  of  $y$  for two extended input points  $(c, x)$  and  $(c', x')$  as

$$k_{cc'}(x, x') = \left( \sum_{m=1}^M k_{cc',m} + \delta_{cc'} k_{\epsilon c} \right) (x, x') \quad (9)$$

$$\text{where } k_{cc',m}(x, x') \triangleq ((a_{cm} \times a_{c'm}) * k_{0,m})(x, x'). \quad (10)$$

For any given  $c$ , the different  $\{f_{cm} \triangleq a_{cm} * f_{0,m}\}_{m=1 \dots M}$  are independent and are GPs with mean functions 0 and covariance functions  $k_{cc,m}(x, x')$ .  $f_{cm}$  will be called the *contribution* of source  $m$  to mixture  $c$ . We can readily perform source separation on  $\mathbf{y}_c$  to recover the different  $\{\mathbf{f}_{cm}\}_{m=1 \dots M}$  using the standard formalism presented in Section II-B. Let  $\hat{\mathbf{f}}_{cm_0}$  be the estimate of  $\mathbf{f}_{cm_0}$ , we have

$$\hat{\mathbf{f}}_{cm_0} = K_{cc,m_0} \left[ \sum_{m=1}^M K_{cc,m} + K_{cc,\epsilon} \right]^{-1} \mathbf{y}_c \quad (11)$$

where  $K_{cc,\epsilon}$  is the covariance matrix of the additive signal  $\epsilon_c$  and where the covariance matrix  $K_{cc,m}$  is defined as  $[K_{cc,m}]_{x,x'} = k_{cc,m}(x, x')$ .

It is important to note here that even if the *sources* are the  $\{f_{0,m}\}_{m=1 \dots M}$ , many systems consider the signals of interest to actually be the different  $\{f_{cm}\}_{c,m}$ . For example, in the case of audio source separation, a stereophonic mixture can be composed of several monophonic sources such as voice, piano, and drums. It is often considered sufficient to be able to separate the different instruments *within* the stereo mixtures and thus to

obtain one stereo signal for each source, rather than trying to recover the original monophonic signals.

Still, for some  $m$ , given the estimates  $\{\hat{\mathbf{f}}_{cm}\}_{c=1\dots C}$  of all the different  $\{\mathbf{f}_{cm}\}_{c=1\dots C}$ , we can for example estimate  $\mathbf{f}_{0,m}$  using standard beamforming techniques.

#### D. Parameter Optimization

Even in complex situations such as those presented in Sections III-B or III-C, we can still use classical optimization methods to maximize the marginal log-likelihood (6) of the observations. Following [2], we will now give a simple way to include multiple output GPs in this framework.

Given a set  $X$  of  $n$  input points and the corresponding  $C$  column vectors  $\{\mathbf{y}_c\}_{c=1\dots C}$ , we can build  $\mathbf{y} \triangleq [\mathbf{y}_1^\top, \dots, \mathbf{y}_C^\top]^\top$  as the  $Cn$  column vector containing all stacked outputs and use (9) to build its covariance matrix  $K$ . We can then proceed to parameters estimation through maximization of the marginal log-likelihood  $\log p(\mathbf{y}|X, \Theta)$  of the observations. Once more, depending on the covariance functions considered, this problem may or may not be convex.

#### E. Conclusion

In this section, we have derived a way to perform underdetermined source separation using GP models when the mixtures are the convolved sums of several independent GPs. Given some covariance functions and mixing filters, we saw that stating the problem in terms of GPs provides a principled way to estimate the source signals that minimize the mean squared error. In the GP framework, optimization of the hyperparameters is done through maximization of the marginal log-likelihood of the mixtures given the model.

To our knowledge, very few references are available to date on the topic. For example, [33] performs source separation using GPs in the determined case, but the covariance functions are therein applied on the outputs of the source signals rather than on the coordinates themselves (i.e., time or spatial position). A successful application of GPs to a subject close to source separation can also be found in [39] for echo cancellation.

### IV. GP APPROXIMATIONS FOR LARGE SIGNALS

#### A. The Need for Approximations

The main issue with GP models is the need to invert the  $n \times n$  covariance matrix of the learning points for inference (4) and for each evaluation of the observation likelihood in (6). In many areas of interest, we cannot afford to handle such a big matrix, since it is not computationally tractable. In audio signal processing for example, values such as  $n \approx 10^7$  are common and GP models cannot be used without a significant reduction of the computational cost of the method.

In order to address this issue, many authors have proposed *sparse* approximation techniques [31], [34], [37], [38], [40] over the years that all aim at making GP inference possible for large datasets. As highlighted in [34], many methods rely on the choice of a small set of input points called *the inducing inputs* to approximate the posterior distribution at test points  $X^*$ . Among those methods, we can mention the fully independent

conditional (FIC) approximation [34], [38], that considers all the test points and the learning points independent given the inducing inputs. This leads to a very important reduction of the computational burden, but heavily relies on the density of the inducing points [37] to yield good estimates. Another approximation called partially independent conditional (PIC) [37] no longer makes the assumption that both the training and test cases are independent given the inducing points, but rather that each of them not only depends on the possibly remote inducing points, but also on a limited number of other learning points nearby. This technique has the advantage of producing better estimates than FIC, while maintaining an easy inversion of the  $n \times n$  covariance matrix that is now block-diagonal. Its main disadvantage is to lead to discontinuities of the estimates between the blocks, which may be problematic for some applications such as audio processing.

Another very attractive direction of research in the last few years has been the consideration of covariance functions with compact support (CS) [31], [40], i.e., covariance functions  $k(x, x')$  such that  $\|x - x'\| > l \Rightarrow k(x, x') = 0$  for some given scale  $l$ . The idea underlying these techniques is to consider that if they are sufficiently far from each other, two points will be independent. If such covariance functions are used, the covariance matrix is sparse and inference through Cholesky decompositions is done much faster [40]. The main issue with this approach is to design covariance functions that correspond to some prior knowledge about the sources and that have CS at the same time.

In Sections IV-B and IV-C, we introduce a general method for fast inference in GP models based on *framing* and that is a direct generalization of the common practice in audio signal processing.

Another important computational simplification is introduced in Sections IV-D and IV-E when the signals are regularly sampled. In that case, we show that when the covariance functions are assumed stationary and separable, approximate inference can be done extremely efficiently in the frequency domain.

When both approaches are combined into so called *locally scaled* and *framewise-independent* stationary covariance functions, we show in Section IV-F that inference and learning of hyperparameters become equivalent to recent and powerful NTF techniques [9].

#### B. Frames

In audio signal processing, it is common to split the signals into overlapping *frames* and to process the frames separately. Formally, the frames  $\{y_i(x')\}_{i \in \mathbb{N}}$  are defined as small portions of the original signal. The advantage of the technique is that the frames are small and can be easily processed. The original signals can then be recovered through a deterministic *overlap-add* procedure: each frame is multiplied by a *weighting function*  $g: \mathcal{X}' \rightarrow \mathbb{R}^+$  to ensure smooth transitions between the frames and is added to the reconstructed signal.  $g$  is often a Hann or a triangular window.

