On the Properties of the Spherical Analogue of the Epanechnikov Kernel for Directional Kernel Density Estimation

Zain Jabbar
Dept. of Mathematics
Univ. of Hawai'i at Manoa
Honolulu, USA
zjabbar@hawaii.edu

Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
Uwe.Hanebeck@kit.edu

Andrey A. Popov
Dept. of Info. & Comp. Sci.
Univ. of Hawai'i at Manoa
Honolulu, USA
apopov@hawaii.edu

***Abstract*—A recent advancement in convergent particle filters is the family of kernel density estimation-based class of ensemble mixture model filters, such as the ensemble Gaussian mixture and ensemble Epanechnikov mixture filters. This work seeks to advance the understanding of directional Kernel Density Estimation (KDE) for later use in directional filtering with directional kernel density estimates. Optimal KDE in high dimensions relies not only on optimal estimates of parameters such as the bandwidth, but also on making use of an optimal kernel. For Euclidean space, the optimal kernel is the Epanechnikov kernel, but for distributions on the sphere this kernel is known, but has largely been forgotten to time. We analyze this kernel and term it the Hall-Watson-Cabrera kernel (HWC) after its original discoverers. The contributions of this work are three-fold: (i) we prove, that the HWC kernel is exactly the restriction of the Epanechnikov kernel onto the sphere, (ii) we provide a practical numerical method to sample from a HWC mixture model, and (iii) we empirically show the HWC kernel's efficiency in the high-dimensional setting.**

***Index Terms*—Kernel density estimation, directional statistics, Epanechnikov kernel, Hall-Watson-Cabrera kernel.**

I. Introduction

Directional statistics [1], [2] is concerned with problems related to estimating quantities of directional data. Rotation, travel direction, and other quantities all fall into these categories. States of systems bounded by conservation of energy also belongs into such a category. Kernel density estimation [3] is concerned with approximating densities from samples in a manner than becomes convergent as the number of samples increases. This worked is concerned specifically with the intersection of these two fields, directional kernel density estimation [2], [4].

In the context of directional state estimation and filtering, there have been some developments towards general non-linear filtering, such as some nonlinear filters [5] and unscented filters [6]. Additionally there has been a lot of effort towards developing machinery adjacent to the filtering problem such as deterministic sampling [7], [8]. This work also aims to develop machinery and methodology for eventual applications to the non-linear directional filtering problem by targeting the research gap of describing the optimal kernel for directional kernel density estimation.

Epanechnikov [9] (read: Yepanechnikov), showed that under certain simplifying assumptions, a compactly supported kernel was more optimal for kernel density estimation than the (still commonly used) Gaussian kernel. Unlike the Epanechnikov kernel, its spherical analogue seems to be a lot more lost to the literature: cited, but somewhat forgotten. For instance the work of Hodges [10] rediscovers this kernel while simultaneously citing the work in which it was actually derived. The work of Kamelä [11] seems to be one of the last works to actually explicitly make use of this kernel to the authors' knowledge. Other works like [12] have incorrectly identified the spherical analogue of the Epanechnikov kernel. The original work deriving the spherical analogue of the Epanechnikov kernel, that of Hall, Watson, and Cabrera [13] claims that:

> We shall show that it is possible to improve on the latter estimate by using an 'optimal' kernel, just as Bartlett (1963) [14] and Epanechnikov (1969) [9] showed that Euclidean-data estimates have optimal kernels.

with references added by the authors. We trivially show that the kernel that they derive is not "just as" the Epanechnikov kernel, but in fact that it is *exactly* the multivariate Epanechnikov kernel restricted to the sphere. We then subsequently derive properties about this kernel and verify them using numerical experiments.

II. Background

In this work, we specifically follow the formulation of directional kernel density estimation given in [4]. Given some distribution f on the sphere $\mathbb{S}^{n-1}$, and some collection of N samples

$$[\mu_1, \mu_2, \dots \mu_N], \tag{1}$$

we can build a kernel density estimate of the form

$$\widetilde{f}(x) = \frac{c_{n,h,\mathcal{K}}}{N} \sum_{i=1}^{N} \mathcal{K}\left(\frac{1 - x^T \mu_i}{h^2}\right), \tag{2}$$

ZJ and AAP are supported by startup funds from the University of Hawai'i at Manoa.

where $\mathcal{K}$ is a kernel, $c_{n,h,\mathcal{K}}$ is a normalizing constant, and h is a bandwidth parameter. Following [3] and [4], as is common in kernel density estimation literature, we wish to find the bandwidth parameter h that minimizes the mean integral squared error (MISE), given by,

$$\mathrm{MISE}(\widetilde{f}, f) = \mathbb{E}_{\mu\sim f}\left[\int_{\mathbb{S}^{n-1}} \left(f(x) - \widetilde{f}(x)\right)^2 \mathrm{d}x\right], \quad (3)$$

where under certain simplifying assumptions we can build an asymptotically converging approximation called the asymptotic mean integral squared error,

$$\mathrm{AMISE}(\mathcal{K}) = \frac{\alpha_{\mathcal{K}}^2}{\beta_{\mathcal{K}}^2} R(f) h^4 + \frac{c_{n,h,\mathcal{K}}}{N}\frac{\gamma_{\mathcal{K}}}{\beta_{\mathcal{K}}} \quad (4)$$

where the asymptotic variable is the number of samples N, meaning that the second term tends to zero, and the constants are defined as,

