2001 年情報論的学習理論ワークショップ
2001 Workshop on Information-Based Induction Sciences (IBIS2001)
Tokyo, Japan, July 30 - August 1, 2001.

FCA: The Fractional Component Analysis

KITAMOTO Asanobu*

Abstract: The goal of the "Fractional component analysis" proposed in this paper is to analyze the observed signal generated by the linear mixing process of unknown fractional component signals. Based on simple assumptions on the probability distributions of component signals and mixing fractions, we derive a new type of distribution, which we call "mixed distribution," and characterize this model in terms of moments and cumulants. Its higher-order cumulants indicate that it is in fact either super-gaussian or sub-gaussian even if the distribution of component signals are gaussian. Finally we show methods for recovering fractional components from the observed signal.

Keywords: fractional component analysis, linear mixing process, mixed distribution, higher-order statistics, mixture density estimation

1 Introduction

This paper describes the "fractional component analysis", whose goal is to solve the inverse problem of recovering the fraction of component signals when the mixing of component signals is unknown. To solve this problem, a simple linear mixing process is assumed as the basic model:

$$\mathbf{x}(t) = A(t)\mathbf{s}(t) + \varepsilon(t), \tag{1}$$

where $\mathbf{x}(t)$, $A(t)$, $\mathbf{s}(t)$ and $\varepsilon(t)$ denotes a N-dimensional row vector called *data vector*, a (N, M)-dimensional matrix called *component matrix*, and a M-dimensional row vector called *mixing vector*, and a noise vector at index t, respectively. Here N dimensions of signals indicate N data channels that observe the same spatio-temporal location.

Although similar "linear mixing process" is widely investigated by many researchers, the uniqueness of our problem lies in the "fractional constraints." Namely, we solve Equation (1) under the constraints that the elements of the mixing vector $s_i(t)$ must satisfy the following fractional constraints,

1. **Positivity Constraint** $s_i(t) > 0$ for all i and t,
2. **Unity Constraint** $\sum_i^M s_i(t) = 1$ for every t.

These constraints are necessary for $s_i(t)$ to represent the fraction of the component class C_i at index t. Here, in a physical sense, this fraction may represent, for example, the amplitude fraction of a signal, the area fractional coverage of a pixel, the volume fractional coverage of a voxel, at the spatio-temporal location of instantaneous observation.

*National Institute of Informatics (NII),
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 JAPAN.
Tel. +81-3-4212-2578, Fax. +81-3-3556-1916 E-mail kitamoto@nii.ac.jp

2 Background

2.1 Mixed Pixel Analysis

One of the real world applications behind this problem is the "mixed pixel analysis" in image analysis [1, 2]. Mixed pixels, or mixels, are present especially on remote sensing images or medical images because of the following two reasons. Firstly, the finite resolution of the sensor causes mixed pixels composed of neighboring objects. For example, the fractal boundary of clouds produce mixed pixels composed of clouds and ocean, lands. Another type of mixed pixels is caused by semi-transparency of objects. For example, thin clouds produce mixed pixels composed of clouds themselves and objects behind. We claim that statistical properties of those heterogeneous pixels are totally different from conventional ones and hence we need to establish appropriate mathematical models to deal with them in a proper manner.

There have been numerous algorithms proposed for solving this problem. One simple approach assumes that $A(t)$ is a constant matrix for all t, neglects the noise vector $\varepsilon(t)$, then uses the Moore-Penrose generalized inverse matrix A^+ to solve directly as $\mathbf{s}(t) = A^+\mathbf{x}(t)$. However, this solution does not guarantee $\mathbf{s}(t)$ to satisfy the fractional constraints, hence sometimes additional approaches such as Lagrange multiplier or regularization are used together. Other approaches, such as fuzzy-based methods or subspace-based methods, do give fractional scalar values for component fractions, but those approaches tend to lack mathematically sound interpretation of the obtained fraction. Recently proposed non-negative matrix factorization (NMF) [3] can also

