Efficient Evaluation of the Probability Density Function of a Wrapped Normal Distribution

Gerhard Kurz, **Igor Gilitschenski**, and **Uwe D. Hanebeck**
Intelligent Sensor-Actuator-Systems Laboratory (ISAS)
Institute for Anthropomatics and Robotics
Karlsruhe Institute of Technology (KIT), Germany
gerhard.kurz@kit.edu, gilitschenski@kit.edu, uwe.hanebeck@ieee.org

***Abstract*—The wrapped normal distribution arises when the density of a one-dimensional normal distribution is wrapped around the circle infinitely many times. At first look, evaluation of its probability density function appears tedious as an infinite series is involved. In this paper, we investigate the evaluation of two truncated series representations. As one representation performs well for small uncertainties, whereas the other performs well for large uncertainties, we show that in all cases a small number of summands is sufficient to achieve high accuracy.**

I. Introduction

The wrapped normal (WN) distribution is one of the most widely used distributions in circular statistics. It is obtained by wrapping the normal distribution around the unit circle and adding all probability mass wrapped to the same point (see Fig. 1). This is equivalent to defining a normally distributed random variable X and considering the wrapped random variable $X \mod 2\pi$.

The WN distribution has been used in a variety of applications. These applications include nonlinear circular filtering [1], [2], constrained object tracking [3], speech processing [4], [5], and bearings-only tracking [6].

However, evaluation of the WN probability density function can appear difficult because it involves an infinite series. This is one of the main reasons why many authors (such as [7], [8], [9], [10], [11]) use the von Mises distribution instead. It is even sometimes referred to as the *circular normal distribution* [12]. Collet et al. published some results on discriminating between wrapped normal and von Mises distributions [13]. Their results were further refined by [14]. These analyses indicate that several hundred samples are necessary to distinguish between the two distributions. Therefore the von Mises distribution can be considered as a sufficiently good approximation in applications where sample sizes are small, but may prove insufficient in applications with large sample sizes.

In this paper, we will show that a very accurate numerical evaluation of the WN probability density function can be performed with little effort. Some authors have briefly discussed approximation of the WN probability density function, but to our knowledge, no one has published any proof for error bounds, even though the WN distribution has been known and used for a long time [15]. Jammalamadaka and Sengupta simply state [12, Sec. 2.2.6]

> *It is clear that the density can be adequately approximated by just the first few terms of the infinite series, depending on the value of* σ^2.

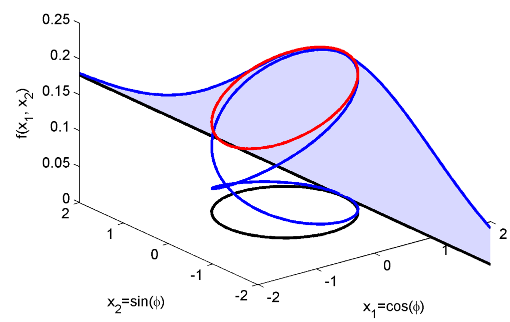

Figure 1. The wrapped normal distribution is obtained by wrapping a normal distribution around the unit circle.

In their book on directional statistics, Mardia and Jupp [16, Sec. 3.5] suggest

> *For practical purposes, the density* ϕ_w *can be approximated adequately by the first three terms of (3.5.66)* [g_n in this paper] *when* $\sigma^2 > 2\pi$ *while for* $\sigma^2 \leq 2\pi$ *the term with* $k =$ *of (3.5.64)* [f_n in this paper] *gives a reasonable approximation.*

While this is practical advice, there is no theoretical justification for using this approximation and there is no quantification of the error incurred by this method. Other classic publications on circular statistics such as the book by Batschelet [17, Sec. 15.4] and the book by Fisher [18, Sec. 3.3.5] do not give any details about evaluation the WN probability density function and suggest the use of the von Mises distribution instead.

