A Generalised Directional Laplacian Distribution for Underdetermined Audio Source Separation

By Mitianoudis N.

Electrical and Computer Engineering Dep., Democritus Univ. of Thrace, Greece

Abstract

The proposed Generalised Directional Laplacian Distribution (DLD) is a hybrid between the Laplacian distribution and the von Mises-Fisher distribution aiming at modelling multidimensional sparse directional or angular data. The essential algorithms to estimate the proposed distribution's parameters as well as Mixtures of DLD (MDLD) are presented. The proposed DLD mixture model is used to cluster sound sources that exist in an underdetermined instantaneous sound mixture, offering a fast and viable solution to the general $K \times L$ ($K < L$) problem.

1. Introduction

Circular Statistics is the branch of statistics that addresses the modeling of circular data, i.e. data with rotating values. The *von Mises distribution* is a continuous probability distribution on the unit circle (the circular equivalent of the normal distribution) and is defined by (see Jammalamadaka & Sengupta (2001)):

$$p(\theta) = \frac{e^{k\cos(\theta - m)}}{2\pi I_0(k)} \quad ,\forall\ \theta \in [0, 2\pi) \tag{1.1}$$

where $I_0(k)$ is the modified Bessel function of the first kind of order 0, m and $k > 0$ describe the mean and the "width" of the distribution. Extending to the multivariate case, a p-dimensional unit random vector $\mathbf{x}$ ($||\mathbf{x}|| = 1$) follows a *von Mises-Fisher* distribution, if

$$p(\mathbf{x}) \propto e^{k\mathbf{m}^T\mathbf{x}} \quad ,\forall ||\mathbf{x}|| \in \mathcal{S}^{p-1} \tag{1.2}$$

where $||\mathbf{m}|| = 1$ defines the centre, $k \geq 0$ and $\mathcal{S}^{p-1}$ is the p dimensional unit hypersphere. In the case of $p = 2$, (1.2) is reduced to the von-Mises distribution of (1.1). One can encounter many methods to fit the von Mises-Fisher distribution or its mixtures to normally distributed circular data (see Jammalamadaka & Sengupta (2001), Mardia et al. (1999), Dhillon & Sra (2003)).

This study proposes a novel distribution to model directional sparse data, i.e. data that are mostly close to their mean value with the exception of several outliers. The Laplacian distribution $p(x) \propto e^{k|x-m|}$ appears to be a strong candidate in modelling sparse data (see Davies (2002)). There were several attempts to model circular sparse signals by wrapping an 1-D or multidimensional Laplace distributions of infinite support (see Jammalamadaka & Kozubowski (2002), Mitianoudis & Stathaki (2007)); however, it is reported to have increased computational cost, as the periodic repetition of a density function is equivalent to a mixture of density functions.

One application where directional statistical modelling is essential is *Underdetermined*

Audio Blind Source Separation (BSS) (see Comon & Jutten (2010)). Assume that a set of K sensors $\mathbf{x}(n) = [x_1(n), \ldots, x_K(n)]^T$ observes a set of L ($K < L$) sound sources $\mathbf{s}(n) = [s_1(n), \ldots, s_L(n)]^T$. The instantaneous (anechoic) mixing model can be expressed by $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n)$, where $\mathbf{A}$ represents a $K \times L$ *mixing matrix* and n the sample index. BSS algorithms provide an estimate of the source signals $\mathbf{s}$ and the mixing matrix $\mathbf{A}$, based on the observed signals $\mathbf{x}$ and some statistical source profile. The underdetermined instantaneous case is challenging. In the two-channel ($K = 2$) BSS scenario, the problem is reduced to an angular clustering problem of sparse data. Many methods have been proposed (see Comon & Jutten (2010)). Mitianoudis & Stathaki (2007) introduced a Laplacian Mixture Model to perform separation. To tackle the angular wrapping at π, they examined the use of Wrapped Laplacian Mixtures (MoWL). Nonetheless, these efforts do not offer a closed form solution and can not be expanded to more than two sensors. Recently, Vincent et al. (2009) used local Gaussian Modelling of minimal constrained variance of the local time-frequency neighbours to perform separation. This study applies the proposed DLD model to address the general $K \times L$ source separation problem, which is rarely tackled in the literature.