$$\begin{aligned}
c_{n,h,\mathcal{K}} &\approx \left(h^{n-1}\lambda_{\mathcal{K}}\right)^{-1}, & \lambda_{\mathcal{K}} &= 2^{\frac{n-3}{2}}\omega_{n-1}\beta_{\mathcal{K}},\\
\omega_{n-1} &= \frac{\sqrt{2\pi}^{n-1}}{\Gamma(\frac{n-1}{2})}, & \alpha_{\mathcal{K}} &= \int_0^\infty v^{\frac{n-1}{2}}\mathcal{K}(v)\mathrm{d}v,\\
\beta_{\mathcal{K}} &= \int_0^\infty v^{\frac{n-3}{2}}\mathcal{K}(v)\mathrm{d}v, & \gamma_{\mathcal{K}} &= \int_0^\infty v^{\frac{n-3}{2}}\mathcal{K}(v)^2\mathrm{d}v,
\end{aligned} \quad (5)$$

and the curvature constant $R(f)$ that depends on the original distribution is typically taken to be that of the Fisher-von Mises-Langevin distribution and is given by

$$R(f) = \frac{\kappa^{n/2}\left(2(n-1)I_{n/2}(2\kappa) + \kappa(n+1)I_{n/2+1}(2\kappa)\right)}{2^{n+1}\pi^{n/2}(n-1)I_{n/2-1}(\kappa)^2}, \quad (6)$$

where I is the modified Bessel function of the first kind and where κ is the concentration defined as

$$\kappa = A_n^{-1}\left(\|\mathbb{E}_{x\sim f}[x]\|_2\right), \quad A_n(\kappa) = \underset{i=0}{\overset{\infty}{K}} \frac{1}{(n+2i)\kappa^{-1}}, \quad (7)$$

where A_n is written in Gauss's continued fraction notation and can be computed through a modification of Lentz's algorithm [15], and where the concentration parameter κ can be approximated by first approximating the expected value, then approximating the inverse of A_n [16]. We note that this results in a *biased* estimator of the concentration [1], [17], and that no unbiased estimator for small sample size is known. More sophisticated estimates of the curvature constant and concentration parameter are outside the scope of this work.

It is important to note that we can compute the optimal bandwidth for an arbitrary kernel by the following formulation,

$$h_{\mathcal{K}} = \left(\frac{\gamma\Gamma\left(\frac{n+3}{2}\right)}{\sqrt{2\pi}^{n-1}(n+1)\alpha^2 N R(f)}\right)^{\frac{1}{n+3}}. \quad (8)$$

It is important to note that the kernel $\mathcal{K}$ is not a distribution, and that when h is taken to be the optimal $h_{\mathcal{K}}$, it is dependent on the scaling of the distribution and distribution parameter. This means that any scaling of the kernel or kernel input defines an equivalent kernel that does not have an effect on the kernel density estimate.

Remark II.1 (Kernel equivalence class)**.** *Thus it is trivial to see that the kernel,*

$$a\,\mathcal{K}(bv), \quad (9)$$

does not change the kernel density estimate, meaning that scalar multiplication of the kernel, and scalar multiplication of the kernel input represent an equivalence class of effectively the same kernel.

Finally we note some properties that this optimal kernel should have:

1) The kernel must be non-negative, and finite, meaning that,
$$\forall v \geq 0,\ (0 \leq \mathcal{K}(v) < \infty). \quad (10)$$
2) The kernel must be monotonically decreasing, meaning that,
$$\forall w \geq v \geq 0,\ (\mathcal{K}(w) \leq \mathcal{K}(v)), \quad (11)$$
as the kernel needs to converge to the Dirac delta as the number of samples goes to infinity (and thus the bandwidth goes to zero).
3) The kernel must satisfy,
$$\beta_{\mathcal{K}} < \infty, \quad (12)$$
which many authors interpret as the fact that the kernel "is rapidly decaying, more rapidly than any polynomial" (from [4] page 58), which is not *strictly* true, as the kernel does not have to be smooth.

Of final note is the standard kernel, that of the Gaussian restricted to the unit sphere,

$$\mathcal{F}v\mathcal{ML}(v) = e^{-v}, \quad (13)$$

which is known as the Fisher-von Mises-Langevin kernel. This kernel is by far the most popular kernel found in the literature for its nice and simple properties that are begotten from its Gaussian lineage.

III. Hall-Watson-Cabrera Kernel

The immediate obvious question that arises is, which functional form, given our constraints, minimizes the AMISE? The kernel identified by Hall, Watson, and Cabrera in [13],

$$\mathcal{HWC}(v) = \begin{cases} 1 - v & v \leq 1 \\ 0 & \text{otherwise} \end{cases}, \quad (14)$$

is precisely that kernel, meaning that

$$\mathcal{HWC} = \underset{\mathcal{K}}{\arg\min}\, \mathrm{AMISE}(\mathcal{K}), \quad (15)$$

which we term the Hall-Watson-Cabrera kernel, visualized in relation to the Fisher-von Mises-Langevin Kernel for $\mathbb{S}^1$ in fig. 1.

We now show that this kernel is precisely the spherical analogue of the Epanechnikov kernel for spherical distributions,

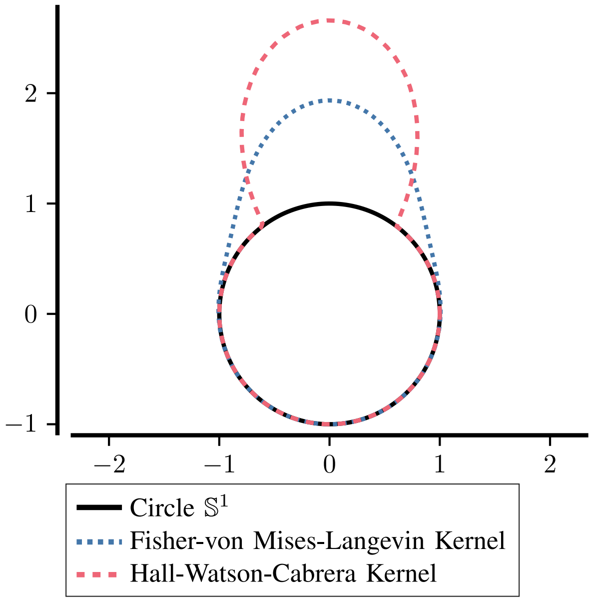

Fig. 1. A visual representation of the Fisher-von Mises-Langevin kernel and the Hall-Watson-Cabrera kernel on the circle $\mathbb{S}^1$. The mean direction is taken to be the point $[0, 1]^T$ and the bandwidth is taken to be $\sqrt{5}^{-1}$.

