Fractional Component Analysis (FCA) for Mixed Signals

Asanobu KITAMOTO

Research Center for Testbeds and Prototyping, National Institute of Informatics (NII)
E-mail: kitamoto@nii.ac.jp || http://research.nii.ac.jp/˜kitamoto/

Abstract

This paper proposes the fractional component analysis (FCA), whose goal is to decompose the observed signal into component signals and recover their fractions. The uniqueness of our idea in comparison with other similar methods is the concept of the virtual PDF (probability distribution function) that models signal mixing on the sensor. In this paper, we derive the virtual PDF based on positivity constraint, unity constraint, and randomness assumption, and we then build it into the mixture density model. In order to learn parameters of this model from data using EM (Expectation-Maximization) algorithm, the key point is to derive the approximation of the virtual PDF using its cumulants. Finally we illustrate experimental results on synthetic data to show the unique decision boundary obtained from our method.

1. Introduction

This paper aims to propose "yet another" method for decomposing the observed signal into unobservable component signals. Our unique approach to this popular problem is to assume two levels of signal mixing, namely *instantaneous mixing* and *density mixing*. The former type of mixing leads to the concept of the *virtual PDF*, while the latter leads to a popular mixture density model to be solved by a variant of EM algorithm [4]. All of the procedures are developed based on probabilistic models and statistical learning.

Our motivation behind this method is the problem of *signal unmixing*, a popular problem in the remote sensing community. Researchers in the field are always frustrated by the presence of heterogeneous pixels that may contain multiple *endmembers* (components) within a single pixel, and they want to estimate components contained in a pixel with their fractions. Here it is important to have in mind that signal mixing depends on how we observe the real world, or, more specifically, on the resolution of the sensor.

We begin with the model of instantaneous mixing in Section 2 and the notion of the virtual PDF. Then we proceed to the model of density mixing in Section 3, in which we introduce the *fractional component analysis (FCA)* and discuss learning issues. Finally in Section 4, we illustrate experimental results and compare our method with similar ones.

2. Instantaneous Mixing

2.1. The Basic Model and Prerequisites

We start our discussion with *instantaneous mixing*, which refers to signal mixing at a single observation instance. For the time being, we are interested in the estimation of a set of components involved in a particular observation instance, and also component fractions. We assume that component fractions are "sparse" (most of the component fractions are zero) in the sense that only a few components are usually involved in instantaneous mixing at a single observation instance. An observation instance may even be "pure", where only a single component is observed without signal mixing. We therefore pursue a generative model capable of generating mixed signals from the combination of pure component signals.

Due to the limited space of the paper, we only consider a linear model for instantaneous mixing as follows:

$$\mathbf{x}(t) = \sum_{m=1}^{M} a_m(t)\mathbf{s}_m(t) + \mathbf{n}(t) = S(t)\mathbf{a}(t) + \mathbf{n}(t). \quad (1)$$

where the observation signal at a sample index t, $\mathbf{x}(t) \in \mathcal{R}^N$, consists of the linear combination of two types of latent variables; namely, pure component signals $\mathbf{s}_m(t) \in \mathcal{R}^N$ weighted by the component fraction, $\mathbf{a}(t) \in \mathcal{R}^M$, where $a_m(t) \in \mathcal{F} = \{a_i(t)|a_i(t) > 0 \quad \text{and} \quad \sum_{i=1}^{M} a_i(t) = 1, \quad \text{for} \quad \forall t\}$, defined for each component $C_m \in C = \{C_1, \ldots, C_M\}$. In a matrix notation, $S(t) \in \mathcal{R}^{N\times M}$ represents an observation instance of M component signals in N observation channels, such as multi-spectral images. In addition, $\mathbf{n}(t) \in \mathcal{R}^N$ is a random vector that usually describes additive noise. Here, we are interested in the special setting of Eq. (1) with the following prerequisites.

1. **Positivity constraint** $a_m(t) > 0 \quad \text{for} \quad \forall m, \forall t$
2. **Unity constraint** $\sum_{m=1}^{M} a_m(t) = 1 \quad \text{for} \quad \forall t$
3. **Randomness assumption** $a_m(t)$ and $\mathbf{s}_m(t)$ are realizations of random variables A_m and $\mathbf{S}_m$.

