

# MULTI-DIMENSIONAL SIGNAL SEPARATION WITH GAUSSIAN PROCESSES

Antoine Liutkus, Roland Badeau, Gaël Richard

Institut Telecom, Telecom ParisTech, CNRS LTCI, France.

## ABSTRACT

Gaussian process (GP) models are widely used in machine learning to account for spatial or temporal relationships between multivariate random variables. In this paper, we propose a formulation of underdetermined source separation in multidimensional spaces as a problem involving GP regression. The advantage of the proposed approach is firstly to provide a flexible means to include a variety of prior information concerning the sources and secondly to lead to minimum mean squared error estimates. We show that if the additive GPs are supposed to be *locally-stationary*, computations can be done very efficiently in the frequency domain. These findings establish a deep connection between GP and nonnegative tensor factorizations with the Itakura-Saito distance and we show that when the signals are monodimensional, the resulting framework coincides with many popular methods that are based on nonnegative matrix factorization and time-frequency masking.

**Index Terms**— Gaussian Processes, Nonnegative Tensor Factorization, Source Separation, Probability Theory, Regression

## 1. INTRODUCTION

Gaussian processes (GPs) [10, 11] are commonly used to model functions whose mean and covariances are known. Given some learning points, they permit to estimate the values taken by the function at any other points of interest. Their advantages are to provide a simple and effective probabilistic framework for regression and classification as well as an effective means to optimize models' parameters. They are thus widely used in many areas and their use can be traced back at least to works by Wiener in 1941 [12].

Source separation is another very intense field of research (see [6] 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. 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. Among the most popular approaches, we can mention Independent Component Analysis [2] that relies both on the probabilistic independence between the source signals and on higher order statistics. We can also cite Non-negative Matrix Factorization (NMF) source separation that models the sources as locally stationary with constant normalized power spectra and time-varying energy [8, 7]. Most of the methods devised so far focus on 1-dimensional signals.

This work is partly funded by the French National Research Agency (ANR) as a part of the DReaM project (ANR-09-CORD-006-03) and partly supported by the Quaero Program, funded by OSEO, French State agency for innovation.

In this study, we consider the case of additive signals defined on multidimensional input spaces and revisit underdetermined source separation as a problem involving GP regression. Due to its heavy computational burden, the GP framework has to come along with effective methods to simplify the computations in order to be of practical use. For the special case of locally stationary and regularly sampled signals, we 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 Non-negative Tensor Factorization (NTF) [5] using the Itakura-Saito divergence.

The article is organized as follows: first, we present the use of GP regression for source separation in section 2. Then, we show in section 3 that when the GPs are assumed *separable* and *locally-stationary*, computations can be done very efficiently in the frequency domain. We finally illustrate the framework through a simple toy example in section 4 and give some conclusions in section 5.

## 2. GAUSSIAN PROCESSES FOR SOURCE SEPARATION

### 2.1. Introduction to Gaussian processes

A Gaussian process [10, 11] 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 in  $\mathbb{R}$ , such that for any *finite* set of inputs  $X = \{x_1 \cdots x_n\} \in \mathcal{X}^n$ ,  $\mathbf{f} \triangleq [f(x_1) \cdots f(x_n)]^\top$  is distributed according 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  $k_f(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x')))]$ . In this study, we will consider centered signals, i.e.  $\forall x \in \mathcal{X}, m(x) = 0$ .

Let  $X$  be a finite set of elements from  $\mathcal{X}$ . The covariance matrix  $K_{f,XX}$  is defined as  $[K_{f,XX}]_{i,j} = k_f(x_i, x_j)$  and the probability of  $\mathbf{f}$  given  $X$  is then given by:

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

which is usually written  $f \sim \mathcal{GP}(0, k_f(x, x'))$ .  $|K_{f,XX}|$  is the determinant of  $K_{f,XX}$ .

The covariance function  $k_f$  of a GP  $f$  must be such that for any  $X$ , the covariance matrix  $K_{f,XX}$  is positive definite<sup>2</sup>. Such a function is called a positive definite function [1]. 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. More generally, covariance functions are parameterized by a set of scalar values that are often called *hyperparameters* and are usually gathered in a *hyperparameter set*  $\Theta$ .