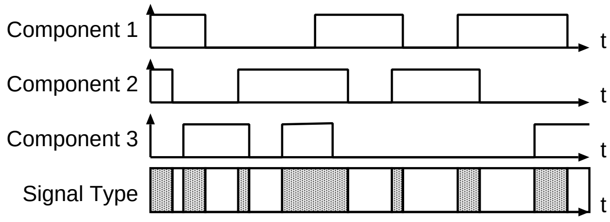

Figure 1: Schematic diagram of the activity function.

decompose the set of observations into non-negative factors, but its goal is to produce sparse representation of the signal, which is not required in our problem.

2.2 Blind Source Separation

Our problem may look similar to the blind source separation (BSS) problem. There is a widely known method called independent component analysis (ICA) for solving this problem [4]. ICA is a signal processing method to extract independent sources given only observed data that are mixtures of the unknown sources. Thus our problem shares some common features with BSS problem, but there are in fact also differences as follows:

The first difference lies in the formulation of Equation (1); in our problem it is the product of a *component matrix* and a *mixing vector*, not the product of a *mixing matrix* and a *source vector* as in BSS problem. This is because our goal is to obtain the set of fractions that satisfy the fractional constraints, and a mixing vector are more compact representation for fractions than a mixing matrix.

The second difference is that the linear mixing process is localized. In other words, linear mixing process varies for every index t, not only in terms of the value of fractions but also in terms of the number of fractional components involved in linear mixing. This assumption leads to the notion of activity function, as we address in the next section.

2.3 Activity Function

The assumption that the activity of each component is localized implies that component signals are active during small parts of the whole signal and remains silent for other parts. Then we can conceptually classify the observed signal into two types of signals; namely pure signals and mixed signals. Pure signals correspond to the case that only one component signal is active, while mixed signals, more than two component signals are active. This is a natural assumption in some image analysis problems, since each component is present for some pixels but not present for all pixels. Fig. 1 illustrates this idea by means of the notion of activity function. Each component is active for some range of index but not active for the whole signal. Pure signals, shown in white at signal type, can be observed for some range of the whole signal, and for other parts mixed signals, shown in gray at signal type, are generated as two-component mixed signals or three-component mixed signals.

Then we further advance this idea to assume that the probability density function (PDF) of the observed signal are the mixture of two types of distributions; namely pure distributions for pure signals and mixed distributions for mixed signals. Pure distributions and mixed distributions can be estimated from appropriate signal type. Here, pure-type distributions are widely used, while mixed-type distributions are hardly known. Hence we need to characterize a probability model behind mixed signals, and that model is what we call the “mixed distribution.”

3 Mixed Distribution

3.1 Definition

This section starts with the more formal definitions of the linear mixing process represented by Equation (1). Suppose that we have a set of data vectors $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))$, $(t = 1, \ldots, T)$ with N data channels, and this (mixed) data vector is constituted of M component classes C_m, $(m = 1, \ldots, M)$ with the fraction $s_m(t)$. Next we define the (N, M) component matrix $A(t) = (\mathbf{a}_1(t), \ldots, \mathbf{a}_m(t))$ and its column vector $\mathbf{a}_m(t)$ is a random vector that represents a component signal from the component class C_m subject to the class-conditional PDF $p_m(\mathbf{a_m}|C_m; \psi_m)$. We then define the mixing vector $\mathbf{s} = (s_1, \ldots, s_M) \in S$ so that its m-th element represents the fraction of the component class C_m at index t, where the space of valid fractions S is

$$S = \left\{ (s_1, \ldots, s_M) \middle| s_i > 0, \sum_{i=1}^{M} s_i = 1 \right\}. \tag{2}$$

Note that the linear mixing of M components has only $M-1$ free parameters due to the fractional constraints, and we can always set $s_M = 1 - \sum_{i=1}^{M-1} s_i$. We denote the prior of the mixing vector as $f(\mathbf{s}; \mathbf{\Theta})$ with parameters $\mathbf{\Theta}$. Finally we define the “mixed distribution” as the PDF of the observed mixed signal $p(\mathbf{x})$ generated by linear mixing process in Equation (1).

