HHS Public Access

Author manuscript

IEEE Trans Signal Process. Author manuscript; available in PMC 2023 January 14.

Published in final edited form as:

IEEE Trans Signal Process. 2022 ; 70: 686–700. doi:10.1109/tsp.2022.3143471.

Projection Filtering with Observed State Increments with Applications in Continuous-Time Circular Filtering

Anna Kutschireiter[*],

Luke Rast,

Jan Drugowitsch

Department of Neurobiology, Harvard Medical School, Boston MA, United States

Abstract

Angular path integration is the ability of a system to estimate its own heading direction from potentially noisy angular velocity (or increment) observations. Non-probabilistic algorithms for angular path integration, which rely on a summation of these noisy increments, do not appropriately take into account the reliability of such observations, which is essential for appropriately weighing one's current heading direction estimate against incoming information. In a probabilistic setting, angular path integration can be formulated as a continuous-time nonlinear filtering problem (circular filtering) with observed state increments. The circular symmetry of heading direction makes this inference task inherently nonlinear, thereby precluding the use of popular inference algorithms such as Kalman filters, rendering the problem analytically inaccessible. Here, we derive an approximate solution to circular continuous-time filtering, which integrates state increment observations while maintaining a fixed representation through both state propagation and observational updates. Specifically, we extend the established projection-filtering method to account for observed state increments and apply this framework to the circular filtering problem. We further propose a generative model for continuous-time angular-valued direct observations of the hidden state, which we integrate seamlessly into the projection filter. Applying the resulting scheme to a model of probabilistic angular path integration, we derive an algorithm for circular filtering, which we term the circular Kalman filter. Importantly, this algorithm is analytically accessible, interpretable, and outperforms an alternative filter based on a Gaussian approximation.

Keywords

Bayesian methods; nonlinear filtering; circular filtering; sensor fusion; continuous-time estimation; stochastic processes

I. Introduction

A compass is an immensely useful tool for a traveler trying to find their way in a barren and featureless landscape. Absent such a tool, the traveler must employ dead reckoning, using

[*]Corresponding author. anna_kutschireiter@hms.harvard.edu.

what they roughly know about how often and how much they have turned, and summing up those turns to maintain an internal sense of their spatial orientation. Angular path integration, i.e., estimation of heading direction or orientation based on angular self-motion cues, plays an essential role in spatial navigation of humans, other animals and robots [1, 2]. Imperfect sensors make angular path integration an inherently noisy process, and inevitably lead to an accumulation of error in the heading estimate over time. Other cues, such as those from visual landmarks, can help correct the estimate's error, despite being also noisy and ambiguous. Importantly, properly combining path integration with these external cues requires a reliability-weighted update of the orientation estimate. Computing with uncertainties in such a strategic way is a hallmark of dynamic Bayesian inference, and calls for a probabilistic description. Our goal in this work is to derive a dynamic probabilistic algorithm for angular path integration.

It is well known that many organisms are able to maintain an internal compass which they update by self-motion cues. Since the discovery of orientation-selective head-direction cells in rats [3], and, more recently, the heading direction circuit in Drosophila [4], theoretical efforts to unravel the mechanism of angular path integration in the brain have highlighted the role of angular velocity observations of self-initiated turns [5, 6], e.g., from proprioceptive or vestibular feedback, or from visual flow. Current theories suggest that these biological systems implement angular path integration by neural network motifs called ring attractors [7, 6]. Such networks maintain a heading direction estimate through sustained neural activity, but lack the ability to simultaneously represent the estimate's certainty. However, the question of whether such biological systems indeed only operate with single point estimates or instead perform probabilistic inference has been hampered by the lack of a *probabilistic* algorithm for path integration in the brain. The reason for this is the complex set of conditions that are required of such an algorithm: (i) the state-space is circular, (ii) the path integration must operate with a continuous time stream of inputs, (iii) it must maintain a fixed representation of the underlying probability distribution, in line with the expectation that, within a certain area of computation, the brain maintains a similarly fixed representation e.g., in terms of a parametric [8] or sampling-based [9] representation, and (iv) in addition to angular velocity observations, (noisy) direct angular observations (e.g., visual landmarks) may be present, which need to be integrated accordingly.

