Tracking Multiple Audio Sources With the von Mises Distribution and Variational EM

Yutong Ban, Xavier Alameda-Pineda, Christine Evers, and Radu Horaud

I. Supplemental Material

We provide four appendices that are referenced in the paper.

Appendix A
Derivation of the E-S step

In order to obtain the formulae for the E-S step, we start from its definition in (9):

$$q(\boldsymbol{s}_t) \propto \exp\Big(\mathrm{E}_{q(\boldsymbol{z}_t)} \log p(\boldsymbol{s}_t, \boldsymbol{z}_t | \boldsymbol{y}_{1:t})\Big). \tag{20}$$

We now use the decomposition in (1) to write:

$$q(\boldsymbol{s}_t) \propto \exp\Big(\mathrm{E}_{q(\boldsymbol{z}_t)} \log p(\boldsymbol{y}_t | \boldsymbol{s}_t, \boldsymbol{z}_t)\Big) p(\boldsymbol{s}_t | \boldsymbol{y}_{1:t-1}). \tag{21}$$

Let us now develop the expectation:

$$\begin{aligned}
&\mathrm{E}_{q(\boldsymbol{z}_t)} \log p(\boldsymbol{y}_t | \boldsymbol{s}_t, \boldsymbol{z}_t) \\
&= \mathrm{E}_{q(\boldsymbol{z}_t)} \sum_{m=1}^{M_t} \log p(y_{tm} | \boldsymbol{s}_t, z_{tm}) \\
&= \sum_{m=1}^{M_t} \mathrm{E}_{q(z_{tm})} \log p(y_{tm} | \boldsymbol{s}_t, z_{tm}) \\
&= \sum_{m=1}^{M_t} \sum_{n=0}^{N} q(z_{tm} = n) \log p(y_{tm} | \boldsymbol{s}_t, z_{tm} = n) \\
&= \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tnm} \log p(y_{tm} | s_{tn}, z_{tm} = n) \\
&= \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tnm} \log \mathcal{M}(y_{tm}; s_{tn}, \omega_{tm}\kappa_y) \\
&\overset{\boldsymbol{s}_t}{=} \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tnm} \omega_{tm} \kappa_y \cos(y_{tm} - s_{tn}),
\end{aligned}$$

where $\overset{\boldsymbol{s}_t}{=}$ denotes the equality up to an additive constant that does *not* depend on $\boldsymbol{s}_t$. Such a constant would become a multiplicative constant after the exponentiation in (21), and therefore can be ignored.

By replacing the developed expectation together with (12) we obtain:

$$q(\boldsymbol{s}_t) \propto \exp\Big(\sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tnm} \omega_{tm} \kappa_y \cos(y_{tm} - s_{tn})\Big) \prod_{n=0}^{N} \mathcal{M}(s_{tn}; \mu_{t-1,n}, \tilde{\kappa}_{t-1,n}),$$

which can be rewritten as:

$$q(\boldsymbol{s}_t) \propto \prod_{n=0}^{N} \exp\Big(\sum_{m=1}^{M_t} \alpha_{tnm} \omega_{tm} \kappa_y \cos(y_{tm} - s_{tn}) \tag{22}$$

$$+ \tilde{\kappa}_{t-1,n} \cos(s_{tn} - \mu_{t-1,n})\Big). \tag{23}$$

(23) is important since it demonstrates that the *a posteriori* pdf of $\boldsymbol{s}_t$ is separable on n and therefore independent for each speaker. In addition, it allows us to rewrite the *a posteriori* pdf for each speaker, i.e., of s_{tn} as a von Mises distribution by using the harmonic addition theorem, thus obtaining

$$q(\boldsymbol{s}_t) = \prod_{n=0}^{N} q(s_{tn}) = \prod_{n=0}^{N} \mathcal{M}(s_{tn}; \mu_{tn}, \kappa_{tn}), \tag{24}$$

with μ_{tn} and κ_{tn} defined as in (14) and (15).