II. The Wrapped Normal Distribution

The wrapped normal distribution [12, Sec. 2.2.6], [16, Sec. 3.5] is defined by the probability density function (pdf)

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(x + 2\pi k - \mu)^2}{2\sigma^2}\right) ,$$

with $x \in [0, 2\pi)$, location parameter $\mu \in [0, 2\pi)$, and uncertainty parameter $\sigma > 0$. Because the summands of the series converge to zero, it is natural to approximate the pdf with a truncated series

$$f(x;\mu,\sigma) \approx f_n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\sum_{k=-n}^{n}\exp\left(-\frac{(x+2\pi k-\mu)^2}{2\sigma^2}\right) ,$$

where only $2n+1$ summands are considered. We will investigate the choice of n (depending on σ) in this paper.

As we will later prove, the series representation defined above yields a good approximation for small values of σ only. For this reason, we introduce a second representation, which yields good approximations for large values of σ. The pdf of a WN distribution is closely related to the Jacobi theta function [19]. This leads to another representation of the pdf [12, eq. (2.2.15)]

$$g(x;\mu,\sigma) = \frac{1}{2\pi}\left(1+2\sum_{k=1}^{\infty}\rho^{k^2}\cos(k(x-\mu))\right) ,$$

where $\rho = \exp(-\sigma^2/2)$. Analogous to f_n, we define a truncated version

$$g(x;\mu,\sigma) \approx g_n(x;\mu,\sigma) = \frac{1}{2\pi}\left(1+2\sum_{k=1}^{n}\rho^{k^2}\cos(k(x-\mu))\right)$$

that only considers the first n summands.[1]

III. Empirical Results

We implemented the truncated series f_n and g_n as well as the exact solution (which increases n until the value of the pdf does not change anymore because of the limited accuracy of the data type). We used the IEEE 754 double data type for all variables. It consists of 1 bit for the sign, 11 bit for the exponent, and 52 bit for the fraction [20]. Thus, it is accurate to approximately 15 decimal digits.

For $x,\mu \in [0,2\pi)$, the error is largest for $\mu = 0$ and $x \to 2\pi$ in both approximations (see Fig. 2). We will later show this fact in the theoretical section. Thus, we compare the error $e_f(n,\sigma) = |f(2\pi;0;\sigma) - f_n(2\pi,0,\sigma)|$ and $e_g(n,\sigma) = |g(2\pi;0;\sigma) - g_n(2\pi,0,\sigma)|$ respectively. The results for $n = 1,2,\ldots,11$ are depicted in Fig. 3. Furthermore, we include a comparison to the uniform distribution with pdf $f_u(x) = \frac{1}{2\pi}$, which is also a special case of g_n for $n = 0$. As can be seen, the uniform distribution is accurate up to numerical precision for approximately $\sigma \geq 9$.

We empirically determined the combined approximation based on f_n and g_n for different accuracies (see Table I).

[1] We treat the parameter n in f_n and g_n the same way, although the evaluation of f_n involves $2n+1$ summands whereas the evaluation of g_n only involves n summands. However, the computational effort for evaluation of a single summand of g_n is higher, which roughly negates this difference.

accuracy	range	approximation
1E-5	$0 < \sigma < 1.34$	$f^0(x;\mu,\sigma)$
	$1.34 \leq \sigma < 2.28$	$f^1(x;\mu,\sigma)$
	$2.28 \leq \sigma < 4.56$	$g^1(x;\mu,\sigma)$
	$4.56 \leq \sigma$	$g^0(x;\mu,\sigma)$
1E-10	$0 < \sigma < 0.93$	$f^0(x;\mu,\sigma)$
	$0.93 \leq \sigma < 1.89$	$f^1(x;\mu,\sigma)$
	$1.89 \leq \sigma < 2.21$	$f^2(x;\mu,\sigma)$
	$2.21 \leq \sigma < 3.31$	$g^2(x;\mu,\sigma)$
	$3.31 \leq \sigma < 6.62$	$g^1(x;\mu,\sigma)$
	$6.62 \leq \sigma$	$g^0(x;\mu,\sigma)$
1E-15	$0 < \sigma < 0.76$	$f^0(x;\mu,\sigma)$
	$0.76 \leq \sigma < 1.53$	$f^1(x;\mu,\sigma)$
	$1.53 \leq \sigma < 2.31$	$f^2(x;\mu,\sigma)$
	$2.31 \leq \sigma < 2.73$	$g^3(x;\mu,\sigma)$
	$2.73 \leq \sigma < 4.09$	$g^2(x;\mu,\sigma)$
	$4.09 \leq \sigma < 8.17$	$g^1(x;\mu,\sigma)$
	$8.17 \leq \sigma$	$g^1(x;\mu,\sigma)$

Table I
Combined approximations for different accuracies.