One approach to designing an algorithm that satisfies these conditions is to consider it in the broader context of continuous-time filtering, which aims to continuously update a posterior distribution over a dynamically evolving hidden state variable from noisy observational data. From condition (i), a circular state-space implies that the underlying filtering task is nonlinear, which precludes the use of popular linear schemes such as the Kalman filter [10, 11]. Furthermore, the solution to this so-called circular filtering problem is analytically intractable [12], and needs to be approximated. The approximation we derive here goes beyond existing circular filtering algorithms [13, 14, 15] (see also the review by Kurz et al. [12]) by both considering increment observations — like observations of angular velocity — and by addressing a continuous stream of observations (condition (ii)). Furthermore, while these previous approaches changed the representations used between prediction and update steps, ours uses a fixed representation, satisfying our condition (iii).

A recent promising approach to circular filtering [16] supports continuous-time state transitions, but is limited to discrete-time and direct (rather than increment) observations. Their method is based on projection filtering [17, 18], a rigorous approach that combines nonlinear filtering with information geometry. The idea behind projection filtering is to approximate the posterior between consecutive discrete-time observations with a parametric distribution. This approximation is chosen to minimize the distance between the true and the approximated posterior, as measured by the Fisher metric. By updating only the values of the parameters, this approach automatically keeps a fixed representation for the posterior in terms of a parametric distribution. Furthermore, if the approximated posterior and the emission probabilities of the observations are conjugate, updating posterior parameters in light of further discrete-time direct observations is straightforward. Projection filtering can be generalized to hidden processes that evolve on arbitrary submanifolds of Euclidean space [19], which makes it applicable to the circular filtering problem. Two challenges hamper the direct application of projection filters to angular path integration. First, no variants currently exist that handle increment observations, either in discrete or continuous time, as we would require to process angular velocity observations. In fact, increment observations have generally received little attention in the filtering literature (but see [20]). Second, there is currently no framework that combines projection-filtering with angular-valued continuous-time observations. We will address both challenges in this work.

We introduce a novel continuous-time nonlinear filtering algorithm based on projection filtering, that includes increment observations and can be applied to circular filtering and to angular path integration, and that meets conditions (i)-(iv) outlined above. To do so, we first describe the general nonlinear filtering problem with observed increment observations in Euclidean space in Section II. In Section III, we review the projection filtering framework [17, 18] as an approximate solution for nonlinear filtering, and extend this approach to account for increment observations. We demonstrate that applying this framework to a *linear* filtering problem recovers the generalized Kalman filter. In Section IV, we revisit the continuous-time circular filtering problem. Therein, we first derive a probabilistic algorithm for angular path integration, i.e., when only increment observations are present. We then account for direct angular-valued observations, in addition to increment observations, by proposing a generative model based on a constant information-rate criterion that supports seamless inclusion into the filtering algorithm. Combining all of the above, we finally retrieve, as a special case of the general framework, a circular filtering algorithm for Gaussian-type increment and angular direct observations, which we term the circular Kalman filter. We demonstrate in numerical simulations that this algorithm performs comparably to an asymptotically-exact particle filter, and outperforms a Gaussian approximation in the estimation of both heading direction as well as its associated certainty.

II. The filtering problem with increment observations

We consider multivariate filtering with observations generated by *increments* of the hidden state, rather than the hidden state itself. We assume that the hidden state variable $\mathbf{X}_t \in \mathbb{R}^N$ evolves according to a stochastic differential equation (SDE) of the form:

$$d\mathbf{X}_t = \mathrm{f}(\mathbf{X}_t, t)dt + \Sigma_x^{1/2} d\mathbf{W}_t, \tag{1}$$