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

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

## 2.2. Source separation with Gaussian processes

Suppose we observe the sum  $y(x)$  of  $M$  signals  $f_m(x)$ :  $y(x) = \sum_{m=1}^M f_m(x)$  for a finite set  $X$  of  $n$  input points from  $\mathcal{X}$ :  $X = \{x_1 \cdots x_n\}$ . We assume that the  $\{f_m(x)\}_{m=1 \dots M}$  are independent and we want to estimate the values taken by one of the signals  $f_{m_0}$  for  $m_0 \in (1 \cdots M)$  on a finite and possibly different set  $X^* = \{x_1^* \cdots x_{n^*}^*\}$  of  $n^*$  input points from  $\mathcal{X}$ .<sup>3</sup> Let us furthermore assume that  $\forall m, f_m \sim \mathcal{GP}(0, k_m(x, x'))$  where the covariance functions  $k_m$  are known. As  $\{f_m(x)\}_{m=1 \dots M}$  are supposed independent, we have:

$$\sum_{m=1}^M f_m \sim \mathcal{GP}\left(0, \sum_{m=1}^M k_m(x, x')\right). \quad (2)$$

Let  $K_{m, XX^*}$  be the covariance matrix defined by  $[K_{m, XX^*}]_{ij} = k_m(x_i, x_j^*)$ . We define  $K_{m, X^*X}$ ,  $K_{m, X^*X^*}$  in the same way. Let  $\mathbf{f}_m \triangleq [f_m(x_1) \cdots f_m(x_n)]^\top$ ,  $\mathbf{f}_m^* \triangleq [f_m(x_1^*) \cdots f_m(x_{n^*}^*)]^\top$  and similarly for  $\mathbf{y}$ . We have:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_{m_0}^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sum_{m=1}^M K_{m, XX} & K_{m_0, XX^*} \\ K_{m_0, X^*X} & K_{m_0, X^*X^*} \end{bmatrix}\right).$$

Classical probability results then assert that the conditional distribution of  $\mathbf{f}_{m_0}^*$  given  $\mathbf{y}$  is (see [10])  $\mathbf{f}_{m_0}^* | \mathbf{y} \sim \mathcal{N}(\overline{\mathbf{f}}_{m_0}^*, \text{cov}\mathbf{f}_{m_0}^*)$  with<sup>4</sup>:

$$\overline{\mathbf{f}}_{m_0}^* = K_{m_0, X^*X} \left[ \sum_{m=1}^M K_{m, XX} \right]^{-1} \mathbf{y} \quad (3)$$

and

$$\text{cov}\mathbf{f}_{m_0}^* = K_{m_0, X^*X^*} - K_{m_0, X^*X} \left[ \sum_{m=1}^M K_{m, XX} \right]^{-1} K_{m_0, XX^*}. \quad (4)$$

The minimum mean squared error (MMSE) estimate  $\hat{\mathbf{f}}_{m_0}^*$  of  $\mathbf{f}_{m_0}^* | \mathbf{y}$  is thus found by setting  $\hat{\mathbf{f}}_{m_0}^* = \overline{\mathbf{f}}_{m_0}^*$ . A problematic issue with this method is the requirement to invert an  $n \times n$  covariance matrix. This  $\mathcal{O}(n^3)$  computational cost is often prohibitive.

## 3. EFFICIENT COMPUTATIONS FOR LARGE SIGNALS

In this section, we assume that the signals are defined on  $\mathcal{X} = \mathbb{R}^D$  for  $D \geq 1$  and that  $x_i \in \mathcal{X}$  can be written  $x_i = (x_{i,1}, \dots, x_{i,D})$ . 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  $\forall (x_i, x_j) \in \mathcal{X}^2, k(x_i, x_j) = \prod_{d=1}^D k^{(d)}(x_{i,d}, x_{j,d})$ . This assumption implies that all the covariance matrices  $K$  considered can be expressed as a Kronecker product (see [5]) of  $D$  covariance matrices  $K^{(d)}$  of lower dimensions:

$$K = K^{(1)} \otimes K^{(2)} \cdots \otimes K^{(D)} \triangleq \bigotimes_{d=1}^D K^{(d)}. \quad (5)$$

From now on, we suppose that the points are regularly sampled. This is equivalent to assuming that any signal  $\mathbf{y}$ ,  $\mathbf{f}_m$  or  $\mathbf{k}$  considered is the vectorization of a corresponding underlying  $D$ -dimensional tensor  $\underline{y}$ ,  $\underline{f}_m$  or  $\underline{k}$ .<sup>5</sup>

<sup>3</sup>In source separation,  $X$  and  $X^*$  are usually equal and correspond to regularly spaced points.