Before the derivation of the mixed distribution, here we introduce characteristic function (CF) $\varphi_{\mathbf{x}}(\mathbf{w})$ of the random variable $\mathbf{x}$, which is defined as the Fourier transform of the PDF $p(\mathbf{x})$,

$$\varphi(\mathbf{w}) = E\left\{ e^{j\mathbf{w}^T\mathbf{x}} \right\} \tag{3}$$

where $j = \sqrt{-1}$ is the imaginary unit. Now we refer to useful properties of the CF that will be used in the following. Firstly, the CF of the random variable $c\mathbf{x}$ when c is a constant can be represented as

$$\varphi_{c\mathbf{x}}(\mathbf{w}) = \varphi_{\mathbf{x}}(c\mathbf{w}). \tag{4}$$

Secondly, suppose the sum of independent random vectors $\mathbf{x} = \sum_i \mathbf{x}_i$, and the CF of $\mathbf{x}$ is obtained by the product of all CFs $\varphi_{\mathbf{x}_i}(\mathbf{w})$ involved in the mixing,

$$\varphi_{\mathbf{x}}(\mathbf{w};\Psi) = \prod_i \varphi_{\mathbf{x}_i}(\mathbf{w};\Psi_i), \tag{5}$$

where Ψ_i is the parameters of i-th random vector, and Ψ is the parameters of $\mathbf{x}$.

3.2 General Mixed Distribution

Another assumption required here is that the column vectors of the component matrix are independent random vectors, and in addition, the noise term are also assumed to be independent random vectors. Then we can derive the CF $\varphi_{\mathbf{x}}(\mathbf{w})$ of data vectors $\mathbf{x}$. We first rewrite Equation (1) into the sum of M terms:

$$\begin{aligned} \mathbf{x}(t) &= A(t)\mathbf{s}(t) + \varepsilon(t) = \sum_{m=1}^{M} \mathbf{a}_m(t)s_m(t) + \varepsilon(t) \\ &= \sum_{m=1}^{M} \mathbf{b}_m(t) + \varepsilon(t) \end{aligned} \tag{6}$$

where $\mathbf{b}_m(t) = \mathbf{a}_m(t)s_m(t)$. Next, using the property of Equation (5), we represent the CF of data vectors with parameters (Ψ, Θ);

$$\varphi_{\mathbf{x}}(\mathbf{w};\Psi,\Theta) = \prod_{m=1}^{M} \varphi_m(\mathbf{w};\psi_m,\Theta)\varphi_{\varepsilon}(\mathbf{w}) \tag{7}$$

The mixed distribution can then be derived in two steps. The first step is to assume that the mixing vector $\mathbf{s}$ has constant values, namely the mixing fractions are constant for the whole signal. Applying Equation (4) to Equation (7) yields the CF conditioned on the fixed mixing vector $\mathbf{s}$:

$$\varphi_{\mathbf{x}}(\mathbf{w}|\mathbf{s};\Psi) = \prod_{m=1}^{M} \varphi_m(\mathbf{w}s_m;\psi_m)\varphi_{\varepsilon}(\mathbf{w}) \tag{8}$$

The CF $\varphi_{\mathbf{x}}(\mathbf{w}|\mathbf{s};\Psi)$ is usually invertible and we have $\varphi^{-1}(\mathbf{x}|\mathbf{s};\Psi)$ which is derived by the inverse Fourier transform;

$$\varphi^{-1}(\mathbf{x}|\mathbf{s};\Psi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\mathbf{w}^T\mathbf{x}}\varphi_{\mathbf{x}}(\mathbf{w}|\mathbf{s};\Psi)d\mathbf{w} \tag{9}$$

Note that the PDF $\varphi^{-1}(\mathbf{x}|\mathbf{s};\Psi)$ cannot, in general, be represented in closed form using elementary functions except for cases that all PDFs $p_m(\cdot)$ belong to the same family of stable distributions. However, at least in any cases, Equation (9) can be computed numerically using FFT and numerical integration.