with $\mathbf{W}_t$ an $\mathbb{R}^N$-Brownian motion (BM) process, a vector-valued drift function $\mathbf{f}\colon \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N$, and a matrix $\sum_x^{1/2} \in \mathbb{R}^{N \times N}$, which determines the error covariance of the hidden state process. In the following, we will use the shorthand $\mathbf{f}_t(\mathbf{x}) \colon = \mathbf{f}(\mathbf{x}, t)$ or skip its argument completely (whenever there is no notational ambiguity). The density $\tilde{p}_t(\mathbf{x}) \colon = p(\mathbf{X}_t = \mathbf{x})$ of this stochastic process evolves according to a partial differential equation, the Fokker–Planck equation (FPE):

$$d\tilde{p}_t = \mathscr{L}^\dagger[\tilde{p}_t]dt, \tag{2}$$

with

$$\begin{aligned} \mathscr{L}^\dagger[\tilde{p}_t(x)] = & -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}(f_i(\mathbf{x}, t)\tilde{p}_t(\mathbf{x})) \\ & + \frac{1}{2}\sum_{i,j}^{N} (\Sigma_x)_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \tilde{p}_t(\mathbf{x}) \,. \end{aligned} \tag{3}$$

Equivalently, expectations of a scalar test function $\phi_t \colon = \phi(\mathbf{X}_t)$ with respect to $\tilde{p}_t(\mathbf{x})$ evolve according to

$$d\mathbb{E}[\phi_t] = \mathbb{E}[\mathscr{L}[\phi_t]]dt, \tag{4}$$

with the propagator

$$\mathscr{L}[\phi_t(\mathbf{x}, t)] = \mathrm{f}_t(\mathbf{x})^T \nabla \phi_t(\mathbf{x}) + \frac{1}{2}\mathrm{Tr}\big(H_\phi \Sigma_\mathbf{x}\big), \tag{5}$$

where ∇ denotes the gradient with respect to $\mathbf{x}$, $\mathrm{Tr}(\cdot)$ denotes the trace operator and H_ϕ is the Hessian matrix with $\big(H_\phi\big)_{ij} = \frac{\partial^2 \phi_t}{\partial x_i \partial x_j}$. Note that $\mathscr{L}$ and $\mathscr{L}^\dagger$ are adjoint operators with respect to the L^2 inner product.

We assume that the hidden state process $\mathbf{X}_t$ in (1) cannot be observed directly, but instead is partially observed through the process $d\mathbf{U}_t$, which is governed by the infinitesimal state *increments* $d\mathbf{X}_t$:

$$\begin{aligned} d\mathbf{U}_t &= C\, d\mathbf{X}_t + \Sigma_u^{1/2} d\mathbf{V}_t \\ &= C\, \mathbf{f}_t(\mathbf{X}_t)dt + C\, \Sigma_x^{1/2} d\mathbf{W}_t + \Sigma_u^{1/2} d\mathbf{V}_t \,. \end{aligned} \tag{6}$$

where $\mathbf{V}_t$ is an $\mathbb{R}^M$-BM process, and $C \in \mathbb{R}^{M \times N}$. The matrix $\Sigma_u \in \mathbb{R}^{M \times M}$ determines the level of noise in the increment observations. Due to its dependency on the increment $d\mathbf{X}$, the observation process $d\mathbf{U}_t$ is effectively governed by two noise sources, $d\mathbf{W}_t$ and $d\mathbf{V}_t$. The first

is correlated with the noise in the hidden state dynamics. The second is independent of it. This is in contrast to classical filtering problems, which consider the noise in hidden states and observations to be independent (cf. Appendix V-A).

