Trigonometric Moment Matching and Minimization of the Kullback–Leibler Divergence

Gerhard Kurz[1] and Uwe D. Hanebeck[1]

***Abstract*—We show an important property of the von Mises distribution on the unit circle. If we approximate an arbitrary circular distribution using a von Mises distribution, the result obtained by trigonometric moment matching also minimizes the Kullback–Leibler divergence (Theorem 1). This result is a justification for circular filtering algorithms based on trigonometric moment matching as the loss of information is minimized. Furthermore, we show that Theorem 1 does not hold for the wrapped normal distribution.**

***Index Terms*—Directional statistics, von Mises distribution, wrapped normal distribution, circular filtering.**

I. INTRODUCTION

It is known that the Gaussian distribution possesses an interesting property: when we approximate an arbitrary density on $\mathbb{R}^n$ with a Gaussian distribution in such a way that the Kullback–Leibler divergence [9] is minimized, we obtain the same parameters as we would obtain from moment matching of the mean and the covariance. In other words, fitting a Gaussian distribution to an arbitrary distribution by matching its mean and covariance constitutes the optimal approximation in terms of the Kullback–Leibler divergence. This classical result has, for example, been stated in [13, Sec. 2], [14, Appendix A.5]. Some further investigations for the exponential family on $\mathbb{R}^n$ can be found in [8]. Note that this property of the Gaussian distribution is lost when the Gaussian distribution is used to approximate densities on manifolds, e.g., in the context of approaches based on modified Kalman filters [18], [5], [3], because the calculation of power moments does not consider the topology of the underlying manifold.

In this paper, we prove that a very similar property holds for the von Mises distribution, a probability distribution defined on the unit circle, if we consider trigonometric moments instead of power moments. The main result of this paper can be stated as follows. If we approximate an arbitrary density on the unit circle with a von Mises distribution, the distribution obtained from trigonometric moment matching also minimizes the Kullback–Leibler divergence. A similar result was previously found for the related von Mises–Fisher distribution on the three-dimensional unit sphere [4, p. 95].

This result is very important because it serves as a justification for the use of (trigonometric) moment matching when performing von Mises-assumed density filtering on the circle. A number of moment-based filters relying on the von Mises distribution have been proposed [2], [19], [11], [7], and our new results indicate that these filters are not only optimal in a moment-sense, but also minimize the Kullback–Leibler divergence, i.e., the information lost as a result of the approximation of the true density with a von Mises density (similar to minimum divergence filtering [6]). These filters can be applied to a variety of relevant applications in the fields of aerospace, robotics, and signal processing. For example, they can be used to estimate the heading of an aircraft, the angle of a robotic joint, and the phase of a received signal.

It should be noted that there are other justifications for the use of trigonometric moment matching. For example, it has been shown that the point estimate that is obtained by considering the complex argument of the first circular moment $m_1^{(p)}$ of a circular distribution $p(\cdot)$ has the following property. The point estimate $x = \operatorname{atan2}(\operatorname{Im}(m_1^{(p)}), \operatorname{Re}(m_1^{(p)}))$ minimizes the expected error given by the error function $d(x, y) = 1 - \cos(x - y)$, where y is a sample drawn from the circular distribution $p(\cdot)$ (see [21, Sec. 2], [22, Example 1]). Further research into optimal circular estimation based on different error criteria (which necessitate the consideration of higher moments) has been presented in [15], [16].