Theorem III.1. *The Hall-Watson-Cabrera kernel eq.* (14) *is the restriction of the Epanechnikov kernel,*

$$\mathcal{E}(x, \mu_i) = \begin{cases} n + 4 - \|x - \mu_i\|_2^2 & \|x - \mu_i\|_2^2 \leq n + 4 \\ 0 & \textit{otherwise} \end{cases}, \tag{16}$$

onto the surface of the sphere, $\mathbb{S}^{n-1}$.

Proof. Observe first that on the sphere $\mathbb{S}^{n-1}$, that $x^T x = 1$, that $\mu_i^T \mu_i$, and that $x^T \mu_i = \mu_i^T x$. From this it is trivial to observe that

$$\mathcal{E}(x, \mu_i) = \begin{cases} n + 2 + 2x^T \mu_i & 2(1 - x^T \mu_i) \leq n + 4 \\ 0 & \text{otherwise} \end{cases}, \tag{17}$$

which we can transform into the kernel with input argument,

$$\mathcal{K}(v) = \begin{cases} n + 4 - 2h^2 v & 2h^2 v \leq n + 4 \\ 0 & \text{otherwise} \end{cases}, \tag{18}$$

from which it is trivial to see that a kernel of the same kernel family, by scaling the function and parameter is the kernel $\mathcal{HWC}$, as required. □

From this simple observation, we can make the following conjecture:

Conjecture III.1 (Restricted Epanechnikov)**.** *Given samples from a probability distribution whose support is some compact set* Ω*, which is a restriction of Euclidean space* $\mathbb{R}^n$*, the optimal kernel in terms of* AMISE *is the Epanechnikov kernel restricted to the same support.*

IV. Properties of the Hall-Watson-Cabrera Kernel

In this section we describe the constants required for deriving the bandwidth, the efficiency of the Fisher-von Mises-Langevin Kernel in terms of the Hall-Watson-Cabrera kernel, a method for sampling from a Hall-Watson-Cabrera mixture model, and a quadrature rule for approximating the Fisher-von Mises-Langevin kernel as a simple Hall-Watson-Cabrera mixture.

For the HWC kernel, the constants from eq. (5) can be computed as follows,

$$\alpha_{\mathcal{HWC}} = \frac{4}{(n+3)(n+1)}, \quad \beta_{\mathcal{HWC}} = \frac{4}{(n+1)(n-1)}, \quad \gamma_{\mathcal{HWC}} = \frac{16}{(n+3)(n+1)(n-1)}. \tag{19}$$

A. Fisher-von Mises-Langevin Efficiency

The efficiency of a multivariate kernel in Euclidean space is determined exactly as a ratio of AMISE in terms of the Epanechnikov kernel [3], [18]. We therefore propose a directional kernel density estimate to the efficiency in terms of a ratio of AMISE eq. (4) of the $\mathcal{HWC}$ kernel

$$\text{Eff}(\mathcal{K}) = \left(\frac{\text{AMISE}(\mathcal{HWC})}{\text{AMISE}(\mathcal{K})} \right)^{\frac{n+3}{4}}, \tag{20}$$

which for the $\mathcal{F}v\mathcal{ML}$ kernel, by simple algebraic manipulation is

$$\text{Eff}(\mathcal{F}v\mathcal{ML}) = 2^{n+2}(n+3)^{-\frac{n}{2}-\frac{3}{2}}\Gamma\left(\frac{n+5}{2}\right), \tag{21}$$

which is the spherical analogue of the efficiency trivially derived in [18].

B. Sampling from a Hall-Watson-Cabrera mixture

Following the work of Ulrich [19], Wood [20], and Pinzón and Jung [21], we derive a method of sampling from a generic isotropic kernel mixture on $\mathbb{S}^{n-1}$. For the Fisher-von Mises-Langevin kernel, Wood [20] provides a rejection sampling technique for generating samples thereof. There are several obstacles preventing us from making use of rejection sampling in the case of the Hall-Watson-Cabrera kernel, chief among them being the fact that rejection sampling is an inherently stochastic affair, meaning that it is not computationally efficient. The other simple fact is that it is actually possible to compute a closed-form solution of some functions that are difficult to compute in the case of the Fisher-von Mises-Langevin kernel.

Observe that, as the class of kernels that we are concerned with is isotropic, that we can express any point x on $\mathbb{S}^{n-1}$ in terms of a central starting point μ, a tangential unit direction ε, and an angle θ, in terms of the exponential map on the sphere,

$$x = \cos(\theta)\mu + \sin(\theta)\varepsilon, \tag{22}$$

where we can then rewrite the kernel parameter v as,

$$\begin{aligned} v &= h^{-2}(1 - x^T\mu) \\ &= h^{-2}\left(1 - (\cos(\theta)\mu + \sin(\theta)\varepsilon)^T\mu\right) \\ &= h^{-2}\left(1 - \cos\theta\right), \end{aligned} \tag{23}$$

meaning that by substituting a synthetic time $\tau = \cos\theta$ for $\tau \in [-1, 1]$, it is possible to rewrite,

$$v = h^{-2}(1 - \tau), \tag{24}$$

and rewrite the state purely in terms of τ as,

$$x = \tau\mu + \sqrt{1 - \tau^2}\varepsilon, \tag{25}$$

begetting the fact that any kernel can now be rewritten purely in terms of mean μ, direction ε and the scalar τ, as,

$$\mathcal{K}_{\mu,\varepsilon,h}(\tau) = \mathcal{K}(h^{-2}(1 - \tau)), \tag{26}$$

as an alternate scalar kernel. Following [19]–[21], a sample from any isotropic distribution can be written in the form,

$$\overline{\mathcal{K}}_{\mu,\varepsilon,h}(\tau) = C_{\mathcal{K}_{\mu,\varepsilon,h}}\mathcal{K}_{\mu,\varepsilon,h}(\tau)\underbrace{(1 - \tau^2)^{(n-3)/2}}_{\text{Surface}}, \tag{27}$$

where the first term is a normalizing constant, the second term is the scalar kernel in the direction ε from μ, and the third term accounts for the change in the area of the surface as τ increases.