This idea can very well be generalized in any dimension  $D$ . Instead of considering the original signal  $y$ , we can split it into overlapping frames of smaller dimension. To this end, we consider a *frame input set*  $\mathcal{X}' \subset \mathcal{X}$ , a summable *weighting function*

$g : \mathcal{X}' \rightarrow \mathbb{R}^+$  and a set of *frame positions*  $\{t_i \in \mathcal{X}\}_{i \in \mathbb{N}}$  such that

$$\forall x \in \mathcal{X}, I_x \triangleq \{i \in \mathbb{N} : x - t_i \in \mathcal{X}'\} \neq \emptyset. \quad (12)$$

$I_x$  is thus the set of frame numbers to which the input point  $x$  is mapped. Condition (12) ensures that each point of the signal is represented in at least one frame. Finally, given some signal  $\{y(x)\}_{x \in \mathcal{X}}$ , we can make the assumption that there is a set of *frames*  $\{y_i(x)\}_{i \in \mathbb{N}, x \in \mathcal{X}'}$ , also noted  $\mathcal{G}\{y\}$  in the following, such that

$$\forall x \in \mathcal{X}, y(x) = \frac{1}{\sum_{i \in I_x} g(x - t_i)} \sum_{i \in I_x} g(x - t_i) y_i(x - t_i). \quad (13)$$

When considering a finite set  $X$  of  $n$  input points, we only need to consider the frames  $I \triangleq \bigcup \{I_x\}_{x \in X}$  that contain at least one input point from  $X$ . Let  $X'_i = \{x' \in \mathcal{X}' \mid \exists x \in X : x' = x - t_i\}$  be the finite set of points from  $\mathcal{X}'$  to which the elements of  $X$  in the scope of frame  $i$  are mapped and let<sup>4</sup>  $L_i = \#(X'_i)$ . Given some signal  $\{y(x)\}_{x \in X}$ , it is always possible to build a set of frames obeying (13). This can be achieved by choosing  $X'_i$  and  $y_i(x)$  such that

$$\forall (i, x') \in I \times \mathcal{X}', (t_i + x' \in X) \Rightarrow \begin{cases} x' \in X'_i \\ y_i(x') = y(t_i + x') \end{cases}. \quad (14)$$

When they make use of framing, usual methods focus on the frames  $\mathcal{G}\{y\}$  as the signals of interest rather than on  $y$ . Indeed, a good model for  $\mathcal{G}\{y\}$  is *de facto* a good model for  $y$  since it can be computed deterministically from  $\mathcal{G}\{y\}$ . In such methods based on framing, the set (14) of frames is usually taken as being the observation. From our point of view, the frames are simply another process which is indexed on  $\mathbb{N} \times \mathcal{X}'$  and from which we can deterministically recover  $y$  which is indexed on  $\mathcal{X}$ .

### C. Frame-Wise Independence Assumption

Given a signal  $\{y(x)\}_{x \in \mathcal{X}}$  and a corresponding set of frames  $\{y_i\}_{i \in \mathbb{N}}$ , a classical assumption consists in writing that the different  $y_i$  are independent. As  $y$  can be deterministically computed from  $\{y_i\}_{i \in \mathbb{N}}$ , this is written

$$\log p(y|X, \Theta) = \sum_{i \in I} \log p(y_i|X'_i, \Theta). \quad (15)$$

If  $\mathcal{G}\{y\}$  is modeled as a GP, the frame-wise independence assumption is equivalent to modeling the covariance function  $k((i, x), (i', x'))$  of  $\mathcal{G}\{y\}$  as

$$k((i, x), (i', x')) = \delta_{ii'} k_i(x, x') \quad (16)$$

with  $k_i$  being the covariance function of the GP  $\{y_i(x)\}_{x \in \mathcal{X}'}$ . Let  $X$  be a finite set of input points from  $\mathcal{X}$ ,  $y$  a process indexed on  $X$  and  $I \triangleq \bigcup \{I_x\}_{x \in X}$  be the corresponding frame indexes for a framing  $\mathcal{G}$ . Let  $n_I$  be the number of frames. If we model  $\mathcal{G}\{y\}$  as a GP, we readily see that it is equivalent to a multiple output GP as seen in Section III-C with  $n_I$  outputs whose input set is  $\mathcal{X}'$ . We can thus stack its outputs and observe that the corresponding covariance matrix is block-diagonal due to the

<sup>4</sup>For a countable set  $X$ ,  $\#(X)$  denotes the number of elements in  $X$ .

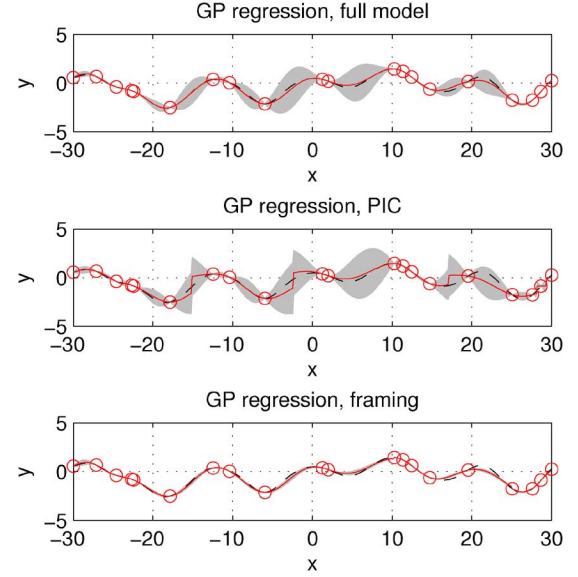


Fig. 2. A simple regression example with a full GP model (up), PIC (middle) and framing (bottom). The dotted and red lines are, respectively, the true and estimated signal. The gray area represents three standard deviations of the estimates around the mean and the circles stand for observed points.

frame-wise independence assumption. Its inverse is thus easily computed.

The main computational trick involved by framing is hence to split the signal into overlapping frames, with a synthesis scheme that allows perfect reconstruction. Then, the frames are supposed to be independent and the corresponding covariance matrix becomes block diagonal. The advantage of this method is that when the frames are overlapping, each point estimate is a smoothing of several estimates computed in the different frames that contain this point, thus avoiding systematic discontinuities. In Fig. 2, we illustrate this advantage of framing over PIC to produce smooth estimates in a very simple regression problem.

The connections between *frame-wise independent* correlation functions and other existing models such as PIC or CS covariance functions [31], [40] are numerous. First, we see that framing without overlap between the frames is very similar to PIC except for the fact that it does not take inducing points outside the frames into account. In framing, such remote inducing points are handled through the use of overlapping frames of different scales. Second, when considering (13), which gives the expression of the signal given its constitutive frames, we can straightforwardly compute the covariance function of the signal  $y$  itself given the covariance functions of the different frames. This computation is actually very similar to that led in [31] where the *basis function* used in the computation becomes the *weighting function*  $g$  we considered here. The basis function proposed in [31] is precisely the HANN window which is a very popular choice for  $g$  in audio processing.

Of course, as in practice the frames are built using (14), there is a duplication of the samples that belong to overlapping frames and the independence assumption between the frames may seem unjustified. Nevertheless, the idea underlying framing is that even if the occurrence of one point from  $X$  in some frame gets duplicated in another, and even if the corresponding observed



values are connected or equal, they are supposed to be produced by two different underlying processes that do not share the same covariance function.