II. KLD PROPERTY OF THE VON MISES DISTRIBUTION

The probability density function of a von Mises distribution [20] is given by

$$g(x; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(x - \mu)) \ ,$$

where $x, \mu \in [0, 2\pi)$, $\kappa \geq 0$, and I_0 is the modified Bessel function of the first kind and of order zero. Its first trigonometric moment is given by the complex number

$$\begin{aligned} m_1^{(g)} &= \int_0^{2\pi} g(x; \mu, \kappa) \exp(ix) dx \\ &= \int_0^{2\pi} g(x; \mu, \kappa) \cos(x) dx + i \int_0^{2\pi} g(x; \mu, \kappa) \sin(x) dx \\ &= \frac{I_1(\kappa)}{I_0(\kappa)} \exp(i\mu) \in \mathbb{C} \end{aligned}$$

according to [12, eq. (3.5.29)]. For a given first trigonometric moment $|m_1^{(g)}| \neq 0$, the parameters of a von Mises distribution are obtained as

$$\mu = \operatorname{atan2}(\operatorname{Im}(m_1^{(g)}), \operatorname{Re}(m_1^{(g)})), \quad \kappa = A_1^{-1}(|m_1^{(g)}|)$$

where $A_1(\kappa) = I_1(\kappa)/I_0(\kappa)$. If $|m_1^{(g)}| = 0$, we have $\kappa = 0$ and μ is undefined, i.e., the resulting distribution is uniform. Based on these fundamentals, we can formulate and prove the main theorem of this paper.

[1] The authors are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany. E-mail: `gerhard.kurz@kit.edu`, `uwe.hanebeck@ieee.org`

Theorem 1. *Consider an arbitrary probability density $p(x)$ on the unit circle that is nowhere zero, i.e., $p(x) > 0\, \forall x \in [0, 2\pi)$, and a von Mises distribution $q(x; \mu, \kappa)$ with parameters μ and κ. Then*

$$[\mu, \kappa] = \arg\min_{\mu,\kappa} \mathrm{D_{KL}}(p||q(x; \mu, \kappa))$$

yields the same result as matching the first circular moment $m_1^{(p)}$ of $p(x)$.

Proof. First, we observe that $q(x; \mu, \kappa) > 0$ for all x, μ, κ, i.e., the KLD is well-defined. Thus, we have

$$\begin{aligned}
\mathrm{D_{KL}}(p||q) &= \int_0^{2\pi} p(x) \log\left(\frac{p(x)}{q(x;\mu,\kappa)}\right) dx \\
&= \int_0^{2\pi} p(x)\log(p(x))dx - \int_0^{2\pi} p(x)\log(q(x;\mu,\kappa))dx \\
&= \int_0^{2\pi} p(x)\log(p(x))dx \\
&\quad - \int_0^{2\pi} p(x)\log\left(\frac{1}{2\pi I_0(\kappa)}\exp(\kappa\cos(x-\mu))\right)dx \\
&= \int_0^{2\pi} p(x)\log(p(x))dx + \int_0^{2\pi} p(x)\log(2\pi I_0(\kappa))dx \\
&\quad - \int_0^{2\pi} p(x)\kappa\cos(x-\mu)dx \\
&= \int_0^{2\pi} p(x)\log(p(x))dx + \log(2\pi I_0(\kappa)) \\
&\quad - \kappa \int_0^{2\pi} p(x)\cos(x-\mu)dx \ .
\end{aligned}$$

a) Location parameter: Now, we take the derivative with respect to μ and set to zero. Interchanging differentiation and integration is allowed according to Lebesgue's dominated convergence theorem as long as $p(x)$ is integrable. This yields

$$\begin{aligned}
&\frac{\partial}{\partial\mu}\mathrm{D_{KL}}(p||q) = -\kappa\int_0^{2\pi} p(x)\frac{\partial}{\partial\mu}\cos(x-\mu)dx \\
&= -\kappa\int_0^{2\pi} p(x)\sin(x-\mu)dx \\
&= -\kappa\int_0^{2\pi} p(x)\big(\sin(x)\cos(\mu) - \cos(x)\sin(\mu)\big)dx \\
&= -\kappa\int_0^{2\pi} p(x)\sin(x)\cos(\mu)dx \\
&\quad + \kappa\int_0^{2\pi} p(x)\cos(x)\sin(\mu)dx \stackrel{!}{=} 0 \\
&\Leftrightarrow \cos(\mu)\int_0^{2\pi} p(x)\sin(x)dx \stackrel{!}{=} \sin(\mu)\int_0^{2\pi} p(x)\cos(x)dx \\
&\Leftrightarrow \cos(\mu)\,\mathrm{Im}(m_1^{(p)}) \stackrel{!}{=} \sin(\mu)\,\mathrm{Re}(m_1^{(p)}) \ .
\end{aligned}$$

