A Directional Laplacian Density for Underdetermined Audio Source Separation

Nikolaos Mitianoudis

School of Science and Technology, International Hellenic University, 14th km Thessaloniki-Moudania, 57001 Thessaloniki, Greece n.mitianoudis@ihu.edu.gr http://www.tech.ihu.edu.gr/

Abstract. In this work, a novel probability distribution is proposed to model sparse directional data. The Directional Laplacian Distribution (DLD) is a hybrid between the linear Laplacian distribution and the von Mises distribution, proposed to model sparse directional data. The distribution's parameters are estimated using Maximum-Likelihood Estimation over a set of training data points. Mixtures of Directional Laplacian Distributions (MDLD) are also introduced in order to model multiple concentrations of sparse directional data. The author explores the application of the derived DLD mixtures to cluster sound sources that exist in an underdetermined two-sensor mixture.

Key words: Audio Source Separation, Mixture Models, Directional Data, Sparse Data Modelling

1 Introduction

Circular Statistics is the branch of statistics that addresses the modeling and inference from *circular* or *directional* data, i.e. data with rotating values. Many interesting circular models can be generated from known probability distributions by either wrapping a linear distribution around the unit circle or transforming a bivariate linear r.v. to its directional component [1]. There also exist several distributions that are periodic by definition and can therefore be employed to model directional data. The *von Mises distribution* (also known as the circular normal distribution) is a continuous probability distribution on the unit circle [1]. It may be considered the circular equivalent of the normal distribution and is defined by:

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

$$\tag{1}$$

where $I_0(k)$ is the modified Bessel function of order 0, m is the mean and k > 0 describes the "width" of the distribution. The von Mises distribution has been extensively studied and many methods that fit the distribution or its mixtures to normally distributed circular data have been proposed [2, 1, 3, 4].

This study proposes a novel distribution to model directional sparse data. Sparsity is mainly used to describe data that are mostly close to their mean

value with the exception of several large values. There are several sparse models that have been proposed for linear sparse data [5] and several attempts to model circular sparse signals by wrapping a Laplace distribution [6,7]. This paper proposes a Directional Laplacian Distribution (DLD) as a direct modelling solution for circular sparse data. The Maximum Likelihood estimates (MLE) of the model's parameters are derived, along with an Expectation-Maximisation (EM) algorithm that estimates the parameters of a Mixture of Directional Laplacian Distributions (MDLD).

One application where directional statistical modelling is essential is Underdetermined Audio Blind Source Separation (BSS) [5,7,8]. 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 mixing model can be expressed in mathematical terms, by $\mathbf{x}(n) = \mathbf{As}(n)$, where \mathbf{A} represents the mixing matrix and n the sample index. The blind source separation problem provides an estimate of the source signals \mathbf{s} , based on the observed microphone signals and some general source statistical profile. The two-channel (K = 2) BSS scenario is often reduced to an angular clustering problem of sparse data, which has been addressed using Wrapped Laplacian Mixtures [7]. The proposed DLD model is tested with several synthetic modelling experiments and in real audio BSS examples.

2 A Directional Laplacian Density

2.1 Definition

Assume a r.v. θ modelling directional data with π -periodicity.

Definition 1. The following probability density function models directional Laplacian data over $[0, \pi)$ and is termed Directional Laplacian Density (*DLD*):

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

$$(2)$$

where $m \in [0,\pi)$ defines the mean, k > 0 defines the "width" (approximate variance) of the distribution, $c(k) = \frac{1}{\pi \hat{I}_0(k)}$ and $\hat{I}_0(k) = \frac{1}{\pi} \int_0^{\pi} e^{-k \sin \theta} d\theta$.

The normalisation coefficient $c(k) = 1/\pi \hat{I}_0(k)$ can be easily derived from the fundamental property of density functions. In Figure 1, the DLD is depicted for m = 0.1 and various values of k. The DLD is a heavy-tailed density that exhibits a π periodicity. The above definition can be amended to reflect a "fully circular" phenomenon (2π) , however, we will continue with the π periodicity since it is required by our source separation application.