<sup>4</sup>In the case of singular covariance matrix  $\sum_{m=1}^M K_{m, XX}$ , numerical methods such as Moore-Penrose pseudo-inversion may be used.

<sup>5</sup>Vectorization is done recursively. For example, with  $D = 2$  where tensors are matrices, it is done one row after the other.

## 3.1. Stationarity assumption

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  $\mathbf{y}$  is given by:  $K_y = \sum_{m=1}^M K_m$  where  $K_m$  is the covariance matrix of source  $m$ . Considering (5), it is given by:  $K_m = \bigotimes_{d=1}^D K_m^{(d)}$  where  $[K_m^{(d)}]_{i,j} = k_m^{(d)}(x_{i,d} - x_{j,d})$  can approximately be considered as circulant<sup>6</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>7</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  $\tau \mapsto k_m^{(d)}(x + \tau, x)$ . We can thus write  $K_y$  as:

$$K_y = \sum_{m=1}^M \bigotimes_{d=1}^D W_F^* \text{diag}S_m^{(d)} W_F. \quad (6)$$

Using classical results from tensor algebra, We can show that (3) can be written using (6) as<sup>8</sup>:

$$\overline{\mathbf{f}}_{m_0}^* = \left( \bigotimes_{d=1}^D W_F^* \right) \left( \frac{\bigotimes_{d=1}^D \text{diag}S_{m_0}^{(d)}}{\sum_{m=1}^M \bigotimes_{d=1}^D \text{diag}S_m^{(d)}} \right) \left( \bigotimes_{d=1}^D W_F \right) \mathbf{y}. \quad (7)$$

Introducing the  $D$ -dimensional tensor<sup>9</sup>

$$\underline{S}_m = S_m^{(1)} \circ S_m^{(2)} \cdots \circ S_m^{(D)} \triangleq \bigcirc_{d=1}^D S_m^{(d)} \quad (8)$$

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 (7) in tensor form as:

$$\mathcal{F}_D \left\{ \overline{\mathbf{f}}_{m_0}^* \right\} = \left( \frac{\underline{S}_{m_0}}{\sum_{m=1}^M \underline{S}_m} \right) \cdot \mathcal{F}_D \{\underline{y}\} \quad (9)$$

which is similar to the classical Wiener filter for stationary processes. The differences between this expression and the classical one is firstly that it is valid for any dimension  $D$  of the input space and secondly 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  $D$ -dimensional Fourier transform of the stationary covariance function  $\tau \mapsto k_m(x + \tau, x)$ . The complexity of this *approximate* GP inference method relying on stationarity of the covariance functions and on regular sampling is  $\mathcal{O}(n \log n)$ . If  $\mathcal{F}_D \{\underline{y}\}$  is known beforehand, the complexity of (9) decreases to  $\mathcal{O}(n)$ , which is remarkable compared to the  $\mathcal{O}(n^3)$  operations required by the basic GP setup presented in section 2.2.

<sup>6</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  $\frac{n_d}{p}$  with  $p \in \mathbb{N}^*$ .

<sup>7</sup> $W_F^*$  denotes the complex conjugate of  $W_F$ .

<sup>8</sup> $\frac{A}{B}$  and  $A.B$  are respectively the element-wise division and multiplication of  $A$  and  $B$ .

<sup>9</sup> $\circ$  denotes the outer product.

### 3.2. Frame-wise correlations

#### 3.2.1. Frames and locally dependent signals

In many areas of interest, we cannot handle matrices of size  $n \times n$  where  $n$  is the number of observations. In audio signal processing, it is common to split the signal into overlapping *frames* and to process the frames independently. The original signal can then be recovered through a deterministic *overlap-add* procedure. This idea can very well be generalized in any dimension  $D$ . Instead of considering the original signal  $y$ , we can split it into  $n_I$  overlapping frames  $\{y_i(x)\}_{i=1 \dots n_I}$  of dimension  $L_1 \times L_2 \times \dots \times L_D$ . A common assumption is to consider that the different frames are independent, and thus that the signals are only locally-correlated.