Consequently, we obtain a unique solution, which is given by $\mu = \mathrm{atan2}(\mathrm{Im}(m_1^{(p)}), \mathrm{Re}(m_1^{(p)}))$ for $|m_1^{(p)}| \neq 0$ and undefined μ otherwise. Considering the second derivative

$$\begin{aligned}
&\frac{\partial^2}{(\partial\mu)^2}\mathrm{D_{KL}}(p||q)\Big|_{\mu=\mathrm{atan2}(\mathrm{Im}(m_1^{(p)}),\mathrm{Re}(m_1^{(p)}))} \\
&= \frac{\partial}{(\partial\mu)}\Bigg(-\kappa\int_0^{2\pi} p(x)\sin(x)\cos(\mu)dx \\
&\quad + \kappa\int_0^{2\pi} p(x)\cos(x)\sin(\mu)dx\Bigg)\Big|_{\mu=\mathrm{atan2}(\mathrm{Im}(m_1^{(p)}),\mathrm{Re}(m_1^{(p)}))} \\
&= \Bigg(\kappa\sin(\mu)\int_0^{2\pi} p(x)\sin(x)dx \\
&\quad + \cos(\mu)\kappa\int_0^{2\pi} p(x)\cos(x)dx\Bigg)\Big|_{\mu=\mathrm{atan2}(\mathrm{Im}(m_1^{(p)}),\mathrm{Re}(m_1^{(p)}))} \\
&= \kappa\sin(\mathrm{atan2}(\mathrm{Im}(m_1^{(p)}),\mathrm{Re}(m_1^{(p)})))\int_0^{2\pi} p(x)\sin(x)dx \\
&\quad + \kappa\cos(\mathrm{atan2}(\mathrm{Im}(m_1^{(p)}),\mathrm{Re}(m_1^{(p)})))\int_0^{2\pi} p(x)\cos(x)dx \\
&= \frac{\kappa}{|m_1^{(p)}|}\left(\left(\mathrm{Im}(m_1^{(p)})\right)^2 + \left(\mathrm{Re}(m_1^{(p)})\right)^2\right) > 0
\end{aligned}$$

proves that this is a minimum (as $\kappa > 0$ for $|m_1^{(p)}| \neq 0$, see below).

b) Concentration parameter: Now, we take the derivative with respect to κ using the identity $\frac{\partial}{\partial\kappa}I_0(\kappa) = I_1(\kappa)$ and set to zero

$$\begin{aligned}
&\frac{\partial}{\partial\kappa}\mathrm{D_{KL}}(p||q) = \frac{I_1(\kappa)}{I_0(\kappa)} - \int_0^{2\pi} p(x)\cos(x-\mu)dx \stackrel{!}{=} 0 \\
&\Leftrightarrow \frac{I_1(\kappa)}{I_0(\kappa)} \stackrel{!}{=} \int_0^{2\pi} p(x)\cos(x-\mu)dx = |m_1^{(p)}| \qquad (1)
\end{aligned}$$

A proof of the identity (1) in is given the appendix. This yields $\kappa = A_1^{-1}(|m_1^{(p)}|)$ where $A_1(t) = I_1(t)/I_0(t)$. Because the second derivative fulfills

$$\frac{\partial^2}{(\partial\kappa)^2}\mathrm{D_{KL}}(p||q) = \frac{\partial}{\partial\kappa}\frac{I_1(\kappa)}{I_0(\kappa)} = A_1'(\kappa) > 0$$

for arbitrary κ as shown by [1, eq. (15)], this value of κ constitutes a minimum. □