2.2 Maximum Likelihood Estimation using Directional Laplacian Priors

Assume a population of angular data $\Theta = \{\theta_1, \dots, \theta_n, \dots, \theta_N\}$ that follow a Directional Laplacian Distribution. Maximum Likelihood Estimation (MLE) can



Fig. 1. The proposed Directional Laplacian Density (DLD) for m = 0.1 and various values of k.

be used to fit the proposed Directional Laplacian Distribution to a set of angular data Θ . Assuming statistical independence between the observed data points, the MLE problem [9] can solve by maximising the log-likelihood function $J(\Theta, m, k)$ in terms of m, k. Assuming a DLD prior, the log-likelihood function $J(\Theta, m, k)$ can be expressed, as follows:

$$J(\Theta, m, k) = -N \log \pi - N \log \hat{I}_0(k) - k \sum_{n=1}^N |\sin(\theta_n - m)|$$
(3)

To estimate the parameters that maximise the log-likelihood, alternating updates for m and k iteratively yields an optimum of the log-likelihood. It is not straighforward to obtain a closed-form solution by solving the equations $\partial J(\Theta, m, k)/\partial m = 0$ and $\partial J(\Theta, m, k)/\partial k = 0$ for m and k respectively. Thus, one has to resort to alternative solutions to estimate m and k.

Estimation of m Iterative optimisation is employed to estimate m. To achieve faster convergence, a Newton-step optimisation will be pursued. The Newton-step updates for m can be given by the following update equation [9]:

$$m^{+} \leftarrow m + \left[\frac{\partial^2 J(\Theta, m, k)}{\partial m^2}\right]^{-1} \frac{\partial J(\Theta, m, k)}{\partial m}$$
 (4)

where m^+ denotes the new update of the estimated parameter and

$$\frac{\partial J(\Theta, m, k)}{\partial m} = \sum_{n=1}^{N} k \operatorname{sgn}(\theta_n - m) \cos(\theta_n - m)$$
(5)

$$\frac{\partial^2 J(\Theta, m, k)}{\partial m^2} = \sum_{n=1}^N k(\delta(\theta_n - m) + |\sin(\theta_n - m)|) \tag{6}$$

The Newton-step for m is independent from k and can be estimated, as follows:

$$m^{+} \leftarrow m + \frac{\sum_{n=1}^{N} \operatorname{sgn}(\theta_{n} - m) \cos(\theta_{n} - m)}{\sum_{n=1}^{N} \delta(\theta_{n} - m) + |\sin(\theta_{n} - m)|}$$
(7)

Estimation of k **In order to avoid the iterative update of** k **via gradient ascent** on the log-likelihood [9], an alternative solution is to solve $\partial J(\Theta, m, k)/\partial k = 0$ numerically. From the first derivative of $J(\cdot)$ along k, and

$$\frac{\partial^p}{\partial k^p}\hat{I}_0(k) = (-1)^p \frac{1}{\pi} \int_0^\pi \sin^p \theta e^{-k\sin\theta} d\theta = (-1)^p \hat{I}_p(k) \tag{8}$$

it is straightforward to derive the following:

$$\frac{\hat{I}_1(k)}{\hat{I}_0(k)} = \frac{1}{N} \sum_{n=1}^{N} |\sin(\theta_n - m)|$$
(9)

Calculating k from the ratio $\hat{I}_1(k)/\hat{I}_0(k)$ analytically is not very straightforward. However, through numerical evaluation, it can be shown that the ratio $\hat{I}_1(k)/\hat{I}_0(k)$ is a monotonic 1-1 function of k. In Figure 2, the ratio $\hat{I}_1(k)/\hat{I}_0(k)$ is estimated for uniformly sampled values of $k \in [0.01, 30]$. Since this ratio is not dependent on data, one can create a look-up table for a variety of k values and use interpolation to estimate k from an arbitrary value of $\hat{I}_1(k)/\hat{I}_0(k)$. This look-up table solution is more efficient compared to iterative optimisation approaches and generally accelerates the model's training.