Appendix B
Derivation of the E-Z step

Similarly to the previous section, and in order to obtain the closed-form solution of the E-Z step, we start from its definition in (8):

$$q(\boldsymbol{z}_t) \propto \exp\Big(\mathrm{E}_{q(\boldsymbol{s}_t)} \log p(\boldsymbol{s}_t, \boldsymbol{z}_t | \boldsymbol{y}_{1:t})\Big), \tag{25}$$

and we use the decomposition in (1),

$$q(\boldsymbol{z}_t) \propto \exp\Big(\mathrm{E}_{q(\boldsymbol{s}_t)} \log p(\boldsymbol{y}_t | \boldsymbol{s}_t, \boldsymbol{z}_t)\Big) p(\boldsymbol{z}_t). \tag{26}$$

Since both the observation likelihood and the prior distribution are separable on z_{tm}, we can write:

$$q(\boldsymbol{z}_t) \propto \prod_{m=1}^{M_t} \exp\Big(\mathrm{E}_{q(\boldsymbol{s}_t)} \log p(y_{tm} | \boldsymbol{s}_t, z_{tm})\Big) p(z_{tm}), \tag{27}$$

proving that the *a posteriori* pdf is also separable on m.

We can thus analyze the posterior of each z_{tm} separately, by computing $q(z_{tm} = n)$:

$$q(z_{tm} = n) \propto \exp\Big(\mathrm{E}_{q(\boldsymbol{s}_t)} \log p(y_{tm} | \boldsymbol{s}_t, z_{tm} = n)\Big) p(z_{tm} = n)$$

Let us first compute the expectation for $n \neq 0$:

$$
\begin{aligned}
&\mathrm{E}_{q(\boldsymbol{s}_t)} \log p(y_{tm}|\boldsymbol{s}_t, z_{tm}=n)\\
&= \mathrm{E}_{q(s_{tn})} \log p(y_{tm}|s_{tn}, z_{tm}=n)\\
&= \mathrm{E}_{q(s_{tn})} \log \mathcal{M}(y_{tm}; s_{tn}, \omega_{tm}\kappa_y)\\
&\overset{z_{tm}}{=} \int_0^{2\pi} q(s_{tn}) \omega_{tm}\kappa_y \cos(y_{tm}-s_{tn}) \mathrm{d}s_{tn}\\
&= \frac{\omega_{tm}\kappa_y}{2\pi I_0(\omega_{tm}\kappa_y)} \int_0^{2\pi} \exp\Big(\cos(s_{tn}-\mu_{tn})\Big) \cos(s_{tn}-y_{tm}) \mathrm{d}s_{tn}\\
&= \omega_{tm}\kappa_y A(\omega_{tm}\kappa_y) \cos(y_{tm}-\mu_{tn}),
\end{aligned}
$$

where for the last line we used the following variable change $\bar{s} = s_{tn} - \mu_{tn}$ and the definition of I_1 and A.

The case $n = 0$ is even easier since the observation distribution is a uniform: $\mathrm{E}_{q(s_{tn})} \log p(y_{tm}|s_{tn}, z_{tm}=n) = \mathrm{E}_{q(s_{tn})} - \log 2\pi = -\log(2\pi)$.

By using the fact that the prior distribution on z_{tm} is denoted by $p(z_{tm}=n) = \pi_n$, we can now write the a posteriori distribution as $q(z_{tm}=n) \propto \pi_n \beta_{tmn}$ with:

$$
\beta_{tmn} = \begin{cases} \omega_{tm}\kappa_y A(\omega_{tm}\kappa_y)\cos(y_{tm}-\mu_{tn}) & n \neq 0 \\ 1/2\pi & n = 0 \end{cases},
$$

thus leading to the results in (16) and (3).