For the Hall-Watson-Cabrera kernel, the marginal kernel around the mean μ in the direction ε is given by,

$$\mathcal{HWC}_{\mu,\varepsilon,h}(\tau) = \begin{cases} 1 - h^{-2}\left(1 - \tau\right) & h^{-2}\left(1 - \tau\right) < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{28}$$

which is a scalar kernel in the parameter τ. Substituting eq. (28) into eq. (27) we get,

$$\overline{\mathcal{HWC}}_{\mu,\varepsilon,h}(\tau) = C_{\mathcal{HWC}_{\mu,\varepsilon,h}}\mathcal{HWC}_{\mu,\varepsilon,h}(\tau)\underbrace{(1 - \tau^2)^{(n-3)/2}}_{\text{Surface}}, \tag{29}$$

where, through the aide of a computer algebra system, the normalizing constant is given by,

$$\begin{aligned} &\left(C_{\mathcal{HWC}_{\mu,\varepsilon,h}}\right)^{-1} = \\ &\begin{cases} \frac{H}{h^2}B_n & h^2 > 2, \\ \frac{H}{2h^2}\left(2\Delta_n(H) + B_n\right) + \frac{h^{n-3}J}{n-1} & h^2 \leq 2 \end{cases} \\ &B_n = \frac{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \\ &\Delta_n(t) = t \cdot {}_2F_1\left(\frac{1}{2}, \frac{3-n}{2}; \frac{3}{2}; t^2\right), \\ &H = h^2 - 1, \\ &J = (2 - h^2)^{\frac{n-1}{2}}, \end{aligned} \tag{30}$$

where ${}_2F_1$ is the Gauss hypergeometric function for which computational algorithms are provided by [22].

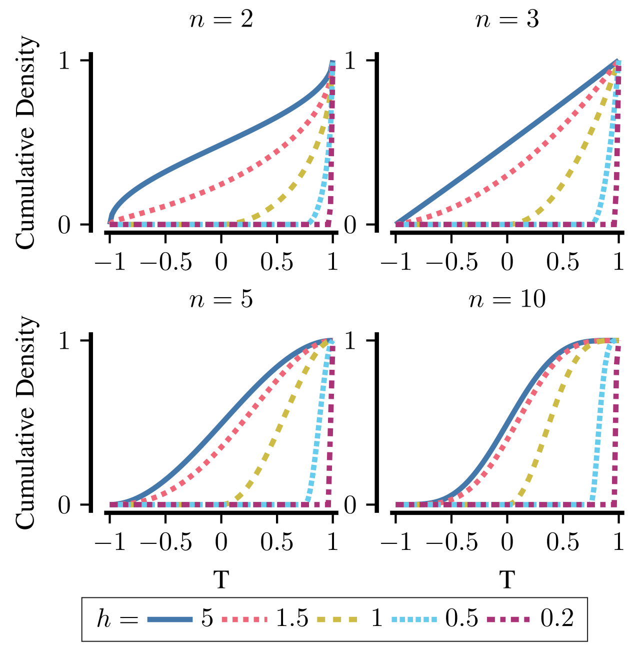

Fig. 2. Cumulative Distribution of the Hall-Watson-Cabrera kernel marginalized to the arbitrary direction ε, for various values of dimension n and bandwidth h. The x-axis spans the CDF input T from -1 corresponding to a point furthest away from the mean, to $T = 1$, which is directly at the mean.

The cumulative distribution function has a closed form, given by,

$$\begin{aligned} &C_{\mathcal{HWC}_{\mu,\varepsilon,h}}F_{\overline{\mathcal{HWC}}_{\mu,\varepsilon,h}}(T) = \\ &\begin{cases} \frac{H(\Delta_n(H)+\Delta_n(T))}{h^2} + \frac{h^{n-1}J-\left(1-T^2\right)^{\frac{n-1}{2}}}{(n-1)h^2} & \begin{matrix} h^2 \leq 2\wedge \\ (h^2+T) > 1 \end{matrix} \\ \frac{H\Delta_n(T)}{h^2} + \frac{H}{2h^2}B_n - \frac{\left(1-T^2\right)^{\frac{n-1}{2}}}{(n-1)h^2} & h^2 > 2 \wedge T > -1 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{31}$$

which can be visually plotted for various values of n and h for all $T \in [-1, 1]$ in fig. 2. The utility of this representation of the cumulative density is in terms of inverse transform (Smirnov) sampling, meaning that a sample from $\overline{\mathcal{HWC}}_{\mu,\varepsilon,h}$ can be obtained as follows,

$$F^{-1}_{\overline{\mathcal{HWC}}_{\mu,\varepsilon,h}}(u), \quad u \sim \mathcal{U}[0, 1], \tag{32}$$

where u is uniformly distributed on the unit interval $[0, 1]$. The inverse of eq. (31) is not readily representable in terms a closed form solution, thus some sort of root finding algorithm is required. There are two simple options for root finding for a monotone continuous function such as eq. (31), one is the binary search method and the other is Newton's method, which requires a derivative of eq. (31), which is exactly the density function, which one can note does not require the evaluation of a hypergeometric function other than for the normalizing constant. This means that in the case of Newton's method, the number of evaluations of a hypergeometric function is exactly one plus the number of steps of the algorithm. In practice, it has been the authors' experience that a simple bisection method has faster convergence properties for larger values of the dimension n, though does require one evaluation of the normalizing constant, one fixed evaluation of the hypergeometric function and one additional hypergeometric function evaluation per step. It is of independent interest to develop alternative, faster, methods for sampling from this probability distribution.