If $U_{0:t} = \{U_\tau : \tau < t\}$ denotes the filtration generated by the process $\mathbf{U}_t$, then the filtering problem is to compute the posterior density $p_t(\mathbf{x}) = p(\mathbf{X}_t = \mathbf{x} \mid U_{0:t})$ or, equivalently, the posterior expectation $\mathbb{E}[\phi_t] := \mathbb{E}[\phi(\mathbf{X}_t) \mid U_{0:t}]$. In an uncorrelated noise setting, the Kushner–Stratonovich equation describes the temporal evolution of the posterior expectation $\mathbb{E}[\phi_t]$ [21, 22, 23] (Appendix V-A). To find a similar formal solution for filtering with observed state *increments* (i.e., correlated noise), we can account for the correlations between observations and the hidden state by introducing a slight modification in the Kushner–Stratonovich equation, resulting in a generalized Kushner–Stratonovich equation (gKSE) [23, Chapter 3.8] (cf. Nüsken et al. [20]). For our particular problem, we show in Appendix V-B that the dynamics of posterior expectations satisfy

$$d\mathbb{E}[\phi_t] = \mathbb{E}[\mathscr{L}[\phi_t]]dt + (\mathrm{cov}(\phi_t, \mathbf{f}_t) + \Sigma_x \mathbb{E}[\nabla \phi_t])^T \cdot \boldsymbol{C}^T \widetilde{\Sigma}_u^{-1} (d\mathbf{U}_t - \boldsymbol{C}\mathbb{E}[\mathbf{f}_t]dt), \tag{7}$$

with $\mathrm{cov}(\phi_t, \mathbf{f}_t) = \mathbb{E}[\phi_t \mathbf{f}_t] - \mathbb{E}[\phi_t]\mathbb{E}[\mathbf{f}_t]$, and

$$\widetilde{\Sigma}_u = \boldsymbol{C}\Sigma_x \boldsymbol{C}^T + \Sigma_u . \tag{8}$$

Note that the right-hand side of (7) does not only depend on $\mathbb{E}[\phi_t]$, but also on $\mathbb{E}[\phi_t \mathbf{f}_t]$, $\mathbb{E}[\nabla \phi_t]$ and other expectations of potentially nonlinear functions, which in general cannot be computed from $\mathbb{E}[\phi_t]$ alone. Therefore, in order to completely characterize the probabilistic solution, we would need one equation (7) for every moment of the posterior, with each moment corresponding to a specific choice of ϕ_t. Thus, except for a few very specific generative models, such as linear ones, the dynamics of posterior expectations in (7) will result in a system of an infinite number of coupled SDEs, which in general is analytically intractable.

Remark 1.

Equation (7) is the gKSE if state increments are the only available type of observations. In Appendix V-B, we extend it to the Kushner–Stratonovich equation when both state increments and (Gaussian-type) direct observations are present.

In what follows, we will approximately solve the continuous-time filtering problem with observed state increments by projecting the gKSE onto a submanifold of parametric densities with a finite number of parameters, resulting in a *finite* system of coupled SDEs for these parameters. For this, we will first review the general projection method, which so far has only been applied to classical filtering problems with uncorrelated state and observation noise, and then extend this framework to filtering problems with observed hidden state increments.

III. Projection filtering for observed continuous-time state increments

A. The general projection filtering method

Projection filtering is a method for approximate nonlinear filtering that is based on differential geometry. In this subsection, we outline briefly the differential geometric setup and derivation of the projection filter, and refer the reader to the seminal papers on projection filters [17, 24, 18], or the intuitive introduction to the subject matter presented in [25], for more detailed derivations and in-depth discussion of the method.

In general, we can interpret the solution of a (stochastic) differential equation (such as the FPE (2)) as a (stochastic) vector field on an infinite-dimensional function space $\mathscr{M}$ of probability density functions p_t. If the vector field is stochastic (which is usually the case in nonlinear filtering [22]), we consider it to be given in Stratonovich form

$$dp_t = \mathscr{A}^\dagger[p_t]dt + \mathscr{B}^\dagger[p_t] \circ d\mathbf{U}_t, \tag{9}$$

which is the standard choice for stochastic calculus on manifolds. Let us further assume a parametrization $p_\theta(\mathrm{x}) := p(\mathbf{x};\theta)$ with a finite set of parameters $\boldsymbol{\theta} = \{\theta_1, \ldots \theta_m\} \in \Theta$, such that the solution to (9) is reasonably well approximated by these parametrized densities.