Initialisation of the ML approach In [2], there exists a methodology to estimate closed-form solutions for m, k in the case of the Von-Mises distribution. The methodology can not yield direct solutions of m, k for the DLD prior, however, one can employ the strategy in [2] to derive an upper bound of the log-likelihood in (3). The derived upper bound can be used to extract closed-form estimates of m_{init} that can be used as initialisations of the iterative solutions proposed in the previous section. Let $C = \sum_n |\cos \theta_n|, S = \sum_n |\sin \theta_n|, R = \sqrt{C^2 + S^2}, \bar{\theta} = \operatorname{atan}(S/C), |\cos \bar{\theta}| = C/R, |\sin \bar{\theta}| = S/R$. Using the triangular inequality $|x| - |y| \leq |x \pm y| \leq |x| + |y|$, we manipulate the last term of the log-likelihood in (3).

$$\sum_{n} |\sin(\theta_{n} - \bar{\theta} + \bar{\theta} - m)| \geq \sum_{n} |\sin(\theta_{n} - \bar{\theta})| |\cos(\bar{\theta} - m)| - \sum_{n} |\cos(\theta_{n} - \bar{\theta})| |\sin(\bar{\theta} - m)|$$

$$= \sum_{n} |\sin\theta_{n}\cos\bar{\theta} - \cos\theta_{n}\sin\bar{\theta}| |\cos(\bar{\theta} - m)| - \sum_{n} |\cos\theta_{n}\cos\bar{\theta} + \sin\theta_{n}\sin\bar{\theta}| |\sin(\bar{\theta} - m)|$$

$$\geq (S|\cos\bar{\theta}| - C|\sin\bar{\theta}|) |\cos(\bar{\theta} - m)| - \sum_{n} (|\cos\theta_{n}||\cos\bar{\theta}| + |\sin\theta_{n}||\sin\bar{\theta}|) |\sin(\bar{\theta} - m)|$$

$$\geq (S\frac{C}{R} - C\frac{S}{R}) - \frac{C^{2} + S^{2}}{R} |\sin(\bar{\theta} - m)| = -R|\sin(\bar{\theta} - m)|$$



Fig. 2. The ratio $\hat{I}_p(k)/\hat{I}_{p-1}(k)$ is a monotonic 1-1 function of k.

Thus, the log-likelihood appears to have the following upper bound:

$$J(m,k) \le kR |\sin(\theta - m)| - N\log\pi - N\log I_0(k) = J_u(m,k)$$
(10)

To find the maximum of the bound, we set

$$\frac{\partial J_u}{\partial m} = -\operatorname{sgn}(\sin(\bar{\theta} - m))\cos(\bar{\theta} - m) = 0 \tag{11}$$

The above equation yields the solutions $m = \bar{\theta}$ or $m = \bar{\theta} \pm \pi/2$. For the second solution set, only the $m = \bar{\theta} + \pi/2$ solution can be valid, since $m \in [0, \pi)$ and $\bar{\theta} \in [0, \pi/2)$. The two solutions arose to account for the use of absolute values in the estimation of $\bar{\theta}$ and thus $\bar{\theta}$ is restricted to $\bar{\theta} \in [0, \pi/2)$. A simple method to define the correct solution is to evaluate the expression $\operatorname{sgn}(\sum_{n=1}^{N} \operatorname{atan} \tan \theta_n)$. If the expression is positive, the correct solution is $m_{init} = \bar{\theta}$. In the opposite case, the correct solution is $m_{init} = \bar{\theta} + \pi/2$. In Figure 3, a comparison between the estimated optimum m_{init} and the actual m for all values of $m \in [0, \pi)$ is performed for two values of k. The accuracy of the bound seems to depend on the value of m. In addition, the tightness of the bound approximates the actual m. For low values of k, the accuracy of the estimated m_{init} depends on the actual m. In general, the optimal value m_{init} that was estimated in this section can only serve as a valid initialisation of the ML approach.

2.3 Mixtures of Directional Laplacian Distributions

In a similar fashion to Gaussian Mixture Models (GMM), one can employ *Mixtures of Directional Laplacian Distributions* (MDLD) in order to model a mixture of directional "heavy-tailed signals".

5



Fig. 3. The estimated optimum m_{init} (red line) compared to the actual m (blue dotted line) for all values of $m \in [0, \pi)$ for k = 2 (left) and k = 12 (right). The tightness of the bound depends on the value of k.