Mixing vectors, however, are in fact unknown and it is natural (especially in image analysis) to assume that the mixing vector is also a random vector distributed over the space of any possible combination of fractions as defined in Equation (2). According to this assumption, we randomize Equation (9) with the prior of the mixing vector represented by $f(\mathbf{s};\Theta)$ and obtain the PDF of $p(\mathbf{x})$.

$$p(\mathbf{x};\Psi,\Theta) = \int_S f(\mathbf{s};\Theta)\varphi^{-1}(\mathbf{x}|\mathbf{s};\Psi)d\mathbf{s} \tag{10}$$

It is clear from above derivation that this general procedures can deal with any types of probability models. However, if we manage to perform above procedures using numerical algorithms, the computational cost is high while the accuracy is low due to high dimensionality involved in the computation. Thus we need a standard model which can be easily computable and can be used to analyze essential properties of this type of models. For this purpose, we propose the standard model of the mixed distribution.

3.3 Standard Mixed Distribution

The standard mixed distribution is composed of two types of probability models, namely the model of component signals and the model of mixing. For the first model, we use the normal distribution (gaussian) as the class-conditional PDF of the component class C_m with parameters Ψ_m:

$$\begin{aligned} p_m(\mathbf{x}|C_m;\Psi_m) &= N(\mathbf{x};\mu_{\mathbf{m}},\Sigma_m) \\ &= \frac{e^{-\frac{1}{2}(\mathbf{x}-\mu_{\mathbf{m}})^T\Sigma_m^{-1}(\mathbf{x}-\mu_{\mathbf{m}})}}{(2\pi)^{N/2}|\Sigma_m|^{1/2}}. \end{aligned} \tag{11}$$

Here the CF of the normal distribution $N(\mu_{\mathbf{m}},\Sigma_m)$ can be simply represented as follows.

$$\varphi(\mathbf{w}) = e^{j\mathbf{w}^T\mu_{\mathbf{m}} - \frac{1}{2}\mathbf{w}^T\Sigma_m\mathbf{w}} \tag{12}$$

For the second model, we use the Dirichlet distribution as the prior of mixing vectors with parameters Θ:

$$f(\mathbf{s};\boldsymbol{\Theta}) = \frac{\Gamma\left(\sum_{m=1}^{M}\theta_m\right)}{\prod_{m=1}^{M}\Gamma(\theta_m)}\prod_{m=1}^{M} s_m^{\theta_m-1}. \tag{13}$$

Note that, this is the equation of $M-1$ dimensional Dirichlet distribution because the M-th fraction s_M can always be set to $s_M = 1 - \sum_{m=1}^{M-1} s_m$. In addition, for later use, let $\theta = \sum_{i=1}^{M}\theta_i$. If $M = 2$, one dimensional Dirichlet distribution is called Beta distribution [2].

In the standard mixed distribution, Equation (8) can be simplified because the normal distribution is a family of stable distributions. That is,

$$\varphi_{\mathbf{x}}(\mathbf{w}|\mathbf{s}) = \prod_{m=1}^{M} \varphi_{\mathbf{a}_m}(\mathbf{w}s_m)\varphi_{\varepsilon}(\mathbf{w}) \tag{14}$$

$$= e^{\left[j\mathbf{w}^T\left(\sum_{m=1}^{M} s_m \mu_{\mathbf{m}}\right) - \frac{1}{2}\mathbf{w}^T\left(\sum_{m=1}^{M} s_m^2 \Sigma_m + \Sigma_\varepsilon\right)\mathbf{w}\right]} \tag{15}$$

where the noise term is assumed to be gaussian $N(0, \Sigma_\varepsilon)$.