Projection filtering provides a solution to the filtering problem by evolving the parameters $\boldsymbol{\theta}$, thereby constraining the approximate posterior to evolve on a finite-dimensional submanifold $\mathcal{S} = \{p_\theta(\mathbf{x}); \boldsymbol{\theta} \in \Theta\}$ of $\mathscr{M}$, rather than on $\mathscr{M}$ itself. This implies that we have to project the vector field in (9) onto the tangent space $T_\theta\mathcal{S}$

$$T_\theta\mathcal{S} = \mathrm{Span}\left(\frac{\partial p_\theta}{\partial \theta_1}, \ldots, \frac{\partial p_\theta}{\partial \theta_m}\right) \subset L^1, \tag{10}$$

where the $\frac{\partial p_\theta}{\partial \theta_i}$ denote the basis vectors of this tangent space. Intuitively, an orthogonal projection minimizes at each timestep the distance between the true posterior p_t and its approximation p_θ with respect to a Riemannian metric which, for probability distributions, corresponds to the Fisher metric [26]

$$g_{ij} = \mathbb{E}_{p_\theta}\left[\frac{\partial \log p_\theta(\mathbf{x})}{\partial \theta_i}\frac{\partial \log p_\theta(\mathbf{x})}{\partial \theta_j}\right]. \tag{11}$$

This allows us to use the general orthogonal projection formula

$$\Pi_\theta[\mathscr{Z}] = \sum_i \sum_j g^{ij}\left\langle \mathscr{Z}, \frac{\partial p_\theta}{\partial \theta_i}\right\rangle_\theta \cdot \frac{\partial p_\theta}{\partial \theta_j}, \tag{12}$$

where Π_θ denotes the projection operator, the g^{ij} are the components of the inverse Fisher metric, and

$$\langle Z_1, Z_2 \rangle_\theta := \int_\Omega d\mathbf{x} \frac{Z_1(\mathbf{x}) Z_2(\mathbf{x})}{p_\theta(\mathbf{x})}, \qquad Z_1, Z_2 \in T_\theta \mathcal{S} \,. \tag{13}$$

is the inner product that is associated with the Fisher metric (see Proposition 1 in Appendix V-C).

To find the parameter updates resulting from this projection, we apply the projection (12) to the dynamics of the probability density p_t in (9). Since our approximate dynamics evolve along tangent vectors to the manifold parametrized by $\boldsymbol{\theta}$, the posterior p_θ will stay on this manifold. Hence, the left-hand side of (9) can be written in terms of the basis vectors of $T_\theta \mathcal{S}$ by using the chain rule:

$$\Pi_\theta[dp_\theta] = \sum_j \frac{\partial p_\theta}{\partial \theta_j} d\theta_j \,. \tag{14}$$

Further, by letting the projection act on the right-hand side of this equation, and consecutively comparing coefficients in front of the basis vectors $\frac{\partial p_\theta}{\partial \theta_j}$, we find the following Stratonovich SDEs for the parameters of the projected density following the evolution in (9):

$$\begin{aligned} d\theta_j = &\sum_i g^{ij} \left\langle \mathscr{A}^\dagger[p_\theta], \frac{\partial p_\theta}{\partial \theta_i} \right\rangle_\theta dt \\ &+ \sum_i g^{ij} \left\langle \mathscr{B}^\dagger[p_\theta], \frac{\partial p_\theta}{\partial \theta_i} \right\rangle_\theta \circ d\mathbf{U}_t \,. \end{aligned} \tag{15}$$

This is the result of Brigo et al. [18, Theorem 4.3].