We now move on to the problem of sampling from the Hall-Watson-Cabrera mixture model, given by

$$\widetilde{f}(x) \propto \sum_{i=1}^{N} w_i \mathcal{HWC}\left(\frac{1 - x^T \mu_i}{h_i^2}\right), \tag{33}$$

where the weights $\{w_i\}$ sum to one which are identical in the case of KDE, μ_i are the mean directions which are the samples in the case of KDE, and h_i are the bandwidth parameters which are typically identical and derived by minimizing the AMISE in the case of KDE. The normalizing factor is of no consequence in the case of sampling.

1) For the jth requested sample, sample a random component index λ from the discrete distribution defined by the weights $\{w_i\}$.
2) Sample a random direction ε_j through the following subprocedure due to [21]:
   a) Sample a normal random vector ε_j of dimension n.
   b) Normalize the vector by $\varepsilon_j \leftarrow \varepsilon_j / \|\varepsilon_j\|$
   c) Orient the vector, $\varepsilon_j \leftarrow \varepsilon_j - \mu_\lambda \mu_\lambda^T \varepsilon_j$, such that it is perpendicular to the mean direction μ_λ.
   d) Normalize again by $\varepsilon_j \leftarrow \varepsilon_j / \|\varepsilon_j\|$.
3) Sample a uniform random sample u_j in the interval $[0, 1]$.
4) Using either Newton's method, or the binary search method, compute, a sample from the marginalized directional distribution,

$$\tau_j = F^{-1}_{\overline{\mathcal{HWC}}_{\mu_\lambda, \varepsilon_j, h_\lambda}}(u_j),$$

   using the explicit formulation given in eq. (31).
5) A sample from eq. (33) is therefore given by the expression,

$$x = \tau_j \mu_\lambda + \sqrt{1 - \tau_j^2}\varepsilon_j.$$

C. Mean direction and Concentration

As the kernel is isotropic, it is trivial to see that the mean direction of the distribution induced by the kernel is μ. What about the concentration? This is non-trivial, but can be computed by looking at the expected value of the directional distribution,

$$\int_{-1}^{1} \tau \overline{\mathcal{HWC}}_{\mu,\varepsilon,h}(\tau) \mathrm{d}\tau = \begin{cases} \frac{1}{Hn} & h^2 \geq 2 \\ \frac{Hn\Gamma\left(\frac{n}{2}\right)\left(hH(n-1)\overline{\Delta}_n(H) + 3h^n J\right) + 3\sqrt{\pi}h\Gamma\left(\frac{n+1}{2}\right)}{3n\left(\sqrt{\pi}hH\Gamma\left(\frac{n+1}{2}\right) + \Gamma\left(\frac{n}{2}\right)(hH(n-1)\Delta_n(H) + h^n J)\right)} & \text{otherwise} \end{cases},$$
$$\overline{\Delta}_n(t) = t \cdot {}_2F_1\left(\frac{3}{2}, \frac{3-n}{2}; \frac{5}{2}; t^2\right) \tag{34}$$

which we can see by observation makes the concentration,

$$\kappa_{\mathcal{HWC}} = A_n^{-1}\left(\int_{-1}^{1} \tau \overline{\mathcal{HWC}}_{\mu,\varepsilon,h}(\tau) \mathrm{d}\tau\right), \tag{35}$$

where A_n is defined in eq. (7). Note that this is the concentration as defined for the $\mathcal{F}v\mathcal{ML}$ kernel, as concentration is analogues to, but not identical to the notation of variance and covariance.

Remark IV.1 (Inverse of A_n)**.** *Suppose we want to find κ such that $A_n(\kappa) = \upsilon$. Observe that,*

$$A_n(\kappa) = \frac{1}{n\kappa^{-1} + A_{n+2}(\kappa)} \tag{36}$$

and further observe that, $\lim_{n\to\infty} A_{n+2}(\kappa) = 1$. *The value of κ can therefore be approximated by the following Picard sequence,*

$$\kappa_0 = \frac{n\upsilon}{1-\upsilon},\ \kappa_{i+1} = \frac{n\upsilon}{1 - \upsilon A_{n+2}(\kappa_i)}. \tag{37}$$

This means that,

$$A_n^{-1}(\upsilon) = \lim_{i\to\infty} \kappa_i. \tag{38}$$

D. Quadrature approximation

Similar to (and inspired by) [23], we wish to derive a quadrature relationship between the Hall-Watson-Cabrera kernel and the Fisher-von Mises-Langevin kernel. We first show that it is possible to represent the $\mathcal{F}v\mathcal{ML}$ kernel eq. (13) as an integral of the $\mathcal{HWC}$ kernel.

Theorem IV.1. *The following holds,*

$$\mathcal{F}v\mathcal{ML}(v) = \int_0^\infty x e^{-x} \mathcal{HWC}(v/x) \mathrm{d}x. \tag{39}$$