IV. Theoretical Results

Before we analyze the approximation error of the different approaches, we prove an inequality for the error function.

Lemma 1. *For $x > 1$, the error function fulfills the inequality $1 - \mathrm{erf}(x) \leq \frac{e^{-x^2}}{\sqrt{\pi}}$.*

Proof: We use the continued fraction representation as given in [19, 7.1.14]

$$\mathrm{erf}(x) = 1 - \frac{e^{-x^2}}{\sqrt{\pi}\left(x+\frac{1}{2x+\frac{2}{x+\frac{3}{2x+\frac{4}{x+\cdots}}}}\right)}$$

$$\Rightarrow 1 - \mathrm{erf}(x) = \frac{e^{-x^2}}{\sqrt{\pi}\left(x+\frac{1}{2x+\frac{2}{x+\frac{3}{2x+\frac{4}{x+\cdots}}}}\right)}$$

$$\underset{x>1}{\Rightarrow} 1 - \mathrm{erf}(x) \leq \frac{e^{-x^2}}{\sqrt{\pi}}$$

■

A. Representation Based on Wrapped Density

We consider the approximation $f_n(x;\mu,\sigma) \approx f(x;\mu,\sigma)$. In the following proposition, we will show that the error decreases exponentially in n.

Proposition 1. *For $x,\mu \in [0,2\pi)$ and $n > 1 + \frac{\sigma}{\sqrt{2\pi}}$, the error $e_f(n,\sigma) = |f_n(x;\mu,\sigma) - f(x;\mu,\sigma)|$ has an upper bound*

$$e_f(n,\sigma) < \frac{\exp\left(-\frac{(\pi\sqrt{2}(n-1))^2}{\sigma^2}\right)}{2\pi^{3/2}} .$$

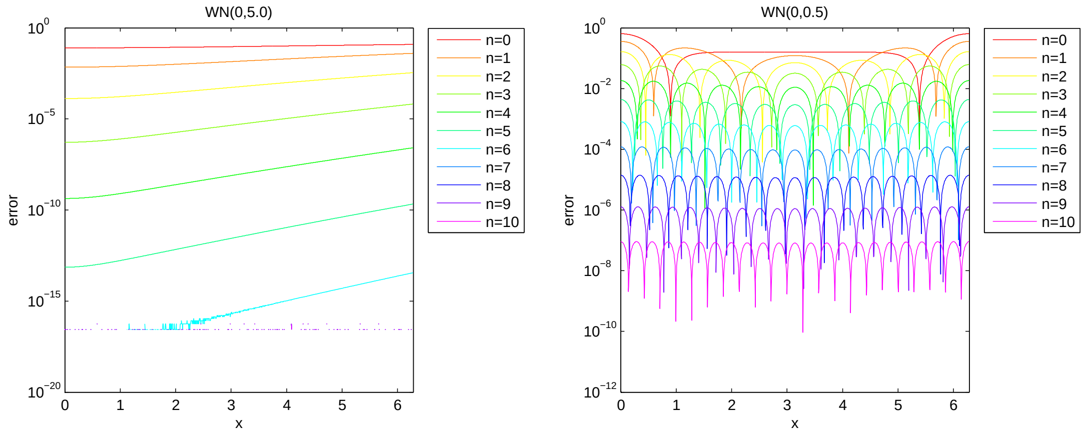

Figure 2. Empirical results depicting the error for different values of n for $e_f(n,\sigma)$ with $\sigma = 5$ (left) and $e_g(n,\sigma)$ with $\sigma = 0.5$ (right). Note that some points are rounded to zero because of the limited accuracy of the floating point arithmetic. These values are not depicted, because it is not possible to display them in a logarithmic plot.