#### 3.2.2. Separation of locally stationary Gaussian Processes

Let  $\{y(x)\}_{x \in \mathcal{X}}$  be a particular signal, observed on a finite input set  $X \in \mathcal{X}^n$  and let  $\{y_i\}_{i=1 \dots n_I}$  be a set of  $n_I$  corresponding frames. We assume that the frames are independent and further suppose that the covariance function  $k_{im}$  of source  $m$  within frame  $i$  is stationary. Each source is thus composed of several small stationary frames, each of which has its own covariance function.

Let us denote  $\underline{\mathbf{Y}}$  the  $(D+1)$ -dimensional tensor whose last dimension goes over the frames index 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 3.1. As this tensor  $\underline{\mathbf{Y}}$  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 tensors  $\underline{\mathbf{F}}_m$  of the sources and the model tensor  $\underline{\mathbf{S}}_m$  of source  $m$  in the same way. We can use (9) for each frame and for source  $m_0$ : the MMSE estimate  $\underline{\mathbf{F}}_{m_0}^*$  of  $\underline{\mathbf{F}}_{m_0}$  is then given by:

$$\underline{\mathbf{F}}_{m_0}^* = \frac{\underline{\mathbf{S}}_{m_0}}{\sum_{m=1}^M \underline{\mathbf{S}}_m} \cdot \underline{\mathbf{Y}}. \quad (10)$$

The sources can then be recovered by first applying an inverse  $D$ -dimensional Fourier transform to the estimate (10) for each frame, and then using the *overlap-add* scheme mentioned in section 3.2 to obtain the estimated sources in the original input space  $\mathcal{X}$ .

Similarly, the marginal likelihood  $\log p(\mathbf{y} | X)$  of the observations can be shown to be equal (up to an additive constant independent of the models  $\Theta = \{\underline{\mathbf{S}}_1, \dots, \underline{\mathbf{S}}_M\}$  of the sources) to  $-\frac{1}{2} D_{\text{IS}}(\underline{\mathbf{Y}} | \sum_{m=1}^M \underline{\mathbf{S}}_m)$  where  $D_{\text{IS}}(\underline{\mathbf{x}} | \underline{\mathbf{y}})$  is the Itakura-Saito (IS) divergence [7] between tensors  $\underline{\mathbf{x}}$  and  $\underline{\mathbf{y}}$ .<sup>10</sup> This expression can be computed in  $\mathcal{O}(n)$  operations when  $\underline{\mathbf{Y}}$  is known.

### 3.3. Putting structures over the covariances

As highlighted by CEMGIL *et al.* in [3, 4] 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  $\underline{\mathbf{S}}_m$ . A solution is to suppose deterministic structures 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 frames are *locally scaled* and thus identical up to an amplification gain depending on the frame. This can be written  $\underline{\mathbf{S}}_{im} = H_{im} \underline{\mathbf{S}}_{0,m}$  where  $i \in \{1 \dots n_I\}$  and  $\underline{\mathbf{S}}_{0,m} \triangleq \bigcirc_{d=1}^D S_{0,m}^{(d)}$  is the  $D$ -dimensional Fourier transform

$$^{10} D_{\text{IS}}(\underline{\mathbf{x}} | \underline{\mathbf{y}}) \triangleq \sum_{i_1 \dots i_{D+1}} \left[ \frac{[\underline{\mathbf{x}}]_{i_1 \dots i_{D+1}}}{[\underline{\mathbf{y}}]_{i_1 \dots i_{D+1}}} - \log \frac{[\underline{\mathbf{x}}]_{i_1 \dots i_{D+1}}}{[\underline{\mathbf{y}}]_{i_1 \dots i_{D+1}}} - 1 \right]$$

of some *template* covariance function  $k_{0,m}$  for source  $m$  that is independent of the frame index  $i$ . We get:

$$\underline{\mathbf{S}}_m = \left( \bigcirc_{d=1}^D S_{0,m}^{(d)} \right) \circ H_m \quad (11)$$

where  $H_m = (H_{1m} \dots H_{n_I m})$  denotes the amplification gains of the covariance function for source  $m$  on the different frames. Considering (11) we readily see that it is equivalent to a classical Nonnegative Tensor Factorization (NTF) model called Canonical Polyadic (CP) decomposition<sup>11</sup>. The different parameters become  $\Theta = \left\{ \left\{ H_m, S_{0,m}^{(1)} \dots S_{0,m}^{(D)} \right\}_{m=1 \dots M} \right\}$  and can be estimated by standard CP algorithms using the IS-divergence function [5, 7].