3.4 Statistics of Mixed Distribution

Now we are interested in characterizing the statistical properties of the standard mixed distribution, and we specifically focus on the moments and cumulants of the standard mixed distribution in this paper, although provided that the moments and cumulants do exist. For this purpose, the following power series expansion of the CF is useful:

$$\varphi(\mathbf{w}) = \sum_{l_1,\ldots,l_N=0}^{\infty} \frac{j^{\sum_{i=1}^{N} l_i}}{\prod_{i=1}^{N} l_i!} M_{\mathbf{l}} \prod_{i=1}^{N} w_i^{l_i}, \tag{16}$$

where $M_{\mathbf{l}} = M_{(l_1,\ldots,l_N)}$ is the moment of order $l = \sum_{i=1}^{N} l_i$. Conversely, the moment $M_{\mathbf{l}}$ can be calculated as follows:

$$M_{\mathbf{l}} = (-j)^l \frac{\partial^l \varphi(\mathbf{w})}{\partial w_1^{l_1} \cdots \partial w_N^{l_N}}\bigg|_{w_1=\cdots=w_N=0} \tag{17}$$

Following equations above, we derive the moments of the PDF $p(\mathbf{x}; \Psi, \Theta)$ with the power series expansion of $\varphi_{\mathbf{x}}(\mathbf{w})$. If we apply Equation (3) to both sides of Equation (10), and replace the CF with power series expansion as in Equation (16), we may obtain the following relation,

$$\sum_{l_1,\ldots,l_N=0}^{\infty} \frac{j^{\sum_{i=1}^{N} l_i}}{\prod_{i=1}^{N} l_i!} M_{\mathbf{l}} \prod_{i=1}^{N} w_i^{l_i}$$

$$= \int_S \sum_{l_1,\ldots,l_N=0}^{\infty} \frac{j^{\sum_{i=1}^{N} l_i}}{\prod_{i=1}^{N} l_i!} R_{\mathbf{l}}(\mathbf{s}; \Psi) \prod_{i=1}^{N} w_i^{l_i} f(\mathbf{s}; \Theta) d\mathbf{s}, \tag{18}$$

where $R_{\mathbf{l}}(\mathbf{s}; \Psi)$ is the moment of the PDF $\varphi^{-1}(\mathbf{x}; \Psi)$. Comparing the terms of same order in both sides in Equation (18), the simple formula is obtained for each $\mathbf{l}$,

$$M_{\mathbf{l}} = \int_S R_{\mathbf{l}}(\mathbf{s}; \Psi) f(\mathbf{s}; \Theta) d\mathbf{s}. \tag{19}$$

Note that these are moments about the origin, different from central moments about the mean.

Central moments for order 1 and 2 are already obtained for arbitrary N [5], but due to limited space, we focus on the case of $N = 1$ in this paper and demonstrate procedures for the first order moment $l = 1$. Firstly R_1 can be derived from Equation (15) and Equation (17),

$$R_1 = \frac{\partial \varphi_x(w|\mathbf{s})}{\partial w}\bigg|_{w=0} = \sum_{m=1}^{M} s_m \mu_m. \tag{20}$$

Then this result is used in Equation (19) to give

$$\begin{aligned} M_1 &= \int_S R_1 f(\mathbf{s}; \Theta) d\mathbf{s} \\ &= \sum_{m=1}^{M} \mu_m \int_S \frac{\Gamma(\theta)}{\prod_{m=1}^{M} \Gamma(\theta_m)} \left(\prod_{\substack{n=1 \\ n\neq m}}^{M} s_n^{\theta_n - 1} \right) s_m^{\theta_m} d\mathbf{s} \\ &= \frac{\sum_{m=1}^{M} \theta_m \mu_m}{\theta}, \end{aligned} \tag{21}$$

where the class-conditional PDF $p_m(\cdot) \sim N(s; \mu_i, \sigma_i^2)$, the noise term is omitted for clarity, and the following relationship is used without proof [5],

$$\int_S \frac{\Gamma(\theta)}{\prod_{m=1}^{M} \Gamma(\theta_m)} \left(\prod_{\substack{n=1 \\ n\neq m}}^{M} s_n^{\theta_n - 1} \right) s_m^{\theta_m - 1 + 1} d\mathbf{s} = \frac{\theta_m}{\theta}. \tag{22}$$

This procedure can easily extend to the second moment case or higher-order moment cases and Fig. 2 illustrates up to fourth order moments of the standard mixed distribution.