Definition 2. Mixtures of Directional Laplacian Distributions are defined by the following pdf:

$$p(\theta, m_i, k_i) = \sum_{i=1}^{M} a_i c(k_i) e^{-k_i |\sin(\theta - m_i)|} \quad , \forall \ \theta \in [0, \pi)$$

$$(12)$$

where a_i denotes the weight of each distribution in the mixture, m_i , k_i denote the mean and the approximate variance of each distribution and M the number of DLDs used in the mixture.

Training via Expectation-Maximisation (EM) A common method that can be employed to train a mixture model is the *Expectation-Maximization* (EM) algorithm. Bilmes estimates Maximum Likelihood mixture density parameters using the EM [10]. Assuming N training samples for θ_n and Directional Laplacian mixtures (12), the log-likelihood of these training samples takes the following form:

$$J(a_i, m_i, k_i) = \sum_{n=1}^{N} \log \sum_{i=1}^{M} \frac{a_i}{\pi \hat{I}_0(k_i)} e^{-k_i |\sin(\theta_n - m_i)|}$$
(13)

Introducing unobserved data items that can identify the components that "generated" each data item, we can simplify the log-likelihood of (13) for MDLD, as follows:

$$I(a_i, m_i, k_i) = \sum_{n=1}^{N} \sum_{i=1}^{M} (\log \frac{a_i}{\pi} - \log \hat{I}_0(k_i) - k_i |\sin(\theta_n - m_i)|) p(i|\theta_n)$$
(14)

where $p(i|\theta_n)$ represents the probability of sample θ_n belonging to the i^{th} Directional Laplacian of the MDLD. In a similar fashion to Gaussian Mixture Models,

the updates for $p(i|\theta_n)$ and α_i can be given by the following equations:

$$p(i|\theta_n) \leftarrow \frac{a_i c(k_i) e^{-k_i |\sin(\theta_n - m_i)|}}{\sum_{i=1}^M a_i c(k_i) e^{-k_i |\sin(\theta_n - m_i)|}}$$
(15)

$$a_i \leftarrow \frac{1}{N} \sum_{n=1}^{N} p(i|\theta_n) \tag{16}$$

Based on the derivation of the MLE estimates for DLD prior, it is straightforward to derive the following updates for the MDLD parameters m_i and k_i , by optimising $I(a_i, m_i, k_i)$. The means m_i are estimated using a Newton-step on $I(a_i, m_i, k_i)$, as follows:

$$m_i^+ \leftarrow m_i + \frac{\sum_{n=1}^N \operatorname{sgn}(\theta_n - m_i) \cos(\theta_n - m_i) p(i|\theta_n)}{\sum_{n=1}^N (\delta(\theta_n - m_i) + |\sin(\theta_n - m_i)|) p(i|\theta_n)}$$
(17)

To estimate k_i , we will resort to the numerical estimation of k_i . The first derivative is given by:

$$\frac{\partial I(a_i, m_i, k_i)}{\partial k_i} = \sum_{n=1}^{N} (\frac{\hat{I}_1(k_i)}{\hat{I}_0(k_i)} - |\sin(\theta_n - m_i)|) p(i|\theta_n)$$
(18)

Equating (18) to zero yields:

$$\frac{\hat{I}_1(k_i)}{\hat{I}_0(k_i)} = \frac{\sum_{n=1}^N |\sin(\theta_n - m_i)| p(i|\theta_n)}{\sum_{n=1}^N p(i|\theta_n)}$$
(19)

Using the lookup table solution that was discussed in the previous section, one can estimate k_i directly from (19).

Training Initialisation The training of a mixture model is dependent on the initialisation of its parameters, especially the means m_i . Depending on different initialisations, in general, the EM algorithm may yield different mixture models that approximate the observed pdf, implying that the EM algorithm may get trapped in several local maxima. In a similar fashion to Gaussian Mixture Model initialisation, a *Directional K-Means* [11] is used to initialise the means m_i of the DLDs in the MDLD EM training, described in the previous section.