To facilitate the comparison to the gKSE (7), we further slightly rewrite this SDE:

$$\begin{aligned} d\theta_j = &\sum_i g^{ij} \left(\int_\Omega dx \mathscr{A}^\dagger[p_\theta] \frac{\partial \log p_\theta}{\partial \theta} \right) dt \\ &+ \sum_i g^{ij} \left(\int_\Omega dx \mathscr{B}^\dagger[p_\theta] \frac{\partial \log p_\theta}{\partial \theta} \right) \circ d\mathbf{U}_t \end{aligned} \tag{16}$$

$$\begin{aligned} = &\sum_i g^{ij} \mathbb{E}_\theta \left[\mathscr{A} \left[\frac{\partial \log p_\theta}{\partial \theta_i} \right] \right] dt \\ &+ \sum_i g^{ij} \mathbb{E}_\theta \left[\mathscr{B} \left[\frac{\partial \log p_\theta}{\partial \theta_i} \right] \right] \circ d\mathbf{U}_t, \end{aligned} \tag{17}$$

where $\mathbb{E}_\theta[\,\cdot\,]$ here denotes the expectation with respect to the projected density p_θ, and $\mathscr{A}$ and $\mathscr{B}$ denote the adjoint of $\mathscr{A}^\dagger$ and $\mathscr{B}^\dagger$ with respect to the L^2 scalar product. Rewriting the SDE in such a way allows us to immediately identify the operators $\mathscr{A}$ and $\mathscr{B}$ on the

right-hand side of this equation with the operators used for propagating expectations rather than densities.

To illustrate how to identify the operator $\mathscr{A}$ concretely, let us consider the filtering problem *without* any observations, i.e., a simple diffusion, which is formally solved by the Fokker–Planck Equation (2). Noting that the operator $\mathscr{L}^\dagger$ propagates the density through time, we identify $\mathscr{A}^\dagger = \mathscr{L}^\dagger$. Similarly, we find $\mathscr{A} = \mathscr{L}$ for the adjoint. Thus, the projection can be immediately determined from the time evolution of the expectation in Eq. (4), with $\phi_t = \frac{\partial \log p_\theta}{\partial \theta_i}$:

$$d\theta_j = \sum_i g^{ij} \mathbb{E}\left[\mathscr{L}\left[\frac{\partial \log p_\theta}{\partial \theta_i}\right]\right] dt\,. \tag{18}$$

If state increments are observed, expectations are propagated by using the gKSE (7), and identification of the operators $\mathscr{A}$ and $\mathscr{B}$ is possible after having transformed the gKSE to Stratonovich form, as we will see in the next section. Since this is usually easier to do than transforming the generalized Kushner equation for the posterior *density* in Stratonovich form, Eq. (17) is a more convenient choice for the parameter dynamics than Eq. (15).

Remark 2.—The derivation in the seminal papers on projection filtering [17, 18] follows a slightly different route but leads to the same result (15). We also refer the reader to the very accessible derivation presented in [25].

B. Projection with observed state increments

As we have seen, the adjunction between the SDE evolution operators for densities and associated expectations allowed us to use the expectation's evolution equation to derive the projection filter for a specific problem (18). The same applies if the evolution of the density (or equivalently, that of the expectations) is a *stochastic* differential equation, as is the case for the KSE and the gKSE (7), as long as these are given in Stratonovich form. This allows us to formulate the projection filter for filtering with observed state increments:

Theorem 1.—The projection filter for the filtering problem with observed state increments is given by the following SDE of the parameters θ of a projected density p_θ:

$$\begin{aligned} d\theta_j = \sum_i g^{ij} \Bigg[& \mathbb{E}_\theta\left[\widetilde{\mathscr{L}}\left[\frac{\partial \log p_\theta}{\partial \theta_i}\right]\right] \\ & - \frac{1}{2}\Bigg(\mathbb{E}_\theta\left[\left\|\widetilde{\Sigma}_u^{-1/2} C\,\mathbf{f}_t\right\|^2 \frac{\partial \log p_\theta}{\partial \theta_i}\right] \\ & + \mathbb{E}_\theta\left[\mathrm{Tr}\left(\widetilde{\Sigma}_u^{-1} C J_f \Sigma_x C^T\right)\frac{\partial \log p_\theta}{\partial \theta_i}\right]\Bigg)\Bigg] dt \\ & + \sum_i g^{ij}\left(\mathbb{E}_\theta\left[\frac{\partial \log p_\theta}{\partial \theta_i}\mathbf{f}_t\right] + \Sigma_x \mathbb{E}_\theta\left[\nabla \frac{\partial \log p_\theta}{\partial \theta_i}\right]\right)^T \\ & \cdot C^T \widetilde{\Sigma}_u^{-1} \circ d\mathbf{U}_t, \end{aligned} \tag{19}$$