Still, there are interesting conceptual issues raised by framing that should be more thoroughly studied in the future. In particular, contrary to PIC, framing as it was exposed here suffers from overconfidence. This can be seen by computing the variance of the estimates for  $y$  given by (13) and then noticing that the more a point will get duplicated in different frames because of overlap, the smaller the *a posteriori* variance of this point will become. This is due to the independence assumption between the frames. If this assumption was strictly legitimate, this diminution of the variance would be justified. However, as the overlap between the frames gets large, independence cannot be a valid assumption anymore and the variances get underestimated. This is illustrated in Fig. 2 where the overlap was large.

Even if overlapping the different frames is common practice in audio signal processing, its consequences on statistical models has been largely neglected. In [26], LE ROUX *et al.* devise practical ways of consistently handling the dependencies between the frames as a postprocessing step. Still, to our knowledge, practical statistical models that fully take the overlap between the frames into account while remaining computationally tractable are yet to be proposed.

#### D. Stationarity Assumption for Regularly Sampled Signals

In this section, we assume that the signals are defined on  $\mathcal{X} = \mathbb{R}^D$  for  $D \geq 1$  and that  $x \in \mathcal{X}$  can be written  $x = (x_1, \dots, x_D)$ . We will moreover assume that all the covariance functions  $k$  that we consider are *separable*, i.e., there are  $D$  covariance functions  $k^{(d)}$  such that

$$\forall(x, x') \in \mathcal{X}^2, k(x, x') = \prod_{d=1}^D k^{(d)}(x_d, x'_d). \quad (17)$$

It is readily shown that this assumption implies that all the covariance matrices  $K$  considered can be expressed as a Kronecker product<sup>5</sup> of  $D$  covariance matrices  $K^{(d)}$  of lower dimensions

$$K = K^{(1)} \otimes K^{(2)} \dots \otimes K^{(D)} \triangleq \bigotimes_{d=1}^D K^{(d)}. \quad (18)$$

From now on, we suppose that the points are regularly sampled. This is equivalent to assuming that any signal  $y$ ,  $\mathbf{f}_m$  or  $\mathbf{k}$  considered is the vectorization<sup>6</sup> of a corresponding underlying  $D$ -dimensional tensor  $\underline{y}$ ,  $\underline{f}_m$  or  $\underline{k}$ . Indeed, we will show in Section IV-D-1 that computations can be very concisely written using these tensors, which are actually natural to consider. For example, when  $D = 2$ , it makes sense to directly think of regularly sampled signals as matrices instead of their vectorized counterpart.

*1) GPs for the Separation of Stationary Mixtures:* As seen in Section II-C, a *stationary* covariance function  $k(x, x')$  between two input points can be expressed as a function of their differ-

ence  $\tau = x - x'$ . It is noted  $k(x - x')$ . If all covariance functions considered are stationary, the computations become particularly simple.

Indeed, let us assume that a mixture  $\{y(x)\}_{x \in \mathcal{X}}$  is the sum of several GPs  $\{f_m(x)\}_{m=1 \dots M, x \in \mathcal{X}}$  whose covariance functions  $k_m(x - x')$  are all stationary, and let us furthermore suppose that we are interested in separating the different sources for all points in  $X$ , thus having  $X^* = X$ . The covariance matrix  $K_y$  of  $y$  is given by:  $K_y = \sum_{m=1}^M K_m$  where  $K_m$  is the covariance matrix of source  $m$ .

Considering (18),  $K_m$  is given by  $K_m = \bigotimes_{d=1}^D K_m^{(d)}$  where  $[K_m^{(d)}]_{i,j} = k_m^{(d)}(x_{i,d} - x_{j,d})$ .  $K_m^{(d)}$  can approximately be considered as circulant.<sup>7</sup> It is readily shown that any circulant matrix  $M$  can be expressed as  $M = W_F^* \Lambda W_F$  where  $W_F$  is the discrete Fourier transform matrix<sup>8</sup> and where  $\Lambda$  is diagonal. Thus, for all  $m$  and  $d$ , there is a diagonal positive semidefinite matrix  $\text{diag} S_m^{(d)}$  such that  $K_m^{(d)} \approx W_F^* \text{diag} S_m^{(d)} W_F$  where the vector  $S_m^{(d)}$  is the discrete Fourier transform of  $\tau \mapsto k_m^{(d)}(\tau)$ . We can thus write  $K_y$  as

$$K_y = \sum_{m=1}^M \bigotimes_{d=1}^D W_F^* \text{diag} S_m^{(d)} W_F. \quad (19)$$

Using classical results from tensor algebra, (19) can be written

$$K_y = \left( \bigotimes_{d=1}^D W_F^* \right) \left( \sum_{m=1}^M \bigotimes_{d=1}^D \text{diag} S_m^{(d)} \right) \left( \bigotimes_{d=1}^D W_F \right). \quad (20)$$

We can use this property to extract a given source  $m_0$ , and write<sup>9</sup> (4) as

$$\overline{\mathbf{f}}_{m_0}^* = \left( \bigotimes_{d=1}^D W_F^* \right) \left( \frac{\bigotimes_{d=1}^D \text{diag} S_{m_0}^{(d)}}{\sum_{m=1}^M \bigotimes_{d=1}^D \text{diag} S_m^{(d)}} \right) \left( \bigotimes_{d=1}^D W_F \right) y. \quad (21)$$

Introducing the  $D$ -dimensional tensor<sup>10</sup>

$$\underline{S}_m = S_m^{(1)} \circ S_m^{(2)} \dots \circ S_m^{(D)} \triangleq \bigcirc_{d=1}^D S_m^{(d)} \quad (22)$$

as the *model for source  $m$*  and  $\mathcal{F}_D\{\underline{y}\}$  as the  $D$ -dimensional Fourier transform of  $\underline{y}$ , we can simply write (21) in tensor form as

$$\mathcal{F}_D\{\overline{\mathbf{f}}_{m_0}^*\} = \left( \frac{\underline{S}_{m_0}}{\sum_{m=1}^M \underline{S}_m} \right) \cdot \mathcal{F}_D\{\underline{y}\} \quad (23)$$

which is similar to the classical Wiener filter for stationary processes. The differences between this expression and the classical one is first that it is valid for any dimension  $D$  of the input space and second that it is not restricted to the case of only two stationary sources. The sources themselves can be recovered through an inverse  $D$ -dimensional Fourier transform. The nonnegative tensor  $\underline{S}_m$  can be understood as the  $D$ -dimen-

<sup>7</sup>If the signal is regularly sampled, this approximation holds when the number  $n_d$  of points along dimension  $d$  tends to infinity or when  $k^{(d)}(\tau)$  is periodic of period  $n_d/p$  with  $p \in \mathbb{N}^*$ .

<sup>8</sup> $W_F^*$  denotes the complex conjugate of  $W_F$ .

<sup>9</sup> $A/B$  and  $A \cdot B$  are, respectively, the element-wise division and multiplication of  $A$  and  $B$ .

<sup>10</sup>The symbol  $\circ$  denotes the outer product.

<sup>5</sup>See [9] for a concise introduction to tensor algebra.

<sup>6</sup>Vectorization is done recursively. For example, with  $D = 2$  where tensors are matrices, it is done one row after the other.

sional Fourier transform of the stationary covariance function  $k_m$ . Note that the complexity of this *approximate* GP inference method relying on stationarity of the covariance functions and on regular sampling is  $\mathcal{O}(n \log n)$ , and it is dominated by the computation of Fourier transforms, for which there exist very efficient and specialized algorithms. If  $\mathcal{F}_D\{y\}$  is known beforehand, the complexity of (23) decreases to  $\mathcal{O}(n)$  which is remarkable for a GP inference technique.