APPENDIX C
DERIVATION OF THE M STEP

In order to derive the M step, we need first to compute the Q function in (10),

$$
\begin{aligned}
Q(\Theta, \tilde{\Theta}) &= \mathrm{E}_{q(\boldsymbol{s}_t)q(\boldsymbol{z}_t)} \Big\{ \log p(\boldsymbol{y}_t, \boldsymbol{s}_t, \boldsymbol{z}_t | \boldsymbol{y}_{1:t-1}, \Theta) \Big\}\\
&= \mathrm{E}_{q(\boldsymbol{s}_t)q(\boldsymbol{z}_t)} \Big\{ \underbrace{\log p(\boldsymbol{y}_t | \boldsymbol{s}_t, \boldsymbol{z}_t, \Theta)}_{\kappa_y} +\\
&= + \underbrace{\log p(\boldsymbol{z}_t|\Theta)}_{\pi_n' s} + \underbrace{\log p(\boldsymbol{s}_t | \boldsymbol{y}_{1:t-1}, \Theta)}_{\kappa_d} \Big\},
\end{aligned}
$$

where each parameter is show below the corresponding term of the Q function. Let us develop each term separately.

A. Optimizing κ_y

$$
\begin{aligned}
Q_{\kappa_y} &= \mathrm{E}_{q(\boldsymbol{s}_t)q(\boldsymbol{z}_t)} \Big\{ \log \prod_{m=1}^{M_t} p(y_{tm}|\boldsymbol{s}_t, z_{tm}) \Big\}\\
&= \sum_{m=1}^{M_t} \mathrm{E}_{q(\boldsymbol{s}_t)q(z_{tm})} \Big\{ \log p(y_{tm}|\boldsymbol{s}_t, z_{tm}) \Big\}\\
&= \sum_{m=1}^{M_t} \mathrm{E}_{q(\boldsymbol{s}_t)} \sum_{n=0}^{N} \alpha_{tmn} \Big\{ \log p(y_{tm}|\boldsymbol{s}_t, z_{tm}=n) \Big\}\\
&= \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tmn} \mathrm{E}_{q(s_{tn})} \Big\{ \log \mathcal{M}(y_{tm}; s_{tn}, \omega_{tm}\kappa_y) \Big\}\\
&= \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tmn} \int_0^{2\pi} q(s_{tn}) (\omega_{tm}\kappa_y \cos(y_{tm}-s_{tn})\\
&\qquad - \log(I_0(\omega_{tm}\kappa_y))) \mathrm{d}s_{tn}\\
&= \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tmn} \Big(\omega_{tm}\kappa_y \cos(y_{tm}-\mu_{tn}) A(\kappa_{tn}) - \log(I_0(\omega_{tm}\kappa_y)) \Big),
\end{aligned}
$$

and by taking the derivative with respect to κ_y we obtain:

$$
\frac{\partial Q}{\partial \kappa_y} = \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tmn} \omega_{tm} \Big(\cos(y_{tm}-\mu_{tn}) A(\kappa_{tn}) - A(\omega_{tm}\kappa_y) \Big),
$$

which corresponds to what was announced in the manuscript.

B. Optimizing π_n's

$$
\begin{aligned}
Q_{\pi_n} &= \mathrm{E}_{q(\boldsymbol{s}_t)q(\boldsymbol{z}_t)} \Big\{ \log \prod_{m=1}^{M_t} p(z_{tm}) \Big\}\\
&= \sum_{m=1}^{M_t} \mathrm{E}_{q(z_{tm})} \Big\{ \log p(z_{tm}) \Big\}\\
&= \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tmn} \Big\{ \log p(z_{tm}=n) \Big\}\\
&= \sum_{m=1}^{M_t} \sum_{n=0}^{N} \alpha_{tmn} \Big\{ \log \pi_n \Big\}
\end{aligned}
$$

This is the same formulae that is correct for any mixture model, and therefore the solution is standard and corresponds to the one reported in the manuscript.