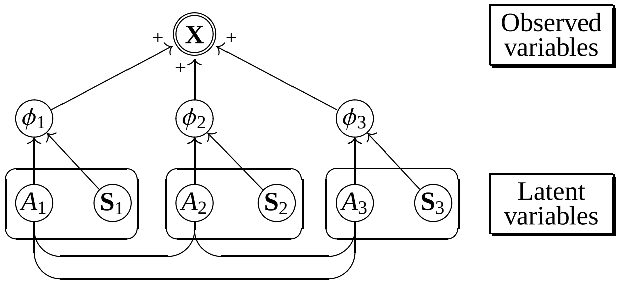


Figure 1. A graphical representation for instantaneous mixing of three components. Note that component fractions A_i are not independent because of the unity constraint.

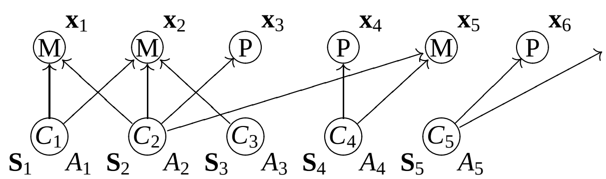


Figure 2. A graphical representation for the density mixing. 'P' and 'M' denotes a pure and a mixed observation, respectively.

Figure 1 illustrates the graphical representation of instantaneous mixing. We assume that 1) all of the latent variables – component signals and component fractions – are random variables, 2) component signals are mutually independent, and 3) the component fraction and the component signal is independent. Component fractions, however, are not independent because of the unity constraint.

2.2. Virtual PDF

We assume that instantaneous mixing occurs at each observation instance with the different combination of components, as illustrated in Figure 2. Now, suppose that we collect only observation instances generated from the instantaneous mixing of the same components. Then it is natural to assume the PDF for this set of observations, or the PDF of mixed signals. We call this PDF the *virtual PDF* in the sense that this PDF does not have a counterpart in the real world; or in other words, it is like a *virtual image* that appears as a result of signal mixing that occurs through the sensor. Hence, we cannot reach this PDF by simply searching for components existing in the real world.

To examine the virtual PDF, we first define a set of components $C = \{C_1, \ldots, C_M\}$ that appear in the whole observation instances, and represents the random variable of the i-th component as S_i and its PDF $S_i \sim p_i(s_i|\Phi_i)$, where Φ_i is the parameter vector. We then consider a subset of C to represent every possible signal mixing that may occur from C. For example, a subset of size 2 is defined as

$$\Omega_2 = \{\omega^{(2)}_{(i,j)} = \omega^{(2)}_k | C_i \in C, C_j \in C, i \neq j\}, \tag{2}$$

for $k = 1, \ldots, \binom{M}{2}$. In the same way, we may define subsets $\Omega_i, (i = 2, \cdots, M)$ with $\binom{M}{i}$ elements.

We next represent the PDF of signal mixing $\omega^{(M)}_k$. Without loss of generality, we may re-number component indices as $\omega^{(M)}_k = \{C_1, \ldots, C_M\}$. Then we obtain the PDF $p(\mathbf{x}|\omega^{(M)}_k)$ as a marginal distribution over component fractions $\mathbf{a}$:

$$p(\mathbf{x}|\omega^{(M)}_k) = \int p(\mathbf{x}, \mathbf{a}|\omega^{(M)}_k) d\mathbf{a} = \int p(\mathbf{x}|\mathbf{a}, \omega^{(M)}_k) p(\mathbf{a}) d\mathbf{a}, \tag{3}$$

where $p(\mathbf{a})$ is the prior of component fractions. To simplify the notation, we hereafter drop $\omega^{(M)}_k$. We then focus on the $p(\mathbf{x}|\mathbf{a})$ term in Eq. (3). Putting Eq. (1) yields

$$p(\mathbf{x}|\mathbf{a}) = p(a_1\mathbf{s}_1 + \cdots + a_M\mathbf{s}_M|\mathbf{a}) = p(\mathbf{s}_1, \ldots, \mathbf{s}_M|\mathbf{a}) \tag{4}$$

in which we omit the noise term $\mathbf{n}$ for brevity, although it can be naturally incorporated into our model. Independence assumptions as represented in Figure 1 further lead to

$$p(\mathbf{x}|\mathbf{a}) = \frac{1}{a_1 \cdots a_M} p_1\left(\frac{\mathbf{s}_1}{a_1}\right) * \cdots * p_M\left(\frac{\mathbf{s}_M}{a_M}\right), \tag{5}$$

where $*$ denotes convolution. Note that we have $a_i > 0$ because, in subset selection, only the components involved in signal mixing are selected. Now, a convenient tool to work with convolution is the characteristic function (CF) $\varphi(\mathbf{w}) = \int e^{j\mathbf{w}^T\mathbf{x}} p(\mathbf{x}) d\mathbf{x}$, where j is an imaginary number and the superscript T denotes the transpose. Then we can simplify the CF $\varphi_{X|A}(\mathbf{w})$ of PDF $p(\mathbf{x}|\mathbf{a})$ as

$$\varphi_{X|A}(\mathbf{w}) = \frac{1}{a_1 \cdots a_M} \prod_{m=1}^{M} a_m \varphi_m(a_m\mathbf{w}) = \prod_{m=1}^{M} \varphi_m(a_m\mathbf{w}). \tag{6}$$