2) *Marginal Likelihood for Stationary Sources:* When all the covariance functions considered are stationary and parameterized by some hyperparameter set  $\Theta$  that consists of their respective  $D$ -dimensional Fourier transforms, i.e.,  $\Theta = \{\underline{S}_1 \cdots \underline{S}_M\}$ , it can readily be shown that the marginal log-likelihood  $\log p(y|X, \Theta)$  of the observations given regularly spaced input points and the hyperparameters simplifies from (6) to

$$\log p(y|X, \{\underline{S}_1 \cdots \underline{S}_M\}) = -\frac{1}{2} \sum_{i_1, \dots, i_D} \left[ \frac{|\mathcal{F}_D\{y\}_{i_1, \dots, i_D}|^2}{\sum_{m=1}^M [\underline{S}_m]_{i_1, \dots, i_D}} + \log \sum_{m=1}^M [\underline{S}_m]_{i_1, \dots, i_D} \right] + \text{Cte.} \quad (24)$$

Considering (24), we see that it is equivalent up to an additive constant independent of  $\Theta$  to half the opposite IS divergence<sup>11</sup> between<sup>12</sup>  $|\mathcal{F}_D\{y\}|^2$  and  $\sum_{m=1}^M \underline{S}_m$ :

$$\log p(y|X) = -\frac{1}{2} D_{\text{IS}} \left( |\mathcal{F}_D\{y\}|^2 \middle| \sum_{m=1}^M \underline{S}_m \right) + \text{Cte.} \quad (25)$$

The evaluation of the likelihood can be done in  $\mathcal{O}(n \log n)$  operations when the signals are regularly sampled and the covariance functions are stationary. If the squared  $D$ -dimensional Fourier transform  $|\mathcal{F}_D\{y\}|^2$  of the signal is known beforehand—it is typically computed only once—the computational complexity is reduced to  $\mathcal{O}(n)$ .

#### E. Locally Stationary Covariance Functions

Let  $\{y(x)\}_{x \in \mathcal{X}}$  be a particular signal, observed on a finite input set  $X \in \mathcal{X}^n$  and let  $\{y_i \in \mathbb{R}^{X_i}\}_{i \in I}$  be a set of  $n_I$  corresponding frames. As in Section IV-C, we can assume that the frames are independent and we can further suppose that the covariance function  $k_{im}$  of source  $m$  within frame  $i$  is stationary. This means that we model each source as being composed of several locally stationary frames, each of which has its own covariance function. The resulting signal is *not* supposed stationary with this assumption, only its restrictions to small regions of the input space  $\mathcal{X}$  are assumed stationary.

Let us denote  $\underline{Y}$  the  $(D+1)$ -dimensional tensor whose last dimension goes over the frames and whose first  $D$  dimensions for a fixed frame contain the  $D$ -dimensional Fourier transform of the signal tensor for this frame as in Section IV-D. As this tensor is called the short term Fourier transform (STFT) of the signal when  $D = 1$ , it will be called the STFT tensor of the mixture. We define the STFT tensor  $\underline{F}_m$  of the sources and the

STFT tensor  $\underline{S}_m$  of the covariance function of source  $m$  in the same way. We can use the results from the previous section for each frame and for source  $m_0$ : the MMSE estimate  $\underline{F}_{m_0}^*$  of  $\underline{F}_{m_0}$  is given by

$$\underline{F}_{m_0}^* = \frac{\underline{S}_{m_0}}{\sum_{m=1}^M \underline{S}_m} \cdot \underline{Y}. \quad (26)$$

The sources can then be recovered by first applying an inverse  $D$ -dimensional Fourier transform to the estimate (26) for each frame, and then using the reconstruction scheme (13) to obtain the estimated sources in the original input space  $\mathcal{X}$ .

Let  $\Theta = \{\underline{S}_1, \dots, \underline{S}_M\}$  be the models for the sources. The marginal likelihood  $\log p(y|X, \Theta)$  of the observations can similarly be shown to be

$$\log p(y|X, \Theta) = -\frac{1}{2} D_{\text{IS}} \left( |\underline{Y}|^2 \middle| \sum_{m=1}^M \underline{S}_m \right) + \text{Cte} \quad (27)$$

where the constant is independent of  $\Theta$ . This very simple expression can be computed in  $\mathcal{O}(n)$  when  $|\underline{Y}|^2$  is known and permits to efficiently proceed to hyperparameters learning as demonstrated in Section IV-F.

#### F. Putting Structures Over the Covariances

Given some regularly sampled signal tensor  $\underline{y}$  and its corresponding STFT tensor  $\underline{Y}$  as defined in Section IV-E, we have seen that source separation can be very efficiently performed provided some  $(D+1)$ -dimensional model  $\underline{S}_m$  is known for every source. As highlighted by CEMGIL *et al.* in [7] or [8] for the case of audio processing ( $D = 1$ ), the important issue raised by this probabilistic framework becomes devising realistic but effective models for the nonnegative sources parameters  $\underline{S}_m$ .

In audio signal processing ( $D = 1$ ), (26) is known as adaptive or generalized Wiener filtering and many methods for source separation such as [8], [32] use this technique in a principled way to recover the sources in the frequency domain. Those studies state their probabilistic model in the frequency domain where the time-frequency bins are supposed to be distributed with respect to independent Gaussian distributions. In our approach, the model is expressed directly in the original input space. The two points of view are actually equivalent: a stationary GP has an independently distributed Gaussian representation in the frequency domain.  $\underline{S}_m$  can hence be seen either as the STFT tensor of a covariance function or as a tensor containing the variances of the independent components of  $\underline{Y}$ .

Focusing on the second interpretation of  $\underline{S}_m$ , recent studies [8], [12], [13] proposed to model these tensors as Gamma Markov random fields (GMRF). This is a sensible choice indeed, because such models guarantee the nonnegativity of all the elements of  $\underline{S}_m$  while implementing the knowledge that for a given source, the spectrum is much likely to exhibit some continuity over time, or over the frequencies, or over both. As GMRF do not provide a closed-form expression for the marginal log-likelihood of the observations, the learning of hyperparameters has to be done using approximate methods. To this end, DIKMEN *et al.* [12] propose to use contrastive divergence [22] and report good results. To our knowledge, no generalization of GMRF has yet been published for input

<sup>11</sup>  $D_{\text{IS}}(\underline{x}|\underline{y}) \triangleq \sum_{i_1, \dots, i_D} [(\underline{x}_{i_1, \dots, i_D} / \underline{y}_{i_1, \dots, i_D}) - \log(\underline{x}_{i_1, \dots, i_D} / \underline{y}_{i_1, \dots, i_D}) - 1]$ .

<sup>12</sup> For a matrix  $M$ ,  $[M^{\cdot 2}]_{ij} \triangleq M_{ij}^2$ .



spaces of dimension greater than 2, but GP modeling may greatly benefit from such an extension.