3.5 Experiments

$$\sum_{k=1}^{\infty} \frac{M_k}{(k-1)!} s^{k-1} = \left[\sum_{n=1}^{\infty} \frac{k_n}{(n-1)!} s^{n-1} \right] \left[\sum_{l=0} \infty \frac{M_l}{l!} s^l \right] \tag{29}$$

Gram-Charlier expansion

To verify equations in Fig. 2, we perform a simple simulation using random number generators. Comparison is made on cumulants, not on moments, so the moments are converted to cumulants using the following relationship,

$$k_1 = M_1 \tag{30}$$

$$k_2 = M_2 - M_1^2 \tag{31}$$

$$k_3 = M_3 - 3M_1M_2 + 2M_1^3 \tag{32}$$

$$k_4 = M_4 - 4M_3M_1 - 3M_2^2 + 12M_2M_1^2 - 6M_1^4 \tag{33}$$

We call these cumulants as theoretical cumulants k_i^t. On the other hand, we also define empirical cumulants k_i^e that can be calculated from simulated data as follows,

$$k_1 = E[X] = \mu \tag{34}$$

$$k_2 = E[(X-\mu)^2] = \sigma^2 \tag{35}$$

$$k_3 = E[(X-\mu)^3] \tag{36}$$

$$k_4 = E[(X-\mu)^4] - 3\sigma^4 \tag{37}$$

Here μ and σ^2 is the mean and the variance of a random variable X. The purpose of the experiments is to compare theoretical cumulants with empirical cumulants and evaluate the magnitude of error between them. The brief description of the experiments follows.

$$R_2 = \left\{\sum_{m=1}^{M} s_m\mu_m\right\}^2 + \left\{\sum_{m=1}^{M} s_m^2\sigma_m^2\right\} \tag{23}$$

$$R_3 = \left\{\sum_{m=1}^{M} s_m\mu_m\right\}^3 + 3\left\{\sum_{m=1}^{M} s_m\mu_m\right\}\left\{\sum_{m=1}^{M} s_m^2\sigma_m^2\right\} \tag{24}$$

$$R_4 = \left\{\sum_{m=1}^{M} s_m\mu_m\right\}^4 + 6\left\{\sum_{m=1}^{M} s_m\mu_m\right\}^2\left\{\sum_{m=1}^{M} s_m^2\sigma_m^2\right\} + 3\left\{\sum_{m=1}^{M} s_m^2\sigma_m^2\right\}^2 \tag{25}$$

$$M_2 = \frac{\sum_{m=1}^{M}\theta_m(\theta_m+1)\left(\mu_m^2+\sigma_m^2\right) + \sum_{\substack{m,n=1\\ m\neq n}}^{M}\theta_m\theta_n\mu_n\mu_m}{\theta(\theta+1)} \tag{26}$$

$$M_3 = \frac{\sum_{m=1}^{M}\theta_m(\theta_m+1)(\theta_m+2)\left(\mu_m^3+3\mu_m\sigma_m^2\right) + \sum_{\substack{m,n=1\\ m\neq n}}^{M}\theta_m\theta_n(\theta_n+1)\left(\mu_m\mu_n^2+3\mu_m\sigma_n^2\right) + \sum_{\substack{m,n,o=1\\ m\neq n\neq o}}^{M}\theta_m\theta_n\theta_o\mu_m\mu_n\mu_o}{\theta(\theta+1)(\theta+2)} \tag{27}$$