2. A Generalised Directional Laplacian model

Assume a r.v. θ modelling directional data with π-periodicity. Mitianoudis (2010) introduced the 1D-DL Density, as follows:

$$p(\theta) = c(k)e^{-k|\sin(\theta - m)|} \quad ,\forall\ \theta \in [0, \pi) \tag{2.1}$$

where $m \in [0, \pi)$ defines the mean, $k > 0$ defines the width ("approximate variance") of the distribution, $c(k) = \frac{1}{\pi I_0(k)}$ and $I_0(k) = \frac{1}{\pi}\int_0^{\pi} e^{-k\sin\theta} d\theta$.

The next step is to derive a generalised definition for the Directional Laplacian model. To generalise the concept of 1D DLD in the p-dimensional space, we will be inspired by the p-D von Mises-Fisher distribution. Since $||\mathbf{x}|| = ||\mathbf{m}|| = 1$, the inner product $\mathbf{m}^T\mathbf{x} = \cos\psi$, where ψ is the angle between the two vectors. Following a similar methodology to the 1D-DLD, we need to formulate the term $-k|\sin\psi|$. But $|\sin\psi| = \sqrt{1 - (\mathbf{m}^T\mathbf{x})^2}$. Thus, the following probability density function models p-D directional Laplacian data and is termed *Generalised Directional Laplacian Distribution* (DLD):

$$p(\mathbf{x}) = c_p(k)e^{-k\sqrt{1-(\mathbf{m}^T\mathbf{x})^2}} \quad ,\forall\ ||\mathbf{x}|| \in \mathcal{S}^{p-1} \tag{2.2}$$

where $\mathbf{m}$ defines the mean, $k \geq 0$ defines the width ("approximate variance") of the distribution, $c_p(k) = \frac{\Gamma((p-1)/2)}{\pi^{(p+1)/2} I_{p-2}(k)}$, $I_p(k) = \frac{1}{\pi}\int_0^{\pi} e^{-k\sin\theta}\sin^p\theta d\theta$ and $\Gamma(\cdot)$ represents the Gamma function.

One can employ *Mixtures of Generalised Directional Laplacians* (MDLD) in order to model multiple concentrations of directional generalised "heavy-tailed signals", as follows:

$$p(\mathbf{x}) = \sum_{i=1}^{K} a_i c_p(k_i) e^{-k_i\sqrt{1-(\mathbf{m}_i^T\mathbf{x})^2}} \quad ,\forall\ ||\mathbf{x}|| \in \mathcal{S}^{p-1} \tag{2.3}$$

where a_i denotes the weight of each distribution in the mixture, K the number of DLDs used in the mixture and $\mathbf{m}_i$, k_i denote the mean and the "width" (approximate variance) of each distribution.

The mixtures of DLD can be trained using the Expectation-Maximisation (EM) algorithm. Following the previous analysis in Mitianoudis & Stathaki (2007), one can yield the following simplified likelihood function:

$$\mathcal{L}(a_i, \mathbf{m}_i, k_i) = \sum_{n=1}^{N} \sum_{i=1}^{K} \left(\log \frac{a_i \Gamma(\frac{p-1}{2})}{\pi^{\frac{p+1}{2}} I_{p-2}(k)} - k\sqrt{1 - (\mathbf{m}^T \mathbf{x}_n)^2} \right) p(i|\mathbf{x}_n) \tag{2.4}$$

where $p(i|\mathbf{x}_n)$ represents the probability of sample $\mathbf{x}_n$ belonging to the i^{th} DLD of the mixture. In a similar fashion to other mixture model estimation, the updates for $p(i|\mathbf{x}_n)$ and α_i can be given by the following equations:

$$p(i|\mathbf{x}_n) \leftarrow \frac{a_i c_p(k_i) e^{-k_i\sqrt{1-(\mathbf{m}_i^T \mathbf{x}_n)^2}}}{\sum_{i=1}^{K} a_i c_p(k_i) e^{-k_i\sqrt{1-(\mathbf{m}_i^T \mathbf{x}_n)^2}}}, \quad a_i \leftarrow \frac{1}{N} \sum_{n=1}^{N} p(i|\mathbf{x}_n) \tag{2.5}$$

Taking the derivatives along $\mathbf{m}_i$ and k_i, it is straightforward to derive the following gradient iterative updates:

$$\mathbf{m}_i^+ \leftarrow \mathbf{m}_i + \eta \sum_{n=1}^{N} \frac{\mathbf{m}^T \mathbf{x}_n}{\sqrt{1 - (\mathbf{m}^T \mathbf{x}_n)^2}} \mathbf{x}_n p(i|\mathbf{x}_n), \quad \mathbf{m}_i^+ \leftarrow \mathbf{m}_i^+ / ||\mathbf{m}_i^+|| \tag{2.6}$$

Setting $\partial\mathcal{L}/\partial k_i = 0$ yields the following ratio and k_i can then be estimated numerically from a lookup table.

$$\frac{I_{p-1}(k_i)}{I_{p-2}(k_i)} = \frac{\sum_{n=1}^{N} \sqrt{1 - (\mathbf{m}_i^T \mathbf{x}_n)^2} p(i|\mathbf{x}_n)}{\sum_{n=1}^{N} p(i|\mathbf{x}_n)} \tag{2.7}$$

The training of this mixture model is also dependent on the initialisation of its parameters, especially the means $\mathbf{m}_i$, which are initiallised by a *Directional K-Means* algorithm (see Mardia et al. (1999)).

3. Source Separation using MDLD

The generalised Directional Laplacian Density offers a faster and complete solution to the problem, as it can be automatically applied to the general $K \times L$ audio source separation scenario. Once the MDLD are fitted to the multichannel directional data, separation can be performed by "hard-thresholding" for the 1D-case (intersections of individual DLDs), or "soft-thresholding" for the general p-D case in a similar manner to Mitianoudis & Stathaki (2007). Hence, the i^{th} source can be associated with those points x_n, for which $p(x_n) \geq (1 - q)\alpha_i c_p(k_i)$, where $q \in [0.7, 0.9]$. Having attributed the points $\mathbf{x}(n)$ to the L sources, using either the "hard" or "soft" thresholding technique, the next step is to reconstruct the sources. Let $S_i \sqsubseteq N$ represent the point indices that have been attributed to the i^{th} source and $\mathbf{m}_i$ the corresponding mean vector, i.e. the corresponding column of the mixing matrix. We initialise $u_i(n) = 0, \forall\ n = 1, \ldots, N$ and $i = 1, \ldots, L$. The source reconstruction is performed by substituting $u_i(S_i) = \mathbf{m}_i^T \mathbf{x}(S_i), \forall\ i = 1, \ldots, L$

Next, we evaluate the proposed MDLD algorithm for audio source separation and compare with the "GaussSep" algorithm by Vincent et al. (2009) (MATLAB code from `http://www.irisa.fr/metiss/members/evincent/software`). In order to quantify the performance of the algorithms, we estimate the *Signal-to-Distortion Ratio* (SDR), the *Signal-to-Interference Ratio* (SIR) and the *Signal-to-Artifact Ratio* (SAR) from the BSS_EVAL Toolbox v.3 (see `http://bass-db.gforge.inria.fr/bss_eval.`). The input signals for the MDLD were sparsified using the *Modified Discrete Cosine Transformation* (MDCT). We tested the algorithms with test signals from the Signal Separation Evaluation Campaigns (see `http://sisec.wiki.irisa.fr/`). In this section, we will attempt to perform separation of the Dev3Female3 set from SiSEC2011 and a 3×5 (3 mixtures - 5 sources)