Proof. Observe the following:

$$\begin{aligned}\int_0^\infty x e^{-x} \mathcal{HWC}(v/x) \mathrm{d}x &= \int_v^\infty (x - v) e^{-x} \mathrm{d}x, \\ &= \mathcal{F}v\mathcal{ML}(v),\end{aligned}$$

as required. □

The integral in eq. (39) can be approximated through a Gauss-Laguerre quadrature [24] in the form,

$$\mathcal{F}v\mathcal{ML}(v) \approx \sum_{i=1}^{M} w_i \mathcal{HWC}(v/\ell_i), \tag{40}$$

where M is the number of components to approximate the integral, w_i are the Gauss-Laguerre weights, and ℓ_i are the roots of the Mth Gauss-Laguerre polynomial. As,

$$\frac{v}{\ell_i} = \frac{1 - x^T\mu}{(h\sqrt{\ell_i})^2}, \tag{41}$$

this means that the approximation in eq. (40) is exactly a Hall-Watson-Cabrera mixture model approximation of the Fisher-von Mises-Langevin kernel with non-equal weights and non-equal bandwidths. Note that the transformation in eq. (39) is the application of an integral transform of Laplace type [25],

$$\mathcal{F}v\mathcal{ML}(v) = v^2\mathcal{L}[x\mathcal{HWC}(x^{-1})](v), \tag{42}$$

meaning that we can observe the following inverse relationship,

$$\mathcal{HWC}(v) = v\mathcal{L}^{-1}[\mathcal{F}v\mathcal{ML}(x)/x^2](1/v), \tag{43}$$

from which a suitable approximation method will get an approximation to the HWC kernel. While various different methods to approxiamte this transform exist [26]–[29], it is not possible for this approximation to be of the same type as that of eq. (40). We now prove this fact:

Theorem IV.2. *There do not exist non-negative weight functions $w(x)$ and a non-negative scaling functions $s(x)$ such that,*

$$\mathcal{HWC}(v) = \int_0^\infty w(x)\mathcal{F}v\mathcal{ML}(s(x)v)\mathrm{d}x. \tag{44}$$

Proof. Observe that the derivative of $\mathcal{HWC}(v)$ in the interval 0 to 1 is a constant -1. Observe also that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}v}\int_0^\infty w(x)\mathcal{F}v\mathcal{ML}(s(x)v)\mathrm{d}x = \\ -\int_0^\infty w(x)s(x)e^{-s(x)v}\mathrm{d}x\end{aligned} \tag{45}$$

meaning that for any $0 < v < 1$, that the gradient is unique unless either the weights and scaling are zero for all x. This means that it is impossible to have a constant gradient no matter the combination of weights and scaling factors. $\square$

In other words, a mixture model of $\mathcal{F}v\mathcal{ML}$ kernels would have to be placed at different offsets, and not just through a reweighting and rescaling. This means that it is possible to efficiently express the $\mathcal{F}v\mathcal{ML}$ kernel in terms of the $\mathcal{HWC}$ kernel through a mixture model no matter the dimension, but the reverse will rely on a dimensionally-dependent quadrature.

V. Numerical Experiments

The first experiment will validate the efficiency of the kernel eq. (21). We construct an experiment where we compute the MISE of two kernel density estimates, that using the $\mathcal{F}v\mathcal{ML}$ kernel and the $\mathcal{HWC}$ kernel. We span the dimension of the sphere from $n = 2$ to $n = 20$ in increments of one. We take $N = 1\,000$ points that are sampled from the $\mathcal{F}v\mathcal{ML}$ distribution with an arbitrary mean direction, and with three different values of the concentration parameter $\kappa = 1, 2, 5$. Two kernel density estimates over these points are constructed: one with the $\mathcal{F}v\mathcal{ML}$ kernel and one with the $\mathcal{HWC}$ kernel. The MISE eq. (3) is computed over 250 000 points sampled from the uniform distribution on the sphere. The true efficiency is computed as a ratio of the MISE similar to the calculated efficency using the AMISE in eq. (20). The experiment is ran over 168 separate Monte Carlo iterations (36 times the number of cores in the machine used to compute these values), and the reported efficiency is taken to be the mean of these runs. The results, with a plot of the theoretical efficiency from eq. (21) are presented in fig. 3.

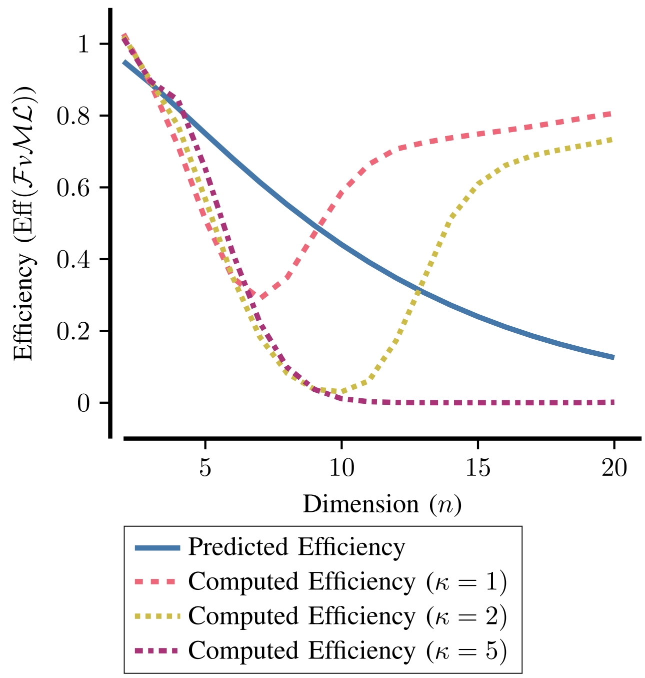

Fig. 3. Efficiency of the Fisher-von Mises-Langevin kernel. The blue line represents the predicted efficiency through eq. (21), while the other lines represent the computed efficiency for $N = 1000$ samples, and various parameters κ.

For all dimensions except $n = 2$, the computed efficiency is smaller than one, meaning that the Hall-Watson-Cabrera kernel requires significantly less samples for the same level of accuracy as compared to the Fisher-von Mises-Langevin kernel. For $\kappa = 5$, past about $n = 10$, the $\mathcal{HWC}$ kernel density estimate effectively makes the $\mathcal{F}v\mathcal{ML}$ estimate irrelevant, as the latter requires more than 100 times the samples than the former. For $\kappa = 1$ and $\kappa = 2$ the result is not as good, but still below one in this regime. We suspect that this result for higher values of dimension is either due to poor calculation of the MISE or due to the estimation of the concentration parameter being highly biased for higher dimensions.