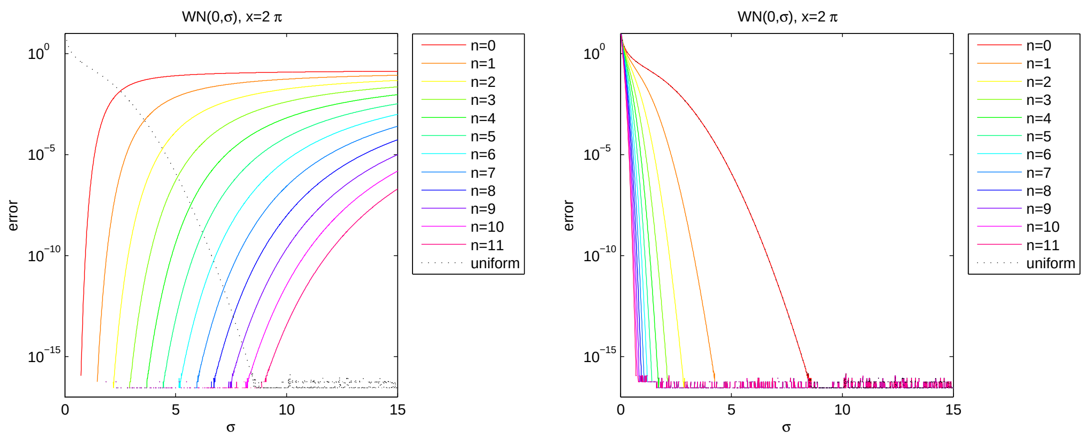

Figure 3. Empirical results depicting the error for different values of n for $e_f(n,\sigma)$ (left) and $e_g(n,\sigma)$ (right). We set the WN parameter $\mu = 0$ and $x = 2\pi$.

Proof: We use the fact that $\sigma > 0$ and $\exp(\cdot) > 0$, and get

$$
\begin{aligned}
e_f(n,\sigma) &= |f_n(x;\mu,\sigma) - f(x;\mu,\sigma)| \\
&= \Bigg| \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-n}^{n} \exp\left(-\frac{(x-\mu-2k\pi)^2}{2\sigma^2}\right) \\
&\quad - \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} \exp\left(-\frac{(x-\mu-2k\pi)^2}{2\sigma^2}\right) \Bigg| \\
&\underset{\sigma>0}{=} \frac{1}{\sqrt{2\pi}\sigma} \Bigg| \sum_{k=-\infty}^{-n-1} \exp\left(-\frac{(x-\mu-2k\pi)^2}{2\sigma^2}\right) \\
&\quad + \sum_{k=n+1}^{\infty} \exp\left(-\frac{(x-\mu-2k\pi)^2}{2\sigma^2}\right) \Bigg| \\
&\underset{\exp(\cdot)>0}{=} \frac{1}{\sqrt{2\pi}\sigma} \Bigg(\sum_{k=-\infty}^{-n-1} \exp\left(-\frac{(x-\mu-2k\pi)^2}{2\sigma^2}\right) \\
&\quad + \sum_{k=n+1}^{\infty} \exp\left(-\frac{(x-\mu-2k\pi)^2}{2\sigma^2}\right) \Bigg) .
\end{aligned}
$$