III. Wrapped Normal Distribution

Another common distribution on the unit circle is the wrapped normal distribution [17] with density

$$f(x; \mu, \sigma) = \sum_{k=-\infty}^{\infty} \mathcal{N}(x + 2k\pi, \mu, \sigma) \ ,$$

where $x, \mu \in [0, 2\pi)$ and $\sigma > 0$. Using a numerical counterexample, we can show that the property given in Theorem 1 for the von Mises distribution does not hold for wrapped normal distributions. Consider the distribution given by the piecewise constant probability density function

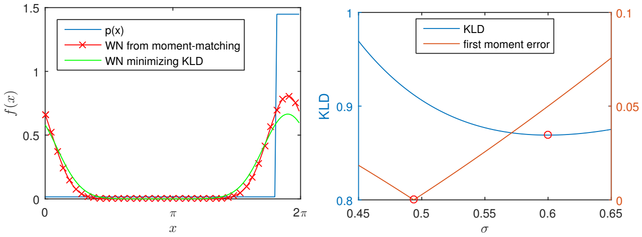
$$p(x) = \begin{cases} \frac{0.1}{2\pi} \ , & x \in [0, \frac{9}{5}\pi) \\ \frac{9.1}{2\pi} \ , & x \in [\frac{9}{5}\pi, 2\pi) \end{cases} \ .$$

Approximating this distribution by trigonometric moment matching (see [11, Lemma 2], [10, Sec. III-A2]) yields

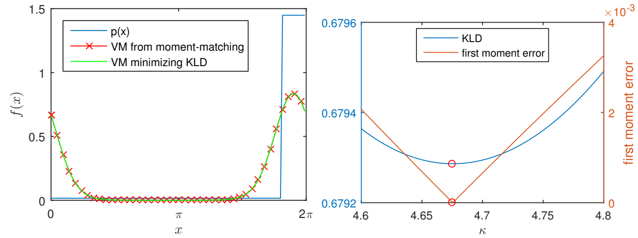
$$\begin{aligned}
\mu_1 &= \mathrm{atan2}(\mathrm{Im}\, m_1^{(p)}, \mathrm{Re}\, m_1^{(p)}) &&= 5.969026, \\
\sigma_1 &= \sqrt{-2\log(|m_1^{(p)}|)} &&= 0.493689,
\end{aligned}$$

whereas approximating by minimizing the Kullback–Leibler divergence using a numerical optimization procedure results in $\mu_2 = 5.969026, \sigma_2 = 0.599728$. Indeed, we find

$$\mathrm{D_{KL}}(p||f(x; \mu_1, \sigma_1)) = 0.9121382830 \ ,$$



(a) WN distribution.



(b) VM distribution.

Fig. 1: Counterexample for the wrapped normal distribution. On the left, we show the true density $p(x)$ and both wrapped normal approximations, whereas on the right, we show the KLD and the absolute error of the first moment as a function of σ. It can be seen that the result from moment-matching does not minimize the KLD. If we perform the same experiment with a von Mises distribution, both methods yields the same result as proven in this paper.

$$\mathrm{D_{KL}}(p||f(x;\mu_2,\sigma_2)) = 0.8690821170 ,$$

i.e., the moment-based solution has a slightly, but unambiguously higher Kullback–Leibler divergence. These results are visualized in Fig. 1. If we perform the same steps with a VM distribution, we obtain $\mu_1 = \mu_2 = 5.969026$, $\kappa_1 = \kappa_2 = 4.675421$ and a KLD of 0.6792864525 in both cases.

IV. Conclusion

We have shown that the von Mises distribution obtained by trigonometric moment matching is also optimal with respect to the Kullback–Leibler divergence. This is an important result, because it serves as a justification for moment-based filters assuming a von Mises distribution. As we have shown, these filters not only retain the trigonometric moments, but also minimize the information that is lost when approximating the true density with a von Mises density. Furthermore, we have shown that the same property does not hold for the wrapped normal density.