Finally, we can derive $p(\mathbf{x}|\mathbf{a})$ from Eq. (6) by the inversion formula $p(\mathbf{x}|\mathbf{a}) = \frac{1}{(2\pi)^N} \int_{\mathbf{w}} e^{-j\mathbf{w}^T\mathbf{x}} \varphi_{X|A}(\mathbf{w}) d\mathbf{w}$, and we reach $p(\mathbf{x})$ after the marginalization of Eq. (3).

We can thus compute the PDF of mixed signals for any combination of components, provided that we can explicitly represent both the PDF of each component and the prior of component fractions. In addition, although this computation is theoretically possible, above procedure requires prohibitively high computational complexity in practice.

3. Density Mixing

3.1. Mixture Density Modeling

Having derived the virtual PDF, we then develop a model for density mixing. Figure 2 suggests that the observation

$$m_X^{(1)} = \frac{\sum_{m=1}^{M} \theta_m \mu_m}{\theta}, \qquad m_X^{(2)} = \frac{\sum_{m=1}^{M} (\theta_m)_1 \left(\mu_m^2 + \sigma_m^2\right) + \sum_{\substack{m,n=1\\ m\neq n}}^{M} \theta_m \theta_n \mu_n \mu_m}{(\theta)_1}$$

$$m_X^{(3)} = \frac{\sum_{m=1}^{M} (\theta_m)_2 \left(\mu_m^3 + 3\mu_m\sigma_m^2\right) + \sum_{\substack{m,n=1\\ m\neq n}}^{M} \theta_m (\theta_n)_1 \left(\mu_m\mu_n^2 + 3\mu_m\sigma_n^2\right) + \sum_{\substack{m,n,o=1\\ m\neq n\neq o}}^{M} \theta_m\theta_n\theta_o\mu_m\mu_n\mu_o}{(\theta)_2}$$

$$m_X^{(4)} = \frac{\sum_{m=1}^{M} (\theta_m)_3 \left(\mu_m^4 + 6\mu_m^2\sigma_m^2 + 3\sigma_m^4\right) + \sum_{\substack{m,n=1\\ m\neq n}}^{M} (\theta_m)_1(\theta_n)_1 \left(\mu_m^2\mu_n^2 + 6\mu_m^2\sigma_n^2 + 3\sigma_m^2\sigma_n^2\right)}{(\theta)_3} +$$
$$\frac{\sum_{\substack{m,n=1\\ m\neq n}}^{M} \theta_m(\theta_n)_2 \left(\mu_m\mu_n^3 + 6\mu_m\mu_n\sigma_n^2\right) + \sum_{\substack{m,n,o=1\\ m\neq n\neq o}}^{M} \theta_m\theta_n(\theta_o)_1 \left(\mu_m\mu_n\mu_o^2 + 6\mu_m\mu_n\sigma_o^2\right) + \sum_{\substack{m,n,o,p=1\\ m\neq n\neq o\neq p}}^{M} \theta_m\theta_n\theta_o\theta_p\mu_m\mu_n\mu_o\mu_p}{(\theta)_3}$$

Figure 3. The moments $m_X^{(k)}$ of order k of the virtual PDF in the case of one-dimensional signals. $m \neq n \neq o \neq p$ represents that none of them take equal values, and $(\theta)_n = \theta(\theta+1)\cdots(\theta+n)$.