$$\begin{aligned} M_4 = {} & \frac{\sum_{m=1}^{M}\theta_m(\theta_m+1)(\theta_m+2)(\theta_m+3)\left(\mu_m^4+6\mu_m^2\sigma_m^2+3\sigma_m^4\right) + \sum_{\substack{m,n=1\\ m\neq n}}^{M}\theta_m(\theta_m+1)\theta_n(\theta_n+1)\left(\mu_m^2\mu_n^2+6\mu_m^2\sigma_n^2+3\sigma_m^2\sigma_n^2\right)}{\theta(\theta+1)(\theta+2)(\theta+3)} + \\ & \frac{\sum_{\substack{m,n=1\\ m\neq n}}^{M}\theta_m\theta_n(\theta_n+1)(\theta_n+2)\left(\mu_m\mu_n^3+6\mu_m\mu_n\sigma_n^2\right) + \sum_{\substack{m,n,o=1\\ m\neq n\neq o}}^{M}\theta_m\theta_n\theta_o(\theta_o+1)\left(\mu_m\mu_n\mu_o^2+6\mu_m\mu_n\sigma_o^2\right)}{\theta(\theta+1)(\theta+2)(\theta+3)} + \\ & \frac{\sum_{\substack{m,n,o,p=1\\ m\neq n\neq o\neq p}}^{M}\theta_m\theta_n\theta_o\theta_p\mu_m\mu_n\mu_o\mu_p}{\theta(\theta+1)(\theta+2)(\theta+3)} \end{aligned} \tag{28}$$

Figure 2: Moments of the standard mixed distribution for order 2, 3 and 4. Here $m \neq n \neq o \neq p$ represents that none of them are equivalent.

Table 1: Comparison between theoretical cumulants and empirical cumulants.

M	k_i	Average	Median
$M = 2$	k_1	1.57×10^{-3}	9.30×10^{-4}
	k_2	1.35×10^{-3}	1.15×10^{-3}
	k_3	2.40×10^{-2}	7.21×10^{-3}
	k_4	1.89×10^{-2}	1.07×10^{-2}
$M = 10$	k_1	1.14×10^{-3}	8.53×10^{-4}
	k_2	1.30×10^{-3}	1.17×10^{-3}
	k_3	2.61×10^{-2}	1.69×10^{-2}
	k_4	1.61×10^{-2}	1.13×10^{-2}

1. The parameters of component distributions Ψ and mixing distributions Θ are randomly generated.

2. The component matrix and the mixing vector are generated according to the normal distribution and the Dirichlet distribution respectively using parameters generated in 1, and a mixed random variable is computed according to Equation (1).

3. This trial (2) is repeated for 1,000,000 times.

4. Then empirical cumulants are calculated for these trials from equations Equation (34) to Equation (37).

5. At the same time, theoretical moments are calculated from equations Equation (30) to Equation (33) using randomly generated parameters in (1).

6. Evaluate the relative error $e_i = |k_i^t - k_i^e|/|k_i^t|$ except for the case $|k_i^t| \sim 0$.

7. Repeat a set of trials for 100 times with collecting e_i, and obtain the average and median of e_i for the cumulant of order i.

Then Table 1 shows the result. Our theoretical cumulants demonstrate relatively small error for all cases. Errors in higher-order cumulants tend to be larger, but we assume that its reason may be the accumulation of error or some bias in random number generators. Based on these results, we can now verify that our theoretical cumulants of the standard mixed distribution is correct.

3.6 Assumption on Mean Vectors

One interesting observation associated with moments derived so far is the effect of mean vectors. If we assume that all mean vectors are non-zero, then the number of terms involved in Equation (28) is $M^4 + M^3 + M^2$, which amounts to 11100 terms for $M = 10$. However, if we assume that all mean vectors are zero, then most of the terms are eliminated and moments can be represented in a simpler form. Moreover, M_1 and M_3 becomes zero and hence $k_1 = k_3 = 0$ for any σ_i^2. But even in this case $k_4 \neq 0$ and the mixed distribution is super-gaussian or sub-gaussian dependent on σ_i^2. This result indicates a unique hypothesis that even if the class-conditional PDFs of component classes are normal distributions, the distribution associated with linear mixing process is a super-gaussian or a sub-gaussian distributions due to the variability of mixing fractions over the signal.