## 4. TOY EXAMPLE

In this section, we set  $D=2$  and  $M=2$ , which means that we aim at separating two additive functions  $f_1(x_1, x_2)$  and  $f_2(x_1, x_2)$  defined on the plane and summed in an observable mixture signal<sup>12</sup>  $y(x_1, x_2)$ . Following the notation introduced in section 3.1, we will thus suppose that the mixture tensor  $\underline{\mathbf{y}}$  is the sum of two source tensors  $\underline{\mathbf{f}}_1$  and  $\underline{\mathbf{f}}_2$ . The corresponding vectors  $\mathbf{y}$ ,  $\mathbf{f}_1$  and  $\mathbf{f}_2$  will denote the vectorization of these tensors one row after the other.

In this experimental setup, the dimensions of the sources and mixture tensors are  $500 \times 500$  each, leading to  $n=250000$ . In the following, we will assume that the covariance function  $k_m^{(d)}(x_d, x'_d)$  of each source  $m$  along each dimension  $d$  is stationary and given by:

$$k_m^{(d)}(x_d, x'_d) = \exp \left( -\frac{2 \sin^2 \frac{\pi(x_d - x'_d)}{T_{m,d}}}{l_{m,d}^2} - \frac{(x_d - x'_d)^2}{2\lambda_{m,d}^2} \right) \quad (12)$$

where  $\{T_{m,d}, l_{m,d}, \lambda_{m,d}\}_{m,d}$  are scalar parameters. This model implements a particular prior knowledge where each source  $m$  is known to be pseudo-periodic of period  $(T_{m,1}, T_{m,2})$  and  $(l_{m,1}, l_{m,2})$  controls its smoothness within one period. A further lengthscale  $(\lambda_{m,1}, \lambda_{m,2})$  controls the global covariance between two input points. In the particular example illustrated in Figure 1, the parameters were:

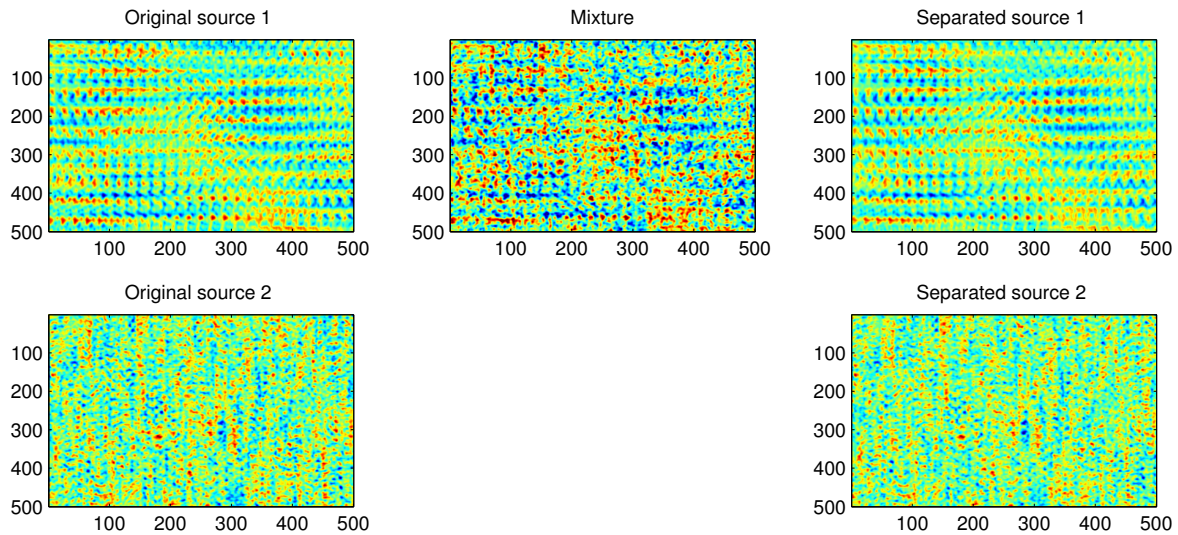
$m$	$\lambda_{m,1}$	$\lambda_{m,2}$	$T_{m,1}$	$T_{m,2}$	$l_{m,1}$	$l_{m,2}$
1	100	100	50	20	0.5	0.7
2	40	4	25	$+\infty$	0.7	N/A