Now we make use of the fact that μ and x are in the same interval of length 2π

$$
\begin{aligned}
e_f(n,\sigma) &\underset{|x-\mu|<2\pi}{<} \frac{1}{\sqrt{2\pi}\sigma} \Bigg(\sum_{k=-\infty}^{-n-1} \exp\left(-\frac{(-2\pi-2k\pi)^2}{2\sigma^2}\right) \\
&\quad + \sum_{k=n+1}^{\infty} \exp\left(-\frac{(2\pi-2k\pi)^2}{2\sigma^2}\right) \Bigg) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \Bigg(\sum_{k=-\infty}^{-n-1} \exp\left(-\frac{(-2(k+1)\pi)^2}{2\sigma^2}\right) \\
&\quad + \sum_{k=n+1}^{\infty} \exp\left(-\frac{(-2(k-1)\pi)^2}{2\sigma^2}\right) \Bigg) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \Bigg(\sum_{k=-\infty}^{-n} \exp\left(-\frac{(2k\pi)^2}{2\sigma^2}\right) \\
&\quad + \sum_{k=n}^{\infty} \exp\left(-\frac{(2k\pi)^2}{2\sigma^2}\right) \Bigg) =: \hat{e}_f(n,\sigma) .
\end{aligned}
$$

This allows us to simplify the expression by combining the two series into a single series

$$\hat{e}_f(n,\sigma) = \frac{2}{\sqrt{2\pi}\sigma} \sum_{k=n}^{\infty} \exp\left(-\frac{(2k\pi)^2}{2\sigma^2}\right) \ .$$

We find an upper bound by integration

$$\begin{aligned}\hat{e}_f(n,\sigma) &\leq \frac{2}{\sqrt{2\pi}\sigma} \int_{k=n-1}^{\infty} \exp\left(-\frac{(2k\pi)^2}{2\sigma^2}\right) dk \\ &\underset{[19, \text{ eq. } (7.1.2)]}{=} \frac{\left(1-\text{erf}\left(\frac{\pi\sqrt{2}\,(n-1)}{\sigma}\right)\right)}{2\pi} \\ &\underset{\text{Lemma 1}}{\leq} \frac{\exp\left(-\frac{(\pi\sqrt{2}(n-1))^2}{\sigma^2}\right)}{2\pi^{3/2}} \ ,\end{aligned}$$

where we use the assumption $\frac{\pi\sqrt{2}\,(n-1)}{\sigma} > 1$ in order to apply Lemma 1. ■

B. Representation Based on Theta Function

In the following, we consider the approximation $g_n(x;\mu,\sigma) \approx g(x;\mu,\sigma)$. In this case, the error decreases exponentially in n as well.

Proposition 2. *For $x,\mu \in [0,2\pi)$ and $n > \sqrt{2}/\sigma$, the error $e_g(n,\sigma) = |g_n(x;\mu,\sigma) - g(x;\mu,\sigma)|$ has an upper bound*

$$e_g(n,\sigma) < \frac{\exp(-n^2\sigma^2/2)}{\sqrt{2\pi}\sigma} \ .$$

Proof: We start with some simplifications

$$\begin{aligned}&e_g(n,\sigma) \\ &= |g_n(x;\mu,\sigma) - g(x;\mu,\sigma)| \\ &= \left|\frac{1}{2\pi}\left(1 + 2\sum_{k=1}^{n} \rho^{k^2} \cos(k(x-\mu))\right)\right. \\ &\left. -\frac{1}{2\pi}\left(1 + 2\sum_{k=1}^{\infty} \rho^{k^2} \cos(k(x-\mu))\right)\right| \\ &= \frac{1}{\pi}\left|\sum_{k=1}^{n} \rho^{k^2} \cos(k(x-\mu)) - \sum_{k=1}^{\infty} \rho^{k^2} \cos(k(x-\mu))\right| \\ &= \frac{1}{\pi}\left|\sum_{k=n+1}^{\infty} \rho^{k^2} \cos(k(x-\mu))\right| \ ,\end{aligned}$$

use the triangle inequality and the fact that $|\cos(\cdot)| \leq 1$

$$e_g(n,\sigma) \leq \frac{1}{\pi} \sum_{k=n+1}^{\infty} \rho^{k^2} =: \hat{e}_g(n,\sigma) \ .$$