	SDR (dB)		**SIR (dB)**		**SAR (dB)**	
	MDLD	GaussSep	MDLD	GaussSep	MDLD	GaussSep
Dev3Female3	6.02	16.93	23.84	22.43	6.17	18.40
Example 3×5	3.91	9.94	17.92	15.21	4.17	11.68
Example 4×8	2.24	-18.63	16.4	-17.58	2.52	9.39

TABLE 1. The MDLD approach is compared with the GaussSep approach ($K = 3, 4$). Scores are averaged for all sources of each experiment.

and a 4×8 (4 mixtures - 8 sources) scenario using the male and female voices from Dev3. After fitting the MDLD model, we employed the soft-thresholding scheme, using a value of $q = 0.8$. The separation results for these experiments are summarised in Table 1. In the case of $K = 3$ mixtures, both algorithms managed to perform separation. The "GaussSep" algorithm featured higher SDR and SAR performances, whereas the proposed MDLD featured higher SIR performance. The image is completely different in the case of $K = 4$ mixtures, where the MDLD manages to separate all 8 sources in contrast to the "GaussSep" that fails to perform separation. This might be due to fact that the sparsest ML solution in the optimisation of "GaussSep" is restricted to vectors with $K \leq 3$ entries, i.e. 3 sources present at each point. In contrast, the proposed MDLD algorithm can operate for any arbitrary number of sensors K. In addition, the MDLD required an average of 7.66 secs to perform separation, which is similar to the $K = 2$ cases. In contrast, the "GaussSep" algorithm's processing has increased considerably with K. For $K = 3$, it required an average of 1310 sec and for $K = 4$, it required 2359 sec which is almost the double processing time for $K = 3$.

4. Conclusions

The problem of modelling multidimensional Directional sparse data is addressed by a novel generalised Directional Laplacian model. The proposed algorithm can also offer a solution for the general multichannel underdetermined source separation problem ($K \geq 2$), offering fast and stable performance and high SIR compared to state-of-the-art methods.

REFERENCES

COMON P. & JUTTEN C. 2010 Handbook of Blind Source Separation: Independent Component Analysis and Applications *Academic Press*, 856 pages.

DAVIES M. 2002 Audio source separation *Mathematics in Signal Processing* **V**, 57-68.

DHILLON I.S. & SRA S. 2003 Modeling Data using Directional Distributions *Tech. Rep., Technical Report TR-03-06, University of Texas at Austin, Austin, TX.*

JAMMALAMADAKA S.R. & KOZUBOWSKI T.J. 2002 A new family of Circular Models: The Wrapped Laplace distributions *Tech. Rep. No. 61, Dep. of Math., Univ. of Nevada.*

JAMMALAMADAKA S.R. & SENGUPTA A. 2001 Topics in Circular Statistics *World Scientific.*

MARDIA K.V., KANTI V. & JUPP P.E. 1999 Directional Statistics *Wiley-Blackwell.*

MITIANOUDIS N. 2010 A Directional Laplacian Density for Underdetermined Audio Source Separation, *Lect. Notes in Comp. Science, Int. Conf. on Artif. Neural Networks* **6352**, 450-459.

MITIANOUDIS N. & STATHAKI T. 2007 Underdetermined Source Separation using Mixtures of Warped Laplacians *Lect. Notes in Comp. Science, Indep. Comp. Analysis and Source Separation* **4666**, 236-243.

VINCENT E. & ARBERET S. & GRIBONVAL R. 2009 Underdetermined instantaneous audio source separation via local Gaussian modeling *Lect. Notes in Comp. Science, Indep. Comp. Analysis and Source Separation* **5441**, 775–782.