signal is in fact composed of pure signals and mixed signals. Intuitively, this idea can be written as follows;

$$\text{Mixture Model} = \text{Real PDFs} + \text{Virtual PDFs} \qquad (7)$$

where "Real PDFs" correspond to a traditional mixture models. Formally, we represent our mixture density model $p(\mathbf{x}|\Phi,\Theta)$ corresponding to Eq. (7) as follows:

$$p(\mathbf{x}|\Phi,\Theta) = \sum_{l=1}^{M} \pi_l p_l(\mathbf{x}|\Phi_l) + \sum_{\omega_k^{(l)}\in\Omega_l}^{|\Omega_l|} \pi_{\omega_k^{(l)}} p_{\omega_k^{(l)}}(\mathbf{x}|\Phi,\Theta),$$
$$\sum_{l=1}^{M} \pi_l + \sum_{l=2}^{M} \sum_{\omega_k^{(l)}\in\Omega_l}^{|\Omega_l|} \pi_{\omega_k^{(l)}} = 1, \qquad (8)$$

where Φ are parameters of component distributions, Θ are parameters related to instantaneous mixing, and $\pi_{\omega_k}^{(l)}$ represents mixing parameters in density mixing. This mixture density model is the central model in *fractional component analysis (FCA)*, and the problem here is to learn those parameters and related latent variables from data.

For that purpose, we exploit EM algorithm [4], which is the standard algorithm for mixture density estimation. One advantage of EM algorithm for the exponential family of distributions is that parameter update rules can be factorized component-wise, which leads to computationally attractive algorithm. Our model, however, cannot take the full advantage of EM algorithm because of the following obstacles.

1. We cannot obtain a closed-form solution for the virtual PDF in general. Moreover, numerical integration involved in the computation of the virtual PDF is sometimes computationally prohibitive.
2. We cannot obtain component-wise rules for parameter updates because of the dependency of the virtual PDF on multiple components as illustrated in Figure 1.

3.2. Approximation of the Virtual PDF

To solve the first obstacle, we derive the approximation of the virtual PDF using its cumulants. The key point is to apply moment expansion to both sides of Eq. (3) to yield

$$m_X^{(l_1,\ldots,l_N)} = \int_F p(\mathbf{a}|\Theta) m_{X|A}^{(l_1,\ldots,l_N)} d\mathbf{a}. \qquad (9)$$

This shows that the moments of the virtual PDF $m_X^{(l_1,\ldots,l_N)}$ can be computed from the prior $p(\mathbf{a}|\Theta)$ and the moments of $p(\mathbf{x}|\mathbf{a})$ of the same order $m_{X|A}^{(l_1,\ldots,l_N)}$, which is usually much easier to compute than $m_X^{(l_1,\ldots,l_N)}$.

To take advantage of this property, we introduce the parametric family of PDFs with realistic descriptive power and mathematical tractability:

$$p_m(\mathbf{x}|\Phi_m) = \frac{1}{(2\pi)^{N/2}|\Sigma_m|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu_m)^T\Sigma_m^{-1}(\mathbf{x}-\mu_m)} \qquad (10)$$
$$p(\mathbf{a}|\Theta) = \frac{\Gamma\left(\sum_{m=1}^{M}\theta_m\right)}{\prod_{m=1}^{M}\Gamma(\theta_m)} \prod_{m=1}^{M} a_m^{\theta_m-1} \qquad (11)$$

That is, we introduce the normal distribution $N(\mu_m, \Sigma_m)$ as the model of component signals $p_m(\mathbf{x}|\Phi_m)$, and the Dirichlet distribution as the model of component fractions $p(\mathbf{a}|\Theta)$. We choose those families of distributions because of both mathematical reasons and empirical justification [2].

With these settings, we can derive the moments and cumulants in a closed form. Due to the limited space, we only summarize analytically derived moments in Figure 3, from which we can easily derive cumulants using such relationship as $c_X^{(1)} = m_X^{(1)}$ and $c_X^{(2)} = m_X^{(2)} - m_X^{(1)} m_X^{(1)^T}$. These cumulants are then required for approximating the virtual PDF with Gram-Charlier expansion [1]. Basically it concerns a deviation from the normal distribution $N(c_X^{(1)}, c_X^{(2)})$ having the same mean vector and covariance matrix.

3.3. EM Algorithm for Our Mixture Density Model

For the second problem, due to the dependency of the virtual PDF on multiple components, simultaneous maximization, or at least simultaneous improvement, of all parameters is required. In so doing, however, many of the advantages in using EM algorithm would be lost. Hence we ignore the interdependency of parameters over multiple components, and apply conditional maximization steps (conditional EM algorithm) [4] against Q function.

E-step: Maximize Q over the latent data parameters.

CM-step 1: Fix $\Phi^{(k)}$, $\Theta^{(k)}$, and $\arg\max Q(\Pi^{(k)})$.

CM-step 2: Fix $\Pi^{(k+1)}$, $\Theta^{(k)}$, and $\arg\max Q(\Phi^{(k)})$.