Another point of view is to introduce some deterministic structure into the covariance functions of the GPs. A simple assumption to this end is to consider that for a given source  $m$ , the covariance functions of the different *independent* frames are *stationary* and *locally scaled*, i.e., identical up to an amplification gain depending on the frame. The model for source  $m$  and frame  $i$  can then be written

$$\mathbf{S}_{im} = H_{im} \underline{S}_{0,m} \quad (28)$$

where  $\underline{S}_{0,m} = \bigcirc_{d=1}^D S_{0,m}^{(d)}$  is the  $D$ -dimensional Fourier transform of some *template* covariance function  $k_{0,m}$  for source  $m$  that is independent of the frame index  $i$ . We get

$$\mathbf{S}_m = \left( \bigcirc_{d=1}^D S_{0,m}^{(d)} \right) \circ H_m \quad (29)$$

where  $H_m = (H_{1m} \cdots H_{nIm})$  denotes the amplification gains of the covariance function for source  $m$  on the different frames. Considering (29) we readily see that it is equivalent to a classical NTF model called Canonical Polyadic (CP) decomposition.<sup>13</sup> The different parameters become  $\Theta = \{ \{ H_m, S_{0,m}^{(1)} \cdots S_{0,m}^{(D)} \}_{m=1 \dots M} \}$  and can be estimated by standard CP algorithms using the IS-divergence function. See [9] for a review of these models and algorithms.

### G. Optimization

We have shown how GP learning can be connected to recent NTF techniques by factorizing the covariance structure of the GP model into a CP decomposition. More generally and depending on the application, the covariances can be factorized in many other ways to account for some prior knowledge we may have concerning the structure of the sources. For example, if the covariances are considered to be the outer product of some shared dictionaries, the tensor decompositions to be used become particular cases of Block Components Decompositions as introduced in [25]. Many very informative models can be designed this way, that decompose the covariance structure of the sources onto sophisticated dictionaries. In music processing ( $D = 1$ ) for example, [41] decomposes the covariances into templates of harmonic bases. Other models of this type have also been used to model and extract singing voice signals from polyphonic mixtures with very promising results [14].

In any case, when an appropriate model has been chosen for  $\{\mathbf{S}_m\}_{m=1 \dots M}$ , we have seen in Section IV-E that hyperparameters learning can be done by minimizing the IS-divergence between  $\sum_m \mathbf{S}_m$  and  $[\mathbf{Y}]^2$  through tensor factorizations. Efficient algorithms for IS-NTF can be found in the literature, for example in [9].

## V. EVALUATION

In this section, we demonstrate the performance of the proposed approach based on GP models for the separation of real-valued mixtures. In Section IV-A, we first show that GP can easily be used for the separation of synthetic 2D random fields, or textures ( $D = 2$ ). Then, we show in Section V-B how GP can

be used for the separation of drums signals in real polyphonic stereo recordings.

### A. Synthetic Additive Textures

In this section, we set  $D = 2$ , which means that we aim at separating additive functions  $f_m(x_1, x_2)$  defined on the plane and summed in an observable mixture signal  $y(x_1, x_2)$ . For this toy example, we will consider the case of one mixture ( $K = 1$ ) that is the sum of  $M = 2$  stationary sources.<sup>14</sup> Following the notations that were introduced in Section IV-D, we will thus suppose that the mixture tensor  $y$  is the sum of two sources tensors  $\underline{f}_1$  and  $\underline{f}_2$ . The corresponding vectors  $y$ ,  $f_1$  and  $f_2$  will denote the vectorization of these tensors one row after the other.  $X$  denotes the corresponding coordinates.

In this experimental setup, the dimensions of the sources and mixture tensors are  $500 \times 500$  each, leading to  $n = 250\,000$ . In the following, we will assume that the covariance function of each source along each dimension is stationary. For the experiment, the covariance functions were arbitrarily set to

$$k_m^{(d)}(x_d, x'_d) = \exp \left( -\frac{2 \sin^2 \frac{\pi(x_d - x'_d)}{T_{m,d}}}{l_{m,d}^2} - \frac{(x_d - x'_d)^2}{2\lambda_{m,d}^2} \right) \quad (30)$$

where  $\{T_{m,d}, l_{m,d}, \lambda_{m,d}\}_{m,d}$  are scalar parameters.

This model implements a particular prior knowledge where the sources are known to exhibit some kind of complex structure. More specifically, source  $m$  is known to be pseudoperiodic of period  $(T_{m,1}, T_{m,2})$  and  $(l_{m,1}, l_{m,2})$  controls the smoothness within one period. A further lengthscale  $(\lambda_{m,1}, \lambda_{m,2})$  controls global covariance between two input points. In the particular example shown in Fig. 3, the parameters were as shown in the table at the bottom of the following page.

1) *Data Synthesis*: Generating a realization of a GP with some known covariance matrix  $K$  is generally addressed through Cholesky or singular value decompositions (SVD) of the covariance matrix [35]. As we have  $n = 250\,000$ , we cannot naively implement this idea here. A simple way to circumvent this problem is to write  $K$  as in (18) and then to perform a Cholesky decomposition of each  $K^{(d)}$  to get  $K^{(d)} = L^{(d)} L^{(d)\top}$ . The Cholesky decomposition of  $K = \bigotimes_{d=1}^D L^{(d)} L^{(d)\top}$  is finally obtained by  $K = (\bigotimes_{d=1}^D L^{(d)}) (\bigotimes_{d=1}^D L^{(d)})^\top$  and a realization of this GP can be very easily generated. Indeed, let  $R$  be a vector of length  $n$  whose entries are i.i.d. Gaussian random variables of unit variance.  $K$  is the covariance matrix of  $(\bigotimes_{d=1}^D L^{(d)}) R$ .<sup>15</sup>

2) *Source Separation*: In this very simple experimental setup, we consider that the 12 hyperparameters for the sources covariance functions (30) are known beforehand.

We can perform separation through the exact method presented in Section IV-D-1. To this purpose, we can build the spec-

<sup>14</sup>This usecase is common in geostatistics: the observed signal is often modeled as the sum of the signal of interest with a contaminating white Gaussian noise [11]. Estimating the value of the target signal through Kriging is hence a special case of source separation with GP priors.

<sup>15</sup>We can further speed up this computation by using the fact that for  $\mathbf{c} = \text{vec}(\underline{C})$  and matrices  $A$  and  $B$  of appropriate size,  $(A \otimes B) \mathbf{c} = \underline{A \underline{C} B^\top}$ . This avoids considering such a big matrix as  $(\bigotimes_{d=1}^D L^{(d)})$ .

<sup>13</sup>CP is also called PARAFAC or CANDECOMP [9].