The results presented in this paper have significant implications in a variety of applications where circular filters based on trigonometric moment-matching can be applied. In particular, a lot of common aerospace problems such as estimation of the heading or, more general, the orientation of an aircraft may benefit from this result.

Appendix

Here, we show the identity used in (1). First we observe that

$$\cos(\mu) = \frac{\mathrm{Re}(m_1^{(p)})}{|m_1^{(p)}|} , \quad \sin(\mu) = \frac{\mathrm{Im}(m_1^{(p)})}{|m_1^{(p)}|}$$

because $\mu = \mathrm{atan2}(\mathrm{Im}(m_1^{(p)}), \mathrm{Re}(m_1^{(p)}))$. Using the cosine addition formula, we obtain

$$\begin{aligned}
&\int_0^{2\pi} p(x)\cos(x-\mu)dx \\
&= \int_0^{2\pi} p(x)(\cos(x)\cos(\mu) + \sin(x)\sin(\mu))dx \\
&= \cos(\mu)\int_0^{2\pi} p(x)\cos(x)dx + \sin(\mu)\int_0^{2\pi} p(x)\sin(x)dx \\
&= \cos(\mu)\,\mathrm{Re}(m_1^{(p)}) + \sin(\mu)\,\mathrm{Im}(m_1^{(p)}) \\
&= \frac{\mathrm{Re}(m_1^{(p)})}{|m_1^{(p)}|}\mathrm{Re}(m_1^{(p)}) + \frac{\mathrm{Im}(m_1^{(p)})}{|m_1^{(p)}|}\mathrm{Im}(m_1^{(p)}) \\
&= \frac{\mathrm{Re}(m_1^{(p)})^2 + \mathrm{Im}(m_1^{(p)})^2}{|m_1^{(p)}|} = \frac{|m_1^{(p)}|^2}{|m_1^{(p)}|} = |m_1^{(p)}| .
\end{aligned}$$

References

[1] Amos, D. E. Computation of modified bessel functions and their ratios. *Mathematics of Computation 28*, 125 (1974), 239–251.

[2] Azmani, M., Reboul, S., Choquel, J.-B., and Benjelloun, M. A recursive fusion filter for angular data. In *IEEE International Conference on Robotics and Biomimetics (ROBIO 2009)* (2009), pp. 882–887.

[3] Bry, A., Bachrach, A., and Roy, N. State estimation for aggressive flight in gps-denied environments using onboard sensing. In *Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA 2012)* (St. Paul, Minnesota, USA, May 2012).

[4] Chiuso, A., and Picci, G. Visual tracking of points as estimation on the unit sphere. In *The Confluence of Vision and Control*. Springer, 1998, pp. 90–105.

[5] Choukroun, D., Bar-Itzhack, I. Y., and Oshman, Y. Novel quaternion kalman filter. *IEEE Transactions on Aerospace and Electronic Systems 42*, 1 (2006), 174–190.

[6] Darling, J. E., and DeMars, K. J. Minimization of the kullback–leibler divergence for nonlinear estimation. In *Proceedings of the Astrodynamics Specialist Conference* (Vail, Colorado, USA, Aug. 2015).

[7] Gilitschenski, I., Kurz, G., and Hanebeck, U. D. Non-Identity Measurement Models for Orientation Estimation Based on Directional Statistics. In *Proceedings of the 18th International Conference on Information Fusion (Fusion 2015)* (Washington D. C., USA, July 2015).

[8] Herbrich, R. Minimising the kullback–leibler divergence. Tech. rep., Microsoft, Aug. 2005.

[9] Kullback, S. *Information Theory and Statistics*. Dover, 1978.

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

[11] Kurz, G., Gilitschenski, I., and Hanebeck, U. D. Recursive Bayesian Filtering in Circular State Spaces. *arXiv preprint: Systems and Control (cs.SY)* (Jan. 2015).