CM-step 3: Fix $\Pi^{(k+1)}$, $\Phi^{(k+1)}$, and $\arg\max Q(\Theta^{(k+1)})$.

The optimization of mixing parameters $\Pi^{(k)}$ in CM-step 1 follows the standard procedure, and it is applied to both real PDFs and virtual PDFs. In CM-step 2, the parameters of real PDFs (normal distributions) $\Phi^{(k)}$ are updated following the standard procedure, then the cumulants of the virtual PDF are updated automatically based on $\Phi^{(k+1)}$ and Figure 3. Finally, in CM-step 3, the parameters of the prior $\Theta^{(k)}$ are updated using a quasi-Newton method. However, fixing Θ to values from simulation studies or other inferences [3] often produces better results, since the optimization of $\Theta^{(k)}$ is difficult to control, and too small $\theta \sim 0$ may deteriorate the quality of Gram-Charlier approximation.

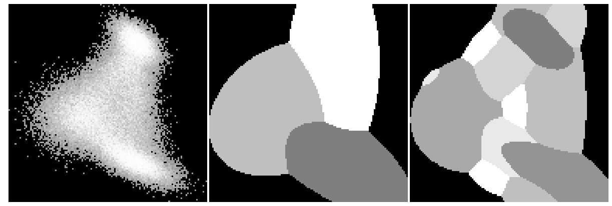
Finally, the result of mixture density estimation is used for estimating component fractions at each observation instance. We perform this task in a two-step procedure: 1) determine a set of components involved in instantaneous mixing, and 2) estimate component fractions given component members. The first part is simply based on Bayes decision rule, while the second part is based on MAP (Maximum A Posteriori) estimate as follows.

$$\mathbf{a} = \arg\max_{\mathbf{a}\in F} p(\mathbf{a}|\mathbf{x}) = \arg\max_{\mathbf{a}\in F} p(\mathbf{x}|\mathbf{a})p(\mathbf{a}) \qquad (12)$$

This maximization is performed using a kind of gradient ascent algorithm. Another choice may be the Bayesian estimate using Monte Carlo methods.

4. Experiments and Discussions

Figure 4 shows the mixture density estimation and Bayesian classification on synthetic signals in $\mathcal{R}^2$ as shown in (a) with a gray-scale. This dataset is generated from three component signals and mixed signals generated from them (three 2-component and one 3-component mixed signals). Case (b), a traditional normal mixture density model, found quadratic decision boundary, while Case (c) found more complex decision boundary between pure components (dark gray) and mixed components (light gray). In our model, the virtual PDF is constrained by the real PDFs, thus it decreases the effective complexity of the model compared to a model with the same number of components.



(a) Data (b) Real (3 PDFs) (c) Real (3 PDFs) + Virtual (4 PDFs)

Figure 4. Mixture density estimation and Bayesian classification on synthetic data.

Spectral unmixing problem has long been studied in the remote sensing community, in which the most popular approaches are based on linear algebra, such as generalized inverse matrix or constrained least-squares regression. They, however, often lack principled mechanisms to deal with uncertainty, such as probability models. They do use neural networks or fuzzy clustering to deal with uncertainty, but we believe the fractional component analysis (FCA) is an alternative principled approach which is based on probability models and learning mechanisms.

Our model also has a strong connection to independent component analysis (ICA) [1], because our model also assumes mutual independence between component signals. In particular, work by Parra et. al. [5] approaches a similar problem with similar prerequisites using ICA. However, they do not formulate their model as a generative one based on a clear distinction between pure and mixed signals.

Our current method is based on the learning of PDF over the whole signal, but it should focus more on localized instantaneous mixing. This is especially true for images, in which signal mixing occur only at a spatially localized regions. Hence our important future work is to incorporate localized or contextual information into the probability model using Markov random field or its variants.

References

[1] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, Inc., 2001.

[2] A. Kitamoto. The moments of the mixel distribution and its application to statistical image classification. In *Advances in Pattern Recognition (SPR'00)*, volume 1876 of *LNCS*, pages 521–531. Springer, 2000.

[3] A. Kitamoto and M. Takagi. Area proportion distribution — relationship with the internal structure of mixels and its application to image classification. *Syst. Comp. Japan*, 31(5):57–76, 2000.

[4] G. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, Inc, 1997.

[5] L. Parra, C. Spence, P. Sajda, A. Ziehe, and K.-R. Müller. Unmixing hyperspectral data. *Advances in Neural Information Processing Systems*, 12:942–948, 2000.