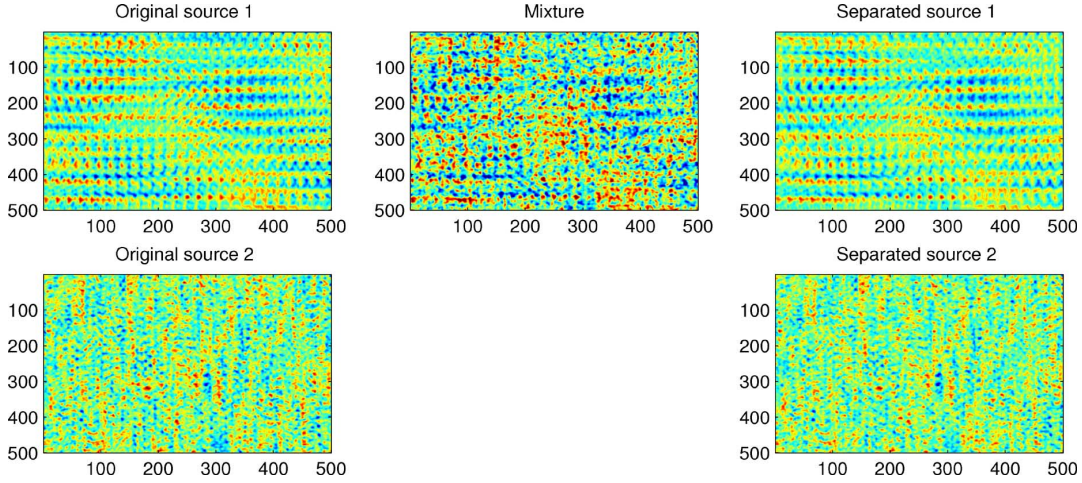


Fig. 3. GP for the separation of two stationary random fields ( $D = 2$ ) using a Gaussian Process model. (left) Original sources. (center) Mixture. (right) Estimated sources.

tral covariance tensor  $\underline{\mathbf{S}}_m$  of each source as the outer product of the Fourier transforms of (30) along each dimension and then perform separation in the frequency domain as in (23). The sources are recovered through an inverse 2-dimensional Fourier transform. It is worth noticing here that the computations are performed extremely rapidly since they only involve element-wise multiplications of  $500 \times 500$  images, instead of the inversion of the  $250\,000 \times 250\,000$  covariance matrix required by the basic GP setup. Overall computations for this example—synthesis and source separation—are achieved in less than 3 s on a standard laptop computer.

Results for one example are shown in Fig. 3. The average Signal to Error ratio obtained on 50 experiments was of 8 dB, which is very encouraging.

### B. Separation of Drums Signal in Polyphonic Music

In this section, we apply the general framework we have presented in Sections III and IV to the separation of drums signals in polyphonic music. The regularly sampled signals we consider are thus defined on the input space  $\mathcal{X} = \mathbb{Z}$  of dimension  $D = 1$ .

Separation of the drums track from polyphonic music is a challenging task that has already been addressed in several studies such as [18] and [21]. Whereas HÉLEN and VIRTANEN [21] perform a NMF of the mixture and then group the different components obtained through a classification procedure, GILLET and RICHARD [18] decompose the mixture signal with spectral templates learned from a drums database.

In Section V-B-1, we introduce a GP model for this task and in Section V-B-2, we compare its performance with the state of the art [18].

1) *GP Model*: The observed mixtures  $y(x)$  are supposed to be the sum of two independent GPs  $s_d(x)$  and  $s_r(x)$  corresponding, respectively, to the *drums* and the *musical residual* tracks. We assume that some framing  $\mathcal{G}\{y\}$  with  $n_I$  frames of same length as defined in Section IV-B is available for the mixtures and we aim at estimating the framings  $\mathcal{G}\{s_d\}$  and  $\mathcal{G}\{s_r\}$  of the different sources such that  $\mathcal{G}\{y\} = \mathcal{G}\{s_d\} + \mathcal{G}\{s_r\}$ .

We suppose that each of the signals  $s_d$  and  $s_r$  are themselves the sum of several independent processes called *components*. In our example, the  $R_d$  different components of  $s_d$  are the five most common sources we find in a drums signal, e.g., kick drum, snare drum, hihat, bells, and clap sounds. The  $R_r$  different components of the musical residual are all the other elements composing the polyphonic mixture. This assumption can be written  $s_d = \sum_{m=1}^{R_d} f_m$  and  $s_r = \sum_{m=R_d+1}^{R_d+R_r} f_m$ .

In the model we are considering, we will assume that all the components  $f_m$  are GP whose covariance functions  $k_m$  are *locally scaled*, *frame-wise independent* and *stationary* as defined in Section IV-F. For some frame  $i$ , they can thus be expressed as

$$k_{im} = H_{im} k_{0,m} \quad (31)$$

where  $k_{0,m}$  denotes the template stationary covariance function for component  $m$  and  $H_m = (H_{1m} \cdots H_{n_I m})$  are the nonnegative activation gains of this component within the frames. Introducing the Fourier transform  $S_{0,m}$  of  $k_{0,m}$  and using the method presented in Section IV-F, the MMSE estimate  $\hat{\underline{F}}_d$  of the STFT  $\underline{F}_d$  of the drums signal is given by

$$\hat{\underline{F}}_d = \frac{\sum_{m=1}^{R_d} S_{0,m} \circ H_m}{\sum_{m=1}^{R_d+R_r} S_{0,m} \circ H_m} \cdot \underline{\mathbf{Y}} \quad (32)$$

$m$	$\lambda_{m,1}$	$\lambda_{m,2}$	$T_{m,1}$	$T_{m,2}$	$l_{m,1}$	$l_{m,2}$
1	100	100	50	20	0.5	0.7
2	40	4	25	$+\infty$	0.7	N/A

where  $\underline{\mathbf{Y}}$  is the STFT of the mixtures. The model  $\underline{\mathbf{S}} \triangleq \sum_m \underline{\mathbf{S}}_m$  becomes  $\underline{\mathbf{S}} = \sum_{m=1}^{R_d+R_r} S_{0,m} \circ H_m$ . Since  $D = 1$ , this can be written in matrix form as  $\underline{\mathbf{S}} = WH$  where  $S_{0,m}$  is the  $m^{\text{th}}$  column of  $W$  and  $H_m$  is the  $m^{\text{th}}$  row of  $H$ . As we have shown in Section IV-F, the optimization of the hyperparameters  $\Theta = \{W, H\}$  through likelihood maximization is thus equivalent to the minimization of the IS distance between the power STFT  $|\underline{\mathbf{Y}}|^2$  and the product  $WH$ , yielding a NMF model as in [17], [27], and [32].

Some other kind of knowledge has now to be put into the model so that it can be useful in practice, since we have not yet made any distinction between the covariance functions of the drums components and those of the musical residual signal. A very simple and computationally cheap solution to this problem is to appropriately initialize some of the hyperparameters. In this experiment, we will focus on the meaning of the activation gains  $H_m$  of the components as introduced in (31).  $H_{im}$  can be understood as a magnitude parameter for component  $m$  into frame  $i$ . A good way to initialize all these parameters  $\{H_m\}_{m=1\dots R_d}$  for the drums signal is simply to use an *onset detector* such as [15]. Indeed, if an onset detector feature has a high magnitude in some frame  $i$ , then some drums component must be active in it. The onset detector of [15] was hence used in  $R_d$  different frequency bands of the mixture STFT, yielding  $R_d$  signals. These signals were used to initialize the activation gains  $\{H_m\}_{m=1\dots R_d}$  and all the other hyperparameters of the model were randomly initialized. A NMF was then applied using this initialization and separation was performed using (32).

2) *Results*: The proposed GP separation method was tested on ten 30-s excerpts sampled at 44.1 kHz from the Quaero<sup>16</sup> source separation corpus. The excerpts featured many different kinds of music signals, including pop, electro pop, rock, reggae, and bossa. For each of these excerpts, the ground truth drums and musical residual signals are known for evaluation but the separation systems can only observe their mixtures. On average, the relative amplitude  $20 \log_{10} \frac{\sum_x |s_r(x)|}{\sum_x |s_d(x)|}$  of the musical residual signal was set to +6 dB compared to the drums signal.