3 Experiments

In this section, we evaluate the efficiency of the derived EM algorithms for the estimation of the Directional Laplacian Density Mixtures. To generate 1D Directional Laplacian data, we employed the Inverse Transform Sampling technique [12]. We created a mixture of 5 concentrations of DLD samples centred at 0.3, 0.9, 1.5, 2.5, 3 respectively. The values of k_i for each DLD were 12, 10, 12, 14, 14.

7

Each of the 5 concentrations contained different number of samples in order to create different contributions to the overall density. The total number of samples was set to 3000. We ran 50 independent runs of the EM-algorithm as described in Section 2.3. An average of the estimated \hat{m}_i , \hat{k}_i and \hat{a}_i is shown in Table 1. In most of the 50 cases, we witnessed accurate estimation of the underlying concentrations. However, there were several cases (5-6 out of 50), where the initialisation offered by the Directional K-Means was not accurate for some of the clusters. Since most mixture estimation algorithms are sensible to initialisation, this resulted into a drop of the average accuracy in mixture training. Nonetheless, most of the clusters were correctly identified and the overall results are promising. Also it is important to mention that two of the cluster centres were chosen to be at 0.3 and 3 which are very close to the borders π and 0 that would be causing problems in the linear case. The MDLD is not affected at all, since it it is circular by definition.

The next step is to evaluate the proposed algorithm for audio source separation. Moving the sensor signals to a sparser domain (MDCT domain), the sources become sparser and thus smaller coefficients are more probable and most of the signal's energy is concentrated in few large values. Therefore, the density of the data in the mixture space shows a tendency to cluster along the directions of the mixing matrix columns and the two-sensor multiple-source separation problem is reduced to an angular clustering problem of $\theta_n = \operatorname{atan}(x_2(n)/x_1(n))$ [7, 8]. In these experiment, we will use Hyvärinen's clustering approach [13], the MoL-EM [7] and the WMoL-EM [8] for comparison. After Mixture fitting with the EM algorithm, separation will be performed using hard thresholding, as described in our previous work [7, 8]. In order to quantify the performance of the algorithms, we are estimating the Signal-to-Distortion Ratio (SDR) from the BSS_EVAL Toolbox [14]. The frame length for the MDCT analysis is set to 64 msec for the test signals sampled at 16 KHz and to 46.4 msec for those at 44.1 KHz. We initialise the parameters of the MoL, MoWL and MDLD as follows: $\alpha_i = 1/N$ and $c_i = 0.001$, T = [-1, 0, 1] (for MoWL only) and $k_i = 15$ (for Circular Laplacian only). The centres m_i were initialised in either case using the Directional *K*-means step.

We tested the algorithms with the *Groove* and the *Latino* dataset, available by (BASS-dB) [15], sampled at 44.1 KHz. The "Groove" dataset features four widely spaced sources: bass (far left), distorted guitar (center left), clean guitar (center right) and drums (far right). The "Latino" dataset features four widely spaced sources: bass (far left), drums (center left), keyboards (center right) and distorted guitar (far right). In Table 2, we can see the results for the four methods in terms of SDR. The average SDR of the two input signals, treating them as estimates of each input signal, is also provided for comparison (Mixed Signals row). The proposed MDLD approach seems to provide the best performance for most of the audio sources with small difference though. The important advantage of the proposed MDLD approach is that it is not susceptible to bordering effects. It is slightly slower compared to the original MoL approach and also faster compared

A Directional Laplacian Density for Underdetermined Source Separation

to the previous MoWL, since the proposed EM is more efficient compared to the one for MoWL.

4 Conclusions

In this paper, the author addresses the problem of modelling Directional Sparse data. Sparsity is often modelled using the Laplacian density for data with infinite linear support, which is not directly applicable in the case of directional or circular data. This work is building on previous work on directional Gaussian models (i.e. the von-Mises densities) to propose a novel Directional Laplacian model for modelling directional sparse data. ML estimates and an EM-algorithm were introduced to handle sparse circular data modelling problems. The proposed circular density was applied to the underdetermined source separation problem. The proposed solution featured a complete solution to the problem, compared to previous efforts. For future work, the author is planning to expand the model to handle multi-dimensional circular data.