Assumption on mean vectors is also an important factor in the estimation process, since the inverse problem of classifying those signals into either pure or mixed seems to be intractable if all component signals are generated from the PDFs with the same mean vector. Hence we assume that mean vectors are different from class to class. This assumption is theoretically not required, but in practice, it may be helpful to classify the observed signal into pure or mixed signals using cumulants. Furthermore, we should mention that this is a natural assumption in image analysis, since scalar values of image pixels represent unique spectral reflectance or emission from particular class.

4 Estimation

4.1 Density Estimation

Now the problem is to label each data vector $\mathbf{x}(t)$ as either "pure" or "mixed" and one method for solving this problem is to model the distribution of the observed data $p(\mathbf{x})$ as a finite mixture density, and regard the unknown label (pure / mixed) as incomplete information that has to be estimated. Then mixture density estimation based on EM algorithm is applied to minimize the log-likelihood $L(\Psi, \Theta)$ of the observed data,

$$L(\Psi, \Theta) = \sum_{t=1}^{T} \log p(\mathbf{x}(t); \Psi, \Theta), \tag{38}$$

where (Ψ, Θ) are parameters involved in the finite mixture density model. Note that we have two types of probabilistic models in a finite mixture density, namely pure distributions and mixed distributions. Some simple cases were already shown in [1, 2], but these results were based on the single modeling for the whole signal. Considering the non-stationarity of the signal, a better strategy would be to estimate cumulants for partial data and classify signals into pure or mixed signals.

4.2 Fraction Estimation

We then estimate the fraction of each component based on probability models, provided that we can now determine which component classes are involved at index t from the result of mixture density estimation. One convenient approach for estimate $\hat{\mathbf{s}}(t)$ is based on maximum likelihood,

$$\hat{\mathbf{s}}(t) = \max_{\mathbf{s} \in S} r(\mathbf{s}|\mathbf{x}(t)), \tag{39}$$

where $r(\mathbf{s}|\mathbf{x}(t))$ is the PDF of $\mathbf{s}$ conditioned on the realization of the observed vector $\mathbf{x}(t)$. Another more mathematical approach is based on the expectation,

$$\hat{\mathbf{s}}(t) = E\{\mathbf{s}|\mathbf{x}(t)\} = \int_S \mathbf{s} r(\mathbf{s}|\mathbf{x}(t)) d\mathbf{s}. \tag{40}$$

Comparing above two approaches, the former approach gives lower computational complexity. Although the latter approach gives more robust estimate of the mixing vector, in usual cases the difference of two approaches has only minor impact on the final result.

5 Conclusion

The main result in this paper is the characterization of our proposed probabilistic model called "mixed distribution." This distribution is characterized in terms of moments and cumulants, which is subsequently verified by simple simulation experiments, and it proposes a unique hypothesis that even if component distributions are gaussian, the distribution after linear mixing process is either super-gaussian or sub-gaussian because of the variability of mixing fractions.

References

[1] Kitamoto, A. and Takagi, M. Image Classification Using Probabilistic Models that Reflect the Internal Structure of Mixels. *Patt. Anal. Appl.*, Vol. 2, No. 2, pp. 31–43, 1999.

[2] Kitamoto, A. and Takagi, M. Area Proportion Distribution — Relationship with the Internal Structure of Mixels and its Application to Image Classification. *Syst. Comp. Japan*, Vol. 31, No. 5, pp. 57–76, 2000.

[3] Lee, D. D. and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, pp. 788–791, 1999.

[4] Lee, T.W. Independent Component Analysis: Theory and Applications. Kluwer Academic Publishers, 1998.

[5] Kitamoto, A. The Moments of the Mixel Distribution and Its Application to Statistical Image Classification. *Advances in Pattern Recognition*, LNCS 1876, pp. 521–531. 2000.