We applied the method proposed by GILLET and RICHARD in [18] on the same mixtures and the quality of the results were quantified through the BSSEVAL toolbox [42]. The separation quality was evaluated both on the drums signals and on the musical residual signal.

The metrics obtained through BSSEVAL include the source to distortion ratio (SDR), the source to artifact ratio (SAR) and the Source to interference ratio (SIR) that are all expressed in dB. Whereas the SDR is a global measure of separation quality, the SAR and SIR, respectively, measure the amount of separation/reconstruction artifacts and the amount of energy from the other sources. Results are given in Fig. 4.

From Fig. 4, we can see that the GP model presented in Section V-B-1 very well manages to separate the drums and musical residual signals on many different kinds of music and both signals are well recovered. A further feature of this technique is that it is extremely fast: on average, 30 seconds are needed to handle a 30-s long excerpt whereas 300 s are needed

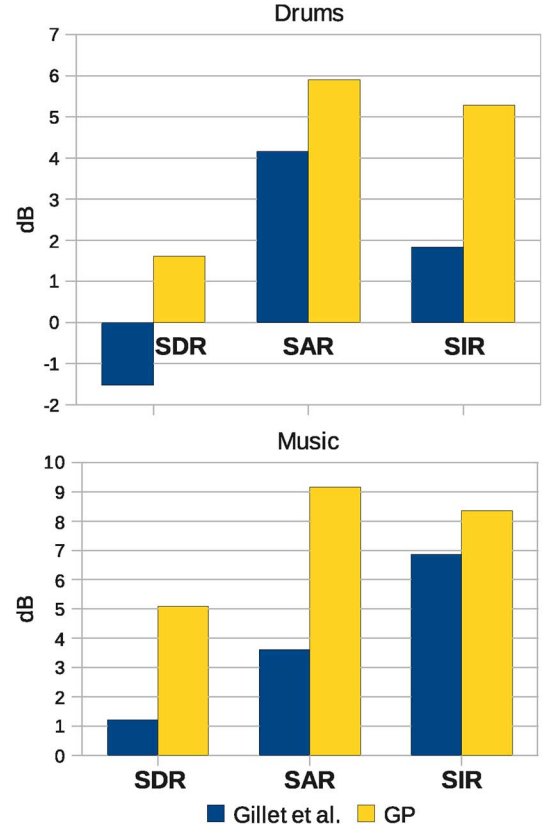


Fig. 4. Evaluation of the separation quality for the extraction of the drums track (top) and the musical residual signal (bottom). Higher is better.

by [18]. Sound excerpts and a full implementation in Python of this separation technique are freely available on our website.<sup>17</sup>

## VI. CONCLUSION

In this paper, we have stated the linear underdetermined, instantaneous, convolutive, and multiple-output source separation problems in terms of Gaussian processes regression and have shown that it leads to simple formulas to optimally proceed to signals separation w.r.t. the MMSE. The advantages of setting out the source separation problem in terms of GP are numerous.

First, there is neither notational burden nor any conceptual issue raised when using input spaces  $\mathcal{X}$  different from  $\mathbb{R}$  or  $\mathbb{Z}$ , thus enabling a vast range of source separation problems to be handled within the same framework. Multidimensional signal separation may include audio, image or video sensor arrays as well as geostatistics.

Second, GP source separation can perfectly be used for the separation of non locally stationary signals. Of course, some important simplifications of the computations as presented in Sections IV-D and IV-E are lost when using non-stationary covariance functions. Still, the frame-wise independence assumption presented in Section IV-C may nonetheless be used in order to make the estimations computationally tractable.

Third, it provides a coherent probabilistic way to take many sorts of relevant prior information into account. Indeed, prior information is encapsulated in the choice of the covariance

<sup>16</sup><http://www.quaero.org>.

<sup>17</sup><http://www.telecom-paristech.fr/~liutkus/GPSS/>.

functions and the framework proposed here thus clearly distinguishes between the optimal separation methods and the particular models considered.

Finally, we have seen that under appropriate assumptions, optimization of the hyperparameters of a GP model is equivalent to a classical NTF using the Itakura-Saito divergence on the spectrogram tensor of the mixtures, thus enabling efficient estimation of the hyperparameters.

Setting the source separation problem in such a unified framework allows it to be considered from a larger perspective where its objective is to separate additive independent functions on arbitrary input spaces that are mathematically characterized by their first and second moments only.