Table 1. Parameter Estimation for a Mixture of Directional Laplacian (M = 5) using the proposed EM algorithm. Average parameter results for 50 independent runs.

	m_i	k_i	a_i	$\hat{m_i}$	$ \hat{k_i} - k_i /k_i$	$\hat{a_i}$
DLD_1	0.3	12	0.1071	0.4743	0.0616	0.0906
DLD_2	0.9	10	0.2143	0.9001	0.1735	0.2377
DLD_3	1.5	12	0.3571	1.5068	0.0403	0.3576
DLD_4	2.5	14	0.1429	2.4262	0.1946	0.1307
DLD_5	3	14	0.1786	2.9650	0.0281	0.1834

Table 2. The proposed MDLD approach is compared in terms of SDR (dB) with MoL-EM_hard, MoWL-EM_hard and Hyvärinen's approach and the average SDR of the mixtures.

	G	roove I	Datase	t	Latino Dataset			
	s_1	s_2	s_3	s_4	s_1	s_2	s_3	s_4
Mixed								
Signals	-30.02	-10.25	-6.14	-21.24	-2.47	-11.8	-2.04	-9.14
MDLD-EM	4.20	-4.32	-0.75	2.07	9.31	-0.04	8.17	3.53
MoWL-EM_hard	4.32	-4.35	-1.16	3.27	8.72	-0.04	8.13	3.03
MoL-EM_hard	2.85	-4.47	-0.86	3.28	8.65	-0.075	8.15	3.51
Hyvärinen	3.79	-3.72	-1.13	1.49	10.03	-1.74	8.16	3.42

References

- Jammalamadaka, S., Sengupta, A.: Topics in Circular Statistics. World Scientific (2001)
- 2. Bentley, J.: Modelling circular data using a mixture of von Mises and uniform distributions. Simon Fraser University, MSc thesis. (2006)
- 3. Mardia, K., Kanti, V., Jupp, P.: Directional Statistics. Wiley (1999)
- Dhillon, I., Sra, S.: Modeling Data using Directional Distributions. Technical report, Technical Report TR-03-06, University of Texas at Austin, Austin, TX (2003)
- Lewicki, M., Sejnowski, T.: Learning Overcomplete Representations. Neural Computation 12 (2000) 337–365
- Jammalamadaka, S., Kozubowski, T.: A new family of Circular Models: The Wrapped Laplace distributions. Technical report, Technical Report No. 61, Department of Mathematics, University of Nevada, Reno, NV, 2002. (2002)
- Mitianoudis, N., Stathaki, T.: Underdetermined Source Separation using Mixtures of Warped Laplacians. In: International Conference on Independent Component Analysis and Source Separation (ICA), London, UK (2007)
- Mitianoudis, N., Stathaki, T.: Batch and Online Underdetermined Source Separation using Laplacian Mixture Models. IEEE Transactions on Audio, Speech and Language Processing 15(6) (2007) 1818–1832
- 9. Moon, T., Stirling, W.: Mathematical Methods and Algorithms for Signal Processing. Prentice Hall, Upper Saddle River, N.J. (2000)
- Bilmes, J.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Mixture Models. Technical report, Department of Electrical Engineering and Computer Science, U.C. Berkeley, California (1998)
- Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S.: Clustering on the Unit Hypersphere using von Mises-Fisher Distributions. Journal of Machine Learning Research 6 (2005) 1345 – 1382
- Devroye, L.: Non-Uniform Random Variate Generation. New York: Springer-Verlag (1986)
- Hyvärinen, A.: Independent Component Analysis in the presence of Gaussian Noise by Maximizing Joint Likelihood. Neurocomputing 22 (1998) 49–67
- Févotte, C., Gribonval, R., Vincent, E.: BSS EVAL Toolbox User Guide. Technical report, IRISA Technical Report 1706, Rennes, France, April 2005, http://www.irisa.fr/metiss/bss eval/
- Vincent, E., Gribonval, R., Fevotte, C., Nesbit, A., Plumbley, M., Davies, M., Daudet, L.: BASS-dB: the blind audio source separation evaluation database. Available at http://bass-db.gforge.inria.fr/BASS-dB/