C. Optimizing κ_d

$$
\begin{aligned}
Q_{\kappa_d} &= \mathrm{E}_{q(\boldsymbol{s}_t)q(\boldsymbol{z}_t)}\Big\{ \log \prod_{n=1}^{N} p(s_{tn}|\boldsymbol{y}_{1:t-1})\Big\} \\
&= \sum_{n=1}^{N} \mathrm{E}_{q(s_{tn})}\Big\{ \log \mathcal{M}(s_{tn};\mu_{t-1,n},\tilde{\kappa}_{t-1,n})\Big\} \\
&= \sum_{n=1}^{N} \mathrm{E}_{q(s_{tn})}\Big\{ -\log I_0(\tilde{\kappa}_{t-1,n}) + \tilde{\kappa}_{t-1,n}\cos(s_{tn}-\mu_{t-1,n})\Big\} \\
&= \sum_{n=1}^{N} -\log I_0(\tilde{\kappa}_{t-1,n}) + \tilde{\kappa}_{t-1,n}\cos(\mu_{tn}-\mu_{t-1,n})A(\kappa_{tn}),
\end{aligned}
$$

where the dependency on κ_d is implicit in $\tilde{\kappa}_{t-1,n} = A^{-1}(A(\kappa_{t-1,n})A(\kappa_d))$.

By taking the derivative with respect to κ_d we obtain:

$$
\frac{\partial Q}{\partial \kappa_d} = \sum_{n=1}^{N}\Big(A(\kappa_{tn})\cos(\mu_{tn}-\mu_{t-1,n}) - A(\tilde{\kappa}_{t-1,n})\Big)\frac{\partial \tilde{\kappa}_{t-1,n}}{\partial \kappa_d}
$$

with

$$
\frac{\partial \tilde{\kappa}_{t-1,n}}{\partial \kappa_d} = \tilde{A}(A(\kappa_{t-1,n})A(\kappa_d))A(\kappa_{t-1,n})\frac{I_2(\kappa_d)I_0(\kappa_d) - I_1^2(\kappa_d)}{I_0^2(\kappa_d)},
$$

where $\tilde{A}(a) = \mathrm{d}A^{-1}(a)/\mathrm{d}a = (2-a^2+a^4)/(1-a^2)^2$.

By denoting the previous derivative as $B(\kappa_d) = \frac{\partial \tilde{\kappa}_{t-1,n}}{\partial \kappa_d}$, we obtain the expression in the manuscript.

Appendix D
Derivation of the Birth Probability

In this section we derive the expression for τ_j by computing the integral (17). Using the probabilistic model defined, we can write (the index j is omitted):

$$
\begin{aligned}
&\int p(\hat{y}_{t-L:t}, s_{t-L:t})\mathrm{d}s_{t-L:t} \\
&\quad = \int \prod_{\tau=-L}^{0} p(\hat{y}_{t+\tau}|s_{t+\tau}) \prod_{\tau=-L+1}^{0} p(s_{t+\tau}|s_{t+\tau-1})p(s_{t-L})\mathrm{d}s_{t-L:t}
\end{aligned}
$$

We will first marginalize s_{t-L}. To do that, we notice that if $p(s_{t-L})$ follows a von Mises with mean $\hat{\mu}_{t-L}$ and concentration $\hat{\kappa}_{t-L}$, then we can write:

$$
\begin{aligned}
&p(\hat{y}_{t-L}|s_{t-L})p(s_{t-L}) \\
&\quad = \mathcal{M}(\hat{y}_{t-L}; s_{t-L}, \hat{\omega}_{t-L}\kappa_y)\mathcal{M}(s_{t-L};\hat{\mu}_{t-L},\hat{\kappa}_{t-L}) \\
&\quad = \mathcal{M}(s_{t-L};\bar{\mu}_{t-L},\bar{\kappa}_{t-L})\frac{I_0(\bar{\kappa}_{t-L})}{2\pi I_0(\hat{\omega}_{t-L}\kappa_y)I_0(\hat{\kappa}_{t-L})}
\end{aligned}
$$