Now we find an upper bound by integration and simplify

$$\begin{aligned}\hat{e}_g(n,\sigma) &\leq \frac{1}{\pi}\int_n^{\infty} \rho^{k^2} dk \\ &\underset{[19, \text{ eq. } (7.1.2)]}{=} \frac{1}{\pi} \cdot \frac{\sqrt{\pi}\ \text{erfc}(n\sqrt{-\log(\rho)})}{2\sqrt{(-\log(\rho))}} \\ &= \frac{1}{\pi} \cdot \frac{\sqrt{\pi}\ \text{erfc}(n\sqrt{\sigma^2/2})}{2\sqrt{\sigma^2/2}} \\ &= \frac{1-\text{erf}(n\sigma/\sqrt{2})}{\sqrt{2\pi}\sigma} \\ &\underset{\text{Lemma 1}}{\leq} \frac{\exp(-n^2\sigma^2/2)}{\sqrt{2\pi}\sigma} \ ,\end{aligned}$$

where we use the assumption $n\sigma/\sqrt{2} > 1$ in order to apply Lemma 1. ■

C. Combination of Both Approaches

For a given error threshold $\tilde{e} > 0$ and a given $\sigma > 0$, we want to obtain the lowest possible n that guarantees that the error threshold is not exceeded. Solving the bound from Proposition 1 for n and taking the precondition for n into account yields

$$n \geq 1 + \frac{\sigma}{\pi}\sqrt{-\log(4\pi^3\tilde{e}^2)} \quad \wedge \quad n > 1 + \frac{\sigma}{\sqrt{2}\pi} \ .$$

By applying the method to the results of Proposition 2, we obtain

$$n \geq \frac{1}{\sigma}\sqrt{-\log(2\pi^2\sigma^2\tilde{e}^2)} \quad \wedge \quad n > \frac{\sqrt{2}}{\sigma} \ .$$

Thus, we define

$$\begin{aligned}n_f &:= \max\left(1 + \frac{\sigma}{\pi}\sqrt{-\log(4\pi^3\tilde{e}^2)}\ ,\ 1 + \frac{\sigma}{\sqrt{2}\pi}\right) \ , \\ n_g &:= \max\left(\frac{1}{\sigma}\sqrt{-\log(2\pi^2\sigma^2\tilde{e}^2)}\ ,\ \frac{\sqrt{2}}{\sigma}\right) \ .\end{aligned}$$

Consequently, we set $n := \lceil \min(n_f, n_g) \rceil$ and choose the according method for approximation. Examples with $\tilde{e} = 1E-5$ and $\tilde{e} = 1E-15$ are given in Fig. 4. Note that the required n is slightly higher than the empirically obtained values given in Table I, because the theoretical bounds are not tight.

V. Conclusion

In this paper, we have shown theoretical bounds on two different representations of the wrapped normal probability density function based on truncated infinite series. In both cases, the error decreases exponentially with increasing number of summands n. Furthermore, we have shown that one representation performs well for small σ whereas the other performs well for large σ. This motivates their combined use depending on the value of σ. Our empirical results match well with the theoretical conclusions. We have proposed piecewise approximations based on the two representations with a varying number of summands for several levels of accuracy.

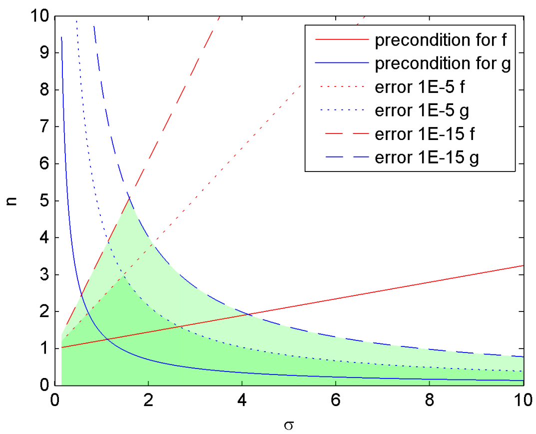

Figure 4. Theoretical results for minimum value of n. We consider $\tilde{e} = 1E-5$ and $\tilde{e} = 1E-15$. The required n by combining both approximations is shaded in dark green and light green respectively.