In this very simple experimental setup and for each experiment, we first synthesize the sources using those hyperparameters. During the separation step, we consider that only the synthesized mixture and the hyperparameters are known and we can perform separation through the method presented in section 3.2.2. To this purpose, we can build the spectral covariance tensor  $\underline{\mathbf{S}}_m$  of each source and then perform separation in the frequency domain as in (10). The sources are recovered through an inverse 2-dimensional Fourier transform. It is worth noticing here that the computations only involve element-wise multiplications of  $500 \times 500$  images, instead of the inversion of the  $250000 \times 250000$  covariance matrix required by the basic GP setup. Results for one example are shown in Figure 1. The average Signal to Error Ratio (SER) obtained on 50 experiments was 8dB and the average computing time was less than  $2s$ <sup>13</sup>.

<sup>11</sup>CP is also called PARAFAC or CANDECOMP [5].

<sup>12</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. Estimating the value of the target signal through Kriging is hence a special case of GPSS.

<sup>13</sup>Since exact computation is not tractable in this case, we also performed



**Fig. 1.** GP for the separation of two stationary random fields ( $D = 2$ ) using GPs. On the left are the original sources. On the center is the mixture and on the right are the estimated sources. A temperature colormap is used : blue indicates large negative values, red indicates large positive values.

## 5. CONCLUSION

In this study, we have stated the linear underdetermined source separation problem in terms of GP regression and we have shown that it leads to simple formulas to optimally proceed to signals separation with respect to the MMSE criterion. We have furthermore noted connections between GP models and Nonnegative Tensor Factorizations when the mixtures are regularly sampled and the sources are locally stationary. It is noticeable that the proposed framework becomes equivalent to popular NMF methods when the signals are 1-dimensional.

Setting the source separation problem in such a unified framework permits to consider it 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. More information on this topic can be found in [9].

## 6. REFERENCES

- [1] P. Abrahamsen. A review of Gaussian random fields and correlation functions. Technical Report 878, Norsk Regnesentral, Oslo, Norway, April 1997.
- [2] J.-F. Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE*, 90:2009–2026, October 1998.
- [3] A. T. Cemgil, S. J. Godsill, P. H. Peeling, and N. Whiteley. *The Oxford Handbook of Applied Bayesian Analysis*, chapter Bayesian Statistical Methods for Audio and Music Processing. Number ISBN13: 978-0-19-954890-3. Oxford University Press, 2010.
- [4] A.T. Cemgil, P. Peeling, O. Dikmen, and S. Godsill. Prior structures for Time-Frequency energy distributions. In *Proc. of the 2007 IEEE Workshop on. App. of Signal Proc. to Audio and Acoust. (WASPAA'07)*, pages 151–154, NY, USA, October 2007.
- [5] 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*. Wiley Publishing, September 2009.
- [6] P. Comon and C. Jutten, editors. *Handbook of Blind Source Separation: Independent Component Analysis and Blind Deconvolution*. Academic Press, 2010.
- [7] C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis. *Neural Computation*, 21(3):793–830, March 2009.
- [8] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems (NIPS)*, volume 13, pages 556–562. MIT Press, April 2001.
- [9] A. Liutkus, R. Badeau, and G. Richard. Gaussian processes for underdetermined source separation (to be published). *Signal Processing, IEEE Transactions on*, PP(99):1, 2011.
- [10] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.
- [11] M. Seeger. Gaussian processes for machine learning. *Int. J. Neural Syst.*, 14(2):69–106, April 2004.
- [12] N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series with engineering applications*. MIT Press, 1949.

the experiment for  $50 \times 50$  signals and obtained an average 4.2dB SER for 0.5s average computing time, compared to 5.7dB for exact inference in approximately 45s.