with

$$
\begin{aligned}
\bar{\mu}_{t-L} &= \tan^{-1}\left(\frac{\hat{\omega}_{t-L}\kappa_y \sin\hat{y}_{t-L} + \hat{\kappa}_{t-L}\sin\hat{\mu}_{t-L}}{\hat{\omega}_{t-L}\kappa_y \cos\hat{y}_{t-L} + \hat{\kappa}_{t-L}\cos\hat{\mu}_{t-L}}\right), \\
\bar{\kappa}_{t-L}^2 &= (\hat{\omega}_{t-L}\kappa_y)^2 + \hat{\kappa}_{t-L}^2 + 2\hat{\omega}_{t-L}\kappa_y\hat{\kappa}_{t-L}\cos(\hat{y}_{t-L}-\hat{\mu}_{t-L}),
\end{aligned}
$$

where we used the harmonic addition theorem.

Now we can effectively compute the marginalization. The two terms involving s_{t-L} are:

$$
\begin{aligned}
&\int \mathcal{M}(s_{t-L+1}; s_{t-L},\kappa_d)\mathcal{M}(s_{t-L};\bar{\mu}_{t-L},\bar{\kappa}_{t-L})ds_{t-L} \\
&\approx \mathcal{M}(s_{t-L+1};\hat{\mu}_{t-L+1},\hat{\kappa}_{t-L+1})
\end{aligned}
$$

with

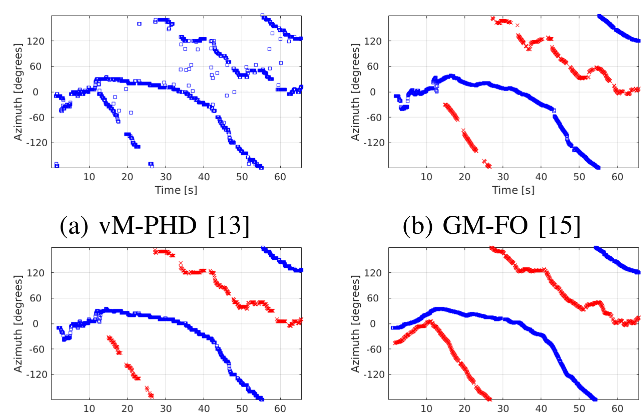
$$
\begin{aligned}
\hat{\mu}_{t-L+1} &= \bar{\mu}_{t-L}, \\
\hat{\kappa}_{t-L+1} &= A^{-1}(A(\bar{\kappa}_{t-L})A(\kappa_d)).
\end{aligned}
$$

Therefore, the marginalization with respect to s_{t-L} yields the following result:

$$
\begin{aligned}
&\int p(\hat{y}_{t-L:t}, s_{t-L:t})\mathrm{d}s_{t-L:t} \\
&\quad = \int \prod_{\tau=-L}^{0} p(\hat{y}_{t+\tau}|s_{t+\tau}) \prod_{\tau=-L+1}^{0} p(s_{t+\tau}|s_{t+\tau-1})p(s_{t-L})\mathrm{d}s_{t-L:t} \\
&\quad = \frac{I_0(\bar{\kappa}_{t-L})}{2\pi I_0(\hat{\omega}_{t-L}\kappa_y)I_0(\hat{\kappa}_{t-L})}\int \prod_{\tau=-L+1}^{0} p(\hat{y}_{t+\tau}|s_{t+\tau})\times \\
&\qquad \prod_{\tau=-L+2}^{0} p(s_{t+\tau}|s_{t+\tau-1})p(s_{t-L+1})\mathrm{d}s_{t-L+1:t}.
\end{aligned}
$$

Since we have already seen that $p(s_{t-L+1})$ is also a von Mises distribution, we can use the same reasoning to marginalize with respecto to s_{t-L+1}. This strategy yields to the recursion presented in the main text.

Appendix E
Results with Errors



(a) vM-PHD [13] (b) GM-FO [15] (c) vM-VEM (proposed) (d) ground-truth trajectories

Fig. 1: Results obtained with recording #3 from Task 6 of the LOCATA dataset. Different colors represent different audio sources. Note that vM-PHD is unable to associate sources with trajectories.