The second numerical experiment will validate the sampling procedure described in section IV-B. We construct a $\mathcal{HWC}$ mixture model as follows: first we take the sample values on the sphere embedded in two dimensions, $\mathbb{S}^1$, namely, $\mu_1 = [\sqrt{2}/2\ -\sqrt{2}/2]^T$ and $\mu_2 = [\sqrt{2}/2\ \sqrt{2}/2]^T$, and assume that we somehow estimated the concentration to be $\kappa = 10$. We then construct a KDE using all the machinery described

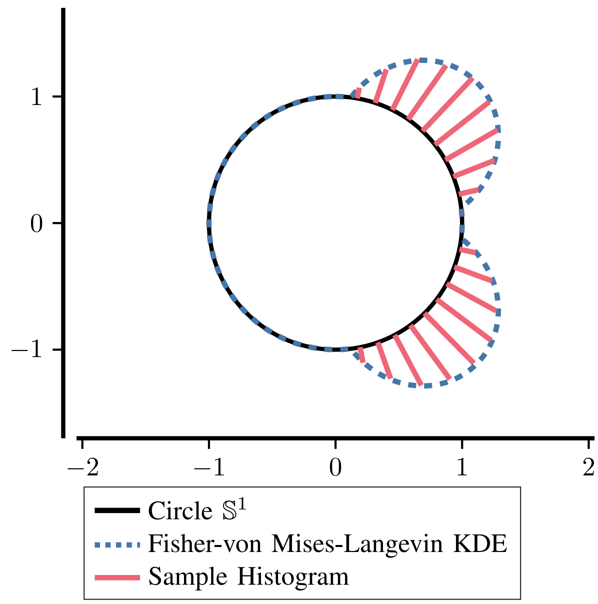

Fig. 4. An example kernel density estimate using the Hall-Watson-Cabrera kernel in two dimensions. The black circle is $\mathbb{S}^1$, the blue dotted plot is the value of the kernel density estimate simulated, and the solid red lines extending from the sphere are representations of the 20-bin histogram over the computed sample points.

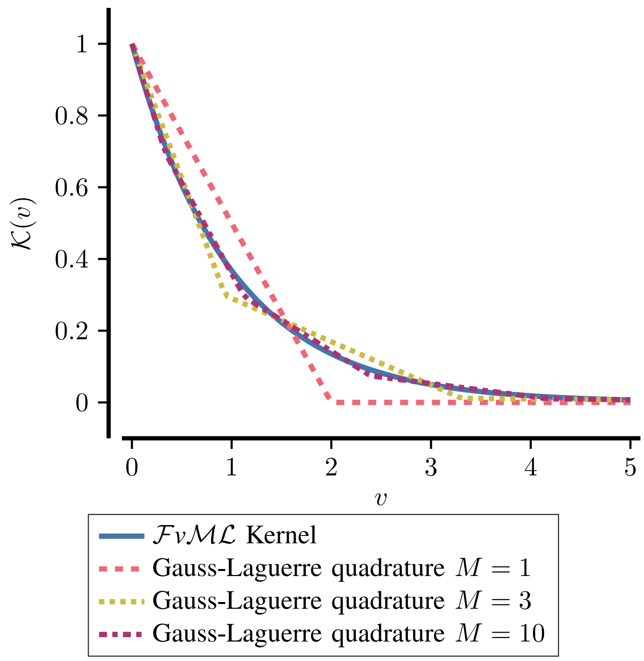

Fig. 5. A representation of the quadrature rule described in eq. (40) for various values of M. The solid blue line represents the $\mathcal{F}v\mathcal{ML}$ kernel, the dashed red line represents its approximation with $M = 1$ quadrature point, the dotted yellow line represents its approximation with $M = 3$ quadrature points, and the dash-dotted purple line represents its approximation by $M = 10$ quadrature points.

herein, meaning that resulting mixture has two $\mathcal{HWC}$ kernels with equal weights and equal bandwidths, centered at the mean directions μ_1 and μ_2. From this KDE mixture model, we sample $N = 10\,000$ points, and compute the histogram over 20 equally spaced bins.

The results of this numerical experiment can be seen in fig. 4. As can be seen, the computed histogram nearly matches the whole distribution over all bins, meaning that the sampling procedure described, and equations derived, are numerically validated.

It is important to note that while the experiment was performed in two dimensions, the derived equations are likely valid for all cases, including the very important $n = 4$ case [30].

The third experiment validates the Gauss-Laguerre quadrature for the $\mathcal{F}v\mathcal{ML}$ kernel in terms of the $\mathcal{HWC}$ kernel derived in section IV-D. We focus on the arbitrary variable v, and derive the Gauss-Laguerre quadrature eq. (40) for $M = 1$, $M = 3$ and $M = 10$.

As can be seen in fig. 5, the quadrature approximations approach the correct function as the number of quadrature points grows. For $M = 1$ the approximation is similar to, but not equivalent to, replacing the $\mathcal{F}v\mathcal{ML}$ kernel by the $\mathcal{HWC}$ kernel with modified bandwidth. For $M = 3$ the dominant behavior of the kernel is approximated well, with significantly less error evident by visual inspection. Finally for $M = 10$ the approximation has very little error, meaning that it is likely that the quadrature rule derived is valid.

It is interesting to note that this quadrature by the Hall-Watson-Cabrera kernel bears a striking resemblance to rectified linear unit [31] universal function approximation [32]. The exploration of which is beyond the scope of this work.

VI. Conclusions

In this work, we rediscover the Hall-Watson-Cabrera kernel for spherical kernel density estimation. We derive the efficiency of the Fisher-von Mises-Langevin kernel in terms of this kernel, how to sample from a $\mathcal{HWC}$ mixture model, and a quadrature rule that expresses the $\mathcal{F}v\mathcal{ML}$ kernel in terms of the $\mathcal{HWC}$ kernel. We then numerically validate all the properties derived. For future work, we will apply the $\mathcal{HWC}$ kernel density estimation techniques to non-linear pose estimation.