## REFERENCES

- [1] P. Abrahamsen, "A review of Gaussian random fields and correlation functions," Norsk Regnesentral, Oslo, Norway, Tech. Rep. 878, Apr. 1997.
- [2] M. Alvarez and N. D. Lawrence, "Sparse convolved Gaussian processes for multi-output regression," in *Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2008, pp. 57–64.
- [3] E. V. Bonilla, K. M. A. Chai, and C. K. I. Williams, "Multi-task Gaussian process prediction," in *Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2007, pp. 153–160.
- [4] P. Boyle and M. Frean, "Multiple output Gaussian process regression," Victoria Univ. Wellington, Apr. 2005.
- [5] P. Boyle and M. R. Frean, "Dependent Gaussian processes," in *Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2004, pp. 217–224.
- [6] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 90, pp. 2009–2026, Oct. 1998.
- [7] A. T. Cemgil, S. J. Godsill, P. H. Peeling, and N. Whiteley, "Bayesian statistical methods for audio and music processing," in *The Oxford Handbook of Applied Bayesian Analysis*. Oxford, U.K.: Oxford Univ. Press, 2010, 13: 978-0-19-954890-3.
- [8] A. T. Cemgil, P. Peeling, O. Dikmen, and S. Godsill, "Prior structures for time-frequency energy distributions," in *Proc. 2007 IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New York, Oct. 2007, pp. 151–154.
- [9] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*. New York: Wiley, Sep. 2009.
- [10] *Handbook of Blind Source Separation: Independent Component Analysis and Blind Deconvolution*, P. Comon and C. Jutten, Eds. New York: Academic, 2010.
- [11] P. J. Diggle and P. J. Ribeiro, *Model-Based Geostatistics*, ser. Springer Series in Statistics, 1st ed. New York: Springer, Mar. 2007.
- [12] O. Dikmen and A. T. Cemgil, "Gamma Markov random fields for audio source modelling," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 589–601, Mar. 2010.
- [13] O. Dikmen and A. T. Cemgil, "Unsupervised single-channel source separation using Bayesian NMF," in *Proc. 2009 IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New York, Oct. 2009, pp. 93–96.
- [14] J.-L. Durrieu, A. Ozerov, C. Févotte, G. Richard, and B. David, "Main instrument separation from stereophonic audio signals using a source/filter model," in *Proc. 17th Eur. Signal Process. Conf. (EUSIPCO)*, Glasgow, U.K., Aug. 2009, pp. 15–19.
- [15] C. Duxbury, J. P. Bello, M. Davies, and M. Sandler, "Complex domain onset detection for musical signals," in *Proc. Digit. Audio Effects Workshop (DAFx)*, London, U.K., Sep. 2003.
- [16] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Comput.*, vol. 21, no. 3, pp. 793–830, Mar. 2009.
- [17] D. FitzGerald, M. Cranitch, and E. Coyle, "On the use of the beta divergence for musical source separation," in *Proc. Irish Signal Syst. Conf. (ISSC)*, 2008.
- [18] O. Gillet and G. Richard, "Transcription and separation of drum signals from polyphonic music," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 529–540, 2008.
- [19] A. Girard, C. E. Rasmussen, J. Quiñero Candela, and R. Murray-Smith, "Gaussian process priors with uncertain inputs—Application to multiple-step ahead time series forecasting," in *Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2002, pp. 529–536.
- [20] P. W. Goldberg, C. K. I. Williams, and C. M. Bishop, "Regression with input-dependent noise: A Gaussian process treatment," in *Neural Information Processing Systems (NIPS)*, M. I. Jordan, M. J. Kearns, and S. A. Solla, Eds. Cambridge, MA: The MIT Press, 1997, pp. 493–499.
- [21] M. Helén and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," in *Proc. 13th Eur. Signal Process. Conf. (EUSIPCO)*, Antalya, Turkey, 2005.
- [22] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.
- [23] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [24] A. G. Journel and C. J. Huijbregts, *Mining Geostatistics*. London, U.K.: Academic, 1978.
- [25] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms—Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, Sep. 2008.
- [26] J. Le Roux, E. Vincent, Y. Mizuno, H. Kameoka, N. Ono, and S. Sagayama, "Consistent Wiener filtering: Generalized time-frequency masking respecting spectrogram consistency," in *Proc. 9th Int. Conf. Latent Variable Anal. Signal Separation (LVA/ICA 2010)*, St. Malo, France, Sep. 2010, pp. 89–96.
- [27] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, Apr. 2001, vol. 13, pp. 556–562.
- [28] D. MacKay, "Gaussian processes—A replacement for supervised neural networks?," in *Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 1997.
- [29] G. Matheron, "The intrinsic random functions and their applications," *Adv. Appl. Probabil.*, vol. 5, no. 3, pp. 439–468, 1973.
- [30] A. Melkumyan and F. Ramos, "Multi-kernel Gaussian processes," in *Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2009.
- [31] A. Melkumyan and F. Ramos, "A sparse covariance function for exact Gaussian process inference in large datasets," in *Proc. 21st Int. Joint Conf. Artif. Intell. (IJCAI)*, San Francisco, CA, Jul. 2009, pp. 1936–1942.
- [32] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 550–563, March 2010.
- [33] S. Park and S. Choi, "Gaussian processes for source separation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Mar. 2008, vol. 18, pp. 1909–1912.
- [34] J. Quiñero-Candela, C. E. Rasmussen, and R. Herbrich, "A unifying view of sparse approximate Gaussian process regression," *J. Mach. Learn. Res.*, vol. 6, pp. 1939–1959, Dec. 2005.
- [35] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Cambridge, MA: The MIT Press, 2005.
- [36] M. Seeger, "Gaussian processes for machine learning," *Int. J. Neural Syst.*, vol. 14, no. 2, pp. 69–106, Apr. 2004.
- [37] E. Snelson, "Local and global sparse Gaussian process approximations," in *Proc. Artif. Intell. Statist. (AISTATS)*, San Juan, Puerto Rico, Mar. 2007, vol. 2, pp. 524–531.
- [38] E. Snelson and Z. Ghahramani, "Sparse Gaussian processes using pseudo-inputs," in *Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2006, pp. 1257–1264.

- [39] J. I. Tomita and Y. Hirai, "Acoustic echo cancellation using Gaussian processes," in *Proc. 15th Int. Conf. Neural Inf. Process. (ICONIP)*, 2008, vol. 5507, Lecture Notes in Comput. Sci., pp. 353–360.
- [40] J. Vanhatalo and A. Vehtari, "Modelling local and global phenomena with sparse Gaussian processes," in *Proc. 24th Conf. Uncertainty Artif. Intell. (UAI)*, Helsinki, Finland, Jul. 2008, pp. 571–578.
- [41] E. Vincent, N. Bertin, and R. Badeau, "Adaptive harmonic spectral decomposition for multiple pitch estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 528–537, Mar. 2010.
- [42] E. Vincent, C. Févotte, and R. Gribonval, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [43] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications*. Cambridge, MA: MIT Press, 1949.
- [44] C. K. I. Williams, "Prediction with Gaussian processes: From linear regression to linear prediction and beyond," in *Learning and Inference in Graphical Models*, M. I. Jordan, Ed. New York: Kluwer, 1999.



**Antoine Liutkus** was born in France on February 23, 1981. He received the State Engineering degree from Telecom ParisTech, France, in 2005, and the M.Sc. degree in acoustics, computer science, and signal processing applied to music (ATIAM) from the Université Pierre et Marie Curie (Paris VI), Paris, in 2005.

He worked as a research engineer on source separation at Audionamix from 2007 to 2010 and is currently pursuing the Ph.D. degree with the Department of Signal and Image Processing, Telecom ParisTech. His research interests include statistical

music processing, source separation and machine learning methods applied to signal processing.



**Roland Badeau** (M'02–SM'10) was born in Marseille, France, on August 28, 1976. He received the State Engineering degree from the École Polytechnique, Palaiseau, France, in 1999, the State Engineering degree from the École Nationale Supérieure des Télécommunications (ENST), Paris, France, in 2001, the M.Sc. degree in applied mathematics from the École Normale Supérieure (ENS), Cachan, France, in 2001, and the Ph.D. degree from the ENST in 2005, in the field of signal processing.

He received the ParisTech Ph.D. Award in 2006, and the Habilitation à Diriger des Recherches degree from the Université Pierre et Marie Curie (UPMC), Paris VI, in 2010.

In 2001, he joined the Department of Signal and Image Processing, Telecom ParisTech (ENST), as an Assistant Professor, where he became Associate Professor in 2005. From November 2006 to February 2010, he was the manager of the DESAM project, funded by the French National Research Agency (ANR), whose consortium was composed of four academic partners. His research interests include high resolution methods, adaptive subspace algorithms, nonnegative factorizations, audio signal processing, and musical applications. He is the author of 17 journal papers and 40 international conference papers.

Dr. Badeau is a Chief Engineer of the French Corps of Mines (foremost of the great technical corps of the French state).



**Gaël Richard** (SM'06) received the State Engineering degree from Telecom ParisTech, France (formerly ENST) in 1990, the Ph.D. degree from LIMSI-CNRS, University of Paris-XI, in 1994 in speech synthesis, and the Habilitation à Diriger des Recherches degree from the University of Paris XI in September 2001.

After the Ph.D. degree, he spent two years with the CAIP Center, Rutgers University, Piscataway, NJ, in the Speech Processing Group of Prof. J. Flanagan, where he explored innovative approaches for speech production. From 1997 to 2001, he successively worked for Matra, Bois d'Arcy, France, and for Philips, Montrouge, France. In particular, he was the Project Manager of several large scale European projects in the field of audio and multimodal signal processing. In September 2001, he joined the Department of Signal and Image Processing, Telecom ParisTech, where he is now a Full Professor in audio signal processing and Head of the Audio, Acoustics, and Waves research group. He is a coauthor of more than 80 papers and inventor in a number of patents. He is also one of the experts of the European commission in the field of speech and audio signal processing.

Prof. Richard is a member of the EURASIP and an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.