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

[13] Minka, T. P. Expectation propagation for approximate bayesian inference. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* (2001), Morgan Kaufmann Publishers Inc., pp. 362–369.

[14] Rasmussen, C. E., and Williams, C. K. I. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[15] Routtenberg, T., and Tabrikian, J. Optimal bayesian parameter estimation with periodic criteria. In *2010 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)* (Oct. 2010), pp. 53–56.

[16] Routtenberg, T., and Tabrikian, J. Bayesian parameter estimation using periodic cost functions. *IEEE Transactions on Signal Processing 60*, 3 (Mar. 2012), 1229–1240.

[17] SCHMIDT, W. Statistische methoden beim gefügestudium krystalliner schiefer. *Sitzungsberichte Akademie der Wissenschaften in Wien 126* (July 1917), 515.

[18] SMITH, G. L., SCHMIDT, S. F., AND MCGEE, L. A. *Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle*. National Aeronautics and Space Administration, 1962.

[19] STIENNE, G., REBOUL, S., AZMANI, M., CHOQUEL, J., AND BENJELLOUN, M. A multi-temporal multi-sensor circular fusion filter. *Information Fusion 18* (July 2013), 86–100.

[20] VON MISES, R. Über die "ganzzahligkeit" der atomgewichte und verwandte fragen. *Physikalische Zeitschrift XIX* (1918), 490–500.

[21] WILLSKY, A. S. Fourier series and estimation on the circle with applications to synchronous communication—part i: Analysis. *IEEE Transactions on Information Theory 20*, 5 (1974), 584–590.

[22] WILLSKY, A. S., AND LO, J. T. Estimation for rotational processes with one degree of freedom—part ii: Discrete-time processes. *IEEE Transactions on Automatic Control 20*, 1 (Feb. 1975), 22–30.

Biographies

Gerhard Kurz

Gerhard Kurz received his diploma in computer science from the Karlsruhe Institute of Technology (KIT), Germany, in 2012. He obtained his Ph.D. in 2015 at the Intelligent Sensor-Actuator-Systems Laboratory, Karlsruhe Institute of Technology (KIT), Germany. His research interests include directional filtering, nonlinear estimation, and medical data fusion. He has authored multiple award-winning publications on these topics.

Uwe D. Hanebeck

Uwe D. Hanebeck is a chaired professor of Computer Science at the Karlsruhe Institute of Technology (KIT) in Germany and director of the Intelligent Sensor-Actuator-Systems Laboratory (ISAS). Since 2005, he is the chairman of the Research Training Group RTG 1194 "Self-Organizing Sensor-Actuator-Networks" financed by the German Research Foundation.

Prof. Hanebeck obtained his Ph.D. degree in 1997 and his habilitation degree in 2003, both in Electrical Engineering from the Technical University in Munich, Germany. His research interests are in the areas of information fusion, nonlinear state estimation, stochastic modeling, system identification, and control with a strong emphasis on theory-driven approaches based on stochastic system theory and uncertainty models. Research results are applied to various application topics like localization, human-robot-interaction, assistive systems, sensor-actuator-networks, medical engineering, distributed measuring system, and extended range telepresence. Research is pursued in many academic projects and in a variety of cooperations with industrial partners.

Uwe D. Hanebeck was the General Chair of the "2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)", Program Co-Chair of the "11th International Conference on Information Fusion (Fusion 2008)", Program Co-Chair of the "2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2008)", Regional Program Co-Chair for Europe for the "2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)", and will be General Chair of the "19th International Conference on Information Fusion (Fusion 2016)". He is a Member of the Board of Directors of the International Society of Information Fusion (ISIF), Editor-in-chief of its Journal of Advances in Information Fusion (JAIF), and associate editor for the letter category of the IEEE Transactions on Aerospace and Electronic Systems (TAES). He is author and coauthor of more than 300 publications in various high-ranking journals and conferences.