Acknowledgment

The authors would like to thank Kyle J. DeMars whose work inspired this line of inquiry.

References

[1] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009.

[2] A. Pewsey and E. García-Portugués, "Recent advances in directional statistics," *Test*, vol. 30, no. 1, pp. 1–58, 2021.

[3] B. W. Silverman, *Density estimation for statistics and data analysis*. Routledge, 2018.

[4] C. Ley and T. Verdebout, *Modern directional statistics*. Chapman and Hall/CRC, 2017.

[5] I. Gilitschenski, G. Kurz, and U. D. Hanebeck, "Non-identity measurement models for orientation estimation based on directional statistics," in *2015 18th International Conference on Information Fusion (Fusion)*. IEEE, 2015, pp. 727–733.

[6] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Unscented von mises–fisher filtering," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 463–467, 2016.

[7] D. Frisch and U. D. Hanebeck, "Deterministic von mises–fisher sampling on the sphere using fibonacci lattices," in *2023 IEEE Symposium Sensor Data Fusion and International Conference on Multisensor Fusion and Integration (SDF-MFI)*. IEEE, 2023, pp. 1–8.

[8] ——, "Deterministic sampling with separation of variables in spherical coordinates," in *2025 28th International Conference on Information Fusion (FUSION)*. IEEE, 2025, pp. 1–8.

[9] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Theory of Probability & Its Applications*, vol. 14, no. 1, pp. 153–158, 1969.

[10] K. Hodges, "Spherical nonparametric estimators applied to the ugamp model integration for amip," *Monthly Weather Review*, vol. 124, no. 12, pp. 2914–2932, 1996.

[11] J. Klemelä, "Estimation of densities and derivatives of densities with directional data," *Journal of Multivariate Analysis*, vol. 73, no. 1, pp. 18–40, 2000.

[12] H. Miao, Y. Jing, Y. Jia, M. Lin, G. Zhang, and G. Wang, "Nonparametric estimations of the sea state bias for a radar altimeter," *Acta Oceanologica Sinica*, vol. 36, no. 9, pp. 108–113, 2017.

[13] P. Hall, G. S. Watson, and J. Cabrera, "Kernel density estimation with spherical data," *Biometrika*, vol. 74, no. 4, pp. 751–762, 1987.

[14] M. S. Bartlett, "Statistical estimation of density functions," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 245–254, 1963.

[15] I. J. Thompson and A. Barnett, "Coulomb and bessel functions of complex arguments and order," *Journal of Computational Physics*, vol. 64, no. 2, pp. 490–509, 1986.

[16] S. Sra, "A short note on parameter approximation for von Mises-Fisher distributions: and a fast implementation of $I_s(x)$," *Computational Statistics*, vol. 27, no. 1, pp. 177–190, 2012.

[17] K. V. Mardia, J. T. Kent, and C. C. Taylor, *Multivariate analysis*. John Wiley & Sons, 2024.

[18] A. A. Popov and R. Zanetti, "The ensemble epanechnikov mixture filter," *arXiv preprint arXiv:2408.11164*, 2024.

[19] G. Ulrich, "Computer generation of distributions on the m-sphere," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 33, no. 2, pp. 158–163, 1984.

[20] A. T. Wood, "Simulation of the von Mises Fisher distribution," *Communications in statistics-simulation and computation*, vol. 23, no. 1, pp. 157–164, 1994.

[21] C. Pinzón and K. Jung, "Fast python sampler for the von Mises Fisher distribution," *hal-04004568v3*, 2023.

[22] J. W. Pearson, "Computation of hypergeometric functions," Ph.D. dissertation, University of Oxford, 2009.

[23] E. M. Zucchelli and B. A. Jones, "A gaussian integral filter with multivariate laplace process noise," in *2023 26th International Conference on Information Fusion (FUSION)*. IEEE, 2023, pp. 1–8.

[24] P. Rabinowitz and G. Weiss, "Tables of abscissas and weights for numerical evaluation of integrals of the form $\int_0^\infty e^{-x} x^n f(x) dx$," *Mathematical Tables and Other Aids to Computation*, pp. 285–294, 1959.

[25] J.-H. He, N. Anjum, C.-H. He, and A. A. Alsolami, "Beyond laplace and fourier transforms: challenges and future prospects," *Thermal Science*, vol. 27, no. 6 Part B, pp. 5075–5089, 2023.

[26] N. Poroshina and V. Ryabov, "Methods for laplace transform inversion," *Vestnik St. Petersburg University: Mathematics*, vol. 44, no. 3, pp. 214–222, 2011.

[27] R. Piessens, "Gaussian quadrature formulas for the numerical integration of bromwich's integral and the inversion of the laplace transform," *Journal of Engineering Mathematics*, vol. 5, no. 1, pp. 1–9, 1971.

[28] A. H. Cheng, P. Sidauruk, and Y. Abousleiman, "Approximate inversion of the laplace transform," *Mathematica Journal*, vol. 4, no. 2, pp. 76–82, 1994.

[29] I. I. Ryabtsev, "Convergent calculation of the original thorugh the values of the transform at equally spaced points on the real axis (in russian)," *Matematika*, no. 3, pp. 139–143, 1966.

[30] M. D. Shuster *et al.*, "A survey of attitude representations," *Navigation*, vol. 8, no. 9, pp. 439–517, 1993.

[31] A. S. Householder, "A theory of steady-state activity in nerve-fiber networks: I. definitions and preliminary lemmas," *The bulletin of mathematical biophysics*, vol. 3, no. 2, pp. 63–69, 1941.

[32] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural networks*, vol. 6, no. 6, pp. 861–867, 1993.