Acknowledgment

This work was partially supported by grants from the German Research Foundation (DFG) within the Research Training Groups RTG 1194 "Self-organizing Sensor-Actuator-Networks" and RTG 1126 "Soft-tissue Surgery: New Computer-based Methods for the Future Workplace".

References

[1] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive Nonlinear Filtering for Angular Data Based on Circular Distributions," in *Proceedings of the 2013 American Control Conference (ACC 2013)*, Washington D. C., USA, Jun. 2013.

[2] ——, "Nonlinear Measurement Update for Estimation of Angular Systems Based on Circular Distributions," in *Proceedings of the 2014 American Control Conference (ACC 2014)*, Portland, Oregon, USA, Jun. 2014.

[3] G. Kurz, F. Faion, and U. D. Hanebeck, "Constrained Object Tracking on Compact One-dimensional Manifolds Based on Directional Statistics," in *Proceedings of the Fourth IEEE GRSS International Conference on Indoor Positioning and Indoor Navigation (IPIN 2013)*, Montbeliard, France, Oct. 2013.

[4] Y. Agiomyrgiannakis and Y. Stylianou, "Wrapped Gaussian Mixture Models for Modeling and High-Rate Quantization of Phase Data of Speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 775–786, 2009.

[5] J. Traa and P. Smaragdis, "A Wrapped Kalman Filter for Azimuthal Speaker Tracking," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1257–1260, 2013.

[6] I. Gilitschenski, G. Kurz, and U. D. Hanebeck, "Circular Statistics in Bearings-only Sensor Scheduling," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, Istanbul, Turkey, Jul. 2013.

[7] N. I. Fisher, "Problems with the Current Definitions of the Standard Deviation of Wind Direction," *Journal of Climate and Applied Meteorology*, vol. 26, no. 11, pp. 1522–1529, 1987.

[8] M. Azmani, S. Reboul, J.-B. Choquel, and M. Benjelloun, "A Recursive Fusion Filter for Angular Data," in *IEEE International Conference on Robotics and Biomimetics (ROBIO 2009)*, 2009, pp. 882–887.

[9] G. Stienne, S. Reboul, M. Azmani, J. Choquel, and M. Benjelloun, "A Multi-temporal Multi-sensor Circular Fusion Filter," *Information Fusion*, vol. 18, pp. 86–100, Jul. 2013.

[10] K. E. Mokhtari, S. Reboul, M. Azmani, J.-B. Choquel, H. Salaheddine, B. Amami, and M. Benjelloun, "A Circular Interacting Multi-Model Filter Applied to Map Matching," in *Proceedings of the 16th International Conference on Information Fusion (Fusion 2013)*, 2013.

[11] I. Markovic and I. Petrovic, "Bearing-Only Tracking with a Mixture of von Mises Distributions," in *International Conference on Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ*, Oct. 2012, pp. 707–712.

[12] S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics*. World Scientific Pub Co Inc, 2001.

[13] D. Collett and T. Lewis, "Discriminating Between the Von Mises and Wrapped Normal Distributions," *Australian Journal of Statistics*, vol. 23, no. 1, pp. 73–79, 1981.

[14] A. Pewsey and M. C. Jones, "Discrimination between the von Mises and wrapped normal distributions: just how big does the sample size have to be?" *Statistics*, vol. 39, no. 2, pp. 81–89, 2005.

[15] E. Breitenberger, "Analogues of the Normal Distribution on the Circle and the Sphere," *Biometrika*, vol. 50, no. 1/2, pp. 81–88, 1963.

[16] K. V. Mardia and P. E. Jupp, *Directional Statistics*, 1st ed. Wiley, 1999.

[17] E. Batschelet, *Circular Statistics in Biology*, ser. Mathematics in biology. London: Academic Press, 1981.

[18] N. I. Fisher, *Statistical Analysis of Circular Data*. Cambridge University Press, 1995.

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. New York: Dover, 1972.

[20] D. Goldberg, "What Every Computer Scientist Should Know About Floating-point Arithmetic," *ACM Comput. Surv.*, vol. 23, no. 1, pp. 5–48, Mar. 1991.