with $(J_f)_{ij} = \frac{\partial f_i}{\partial x_j}$ denoting the Jacobian matrix, and with the short-hand modified generator

$$\begin{aligned}\widetilde{\mathscr{L}}[\phi_t] :&= (\nabla \phi_t(\mathbf{x}))^T C^{-1} \Sigma_u \widetilde{\Sigma}_u^{-1} C\, \mathbf{f}_t(\mathbf{x}) \\ &+ \frac{1}{2} Tr\Big(H_\phi \Sigma_x C^{-1} \Sigma_u \widetilde{\Sigma}_u^{-1} C \Big).\end{aligned} \tag{20}$$

Proof.: As a first step, let us rewrite the gKSE (7) in Stratonovich form (Corollary 3 in V-B):

$$\begin{aligned}d\mathbb{E}[\phi_t] = \Big[&\mathbb{E}\big[\widetilde{\mathscr{L}}[\phi_t]\big] - \frac{1}{2}\Big(\mathrm{cov}\Big(\phi_t, \big\| \widetilde{\Sigma}_u^{-1/2} C \mathrm{f}_t \big\|^2\Big) \\ &+ \mathrm{Tr}\Big(\widetilde{\Sigma}_u^{-1} C\, \mathrm{cov}(\phi_t, J_f) \Sigma_x C^T \Big)\Big)\Big] dt \\ &+ \big[\mathrm{cov}(\phi_t, \mathbf{f}_t) + \Sigma_x \mathbb{E}[\nabla \phi_t]\big]^T C^T \widetilde{\Sigma}_u^{-1} \circ d\mathbf{U}_t\,.\end{aligned} \tag{21}$$

This equation allows us to identify the operators $\mathscr{A}$ and $\mathscr{B}$ in Eq. (17) as the operations acting on ϕ_t in front of dt and $d\mathbf{U}_t$, respectively. By letting these operators act on $\phi_t = \frac{\partial \log p_\theta}{\partial \theta_j}$ in Eq. (17), and evaluating the expectations under the projected density p_θ, we obtain SDEs for the desired parameters. These can be further simplified by using

$$\mathbb{E}_\theta\Big[\frac{\partial \log p_\theta}{\partial \theta_j}\Big] = \frac{\partial}{\partial \theta_j} \int_\Omega dx\; p_\theta(x) = 0,$$

which yields (19). □

The 1D special case follows directly from (19).

Corollary 1.—*For univariate filtering problems, i.e. a filtering problem with $N = M = 1$ in* (1) *and* (6), *with $C = c$, $\Sigma_x = \sigma_x^2$, $\Sigma_u = \sigma_u^2$ and $\widetilde{\Sigma}_u = \tilde{\sigma}_u^2 = c^2\sigma_x^2 + \sigma_u^2$, the projection filter with observed state increments reads:*

$$\begin{aligned}d\theta_j = &\sum_i g^{ij} \Big[\frac{\sigma_u^2}{\tilde{\sigma}_u^2} \mathbb{E}_\theta\Big[\mathscr{L}\Big[\frac{\partial \log p_\theta}{\partial \theta_i} \Big]\Big] \\ &- \frac{c^2}{2\tilde{\sigma}_u^2}\Big(\mathbb{E}_\theta\Big[f_t^2 \frac{\partial \log p_\theta}{\partial \theta_i}\Big] + \sigma_x^2 \mathbb{E}_\theta\Big[\frac{\partial f_t}{\partial x} \cdot \frac{\partial \log p_\theta}{\partial \theta_i}\Big]\Big)\Big] dt \\ &+ \sum_i g^{ij} \Big[\frac{c}{\tilde{\sigma}_u^2}\Big(\mathbb{E}_\theta\Big[f_t \frac{\partial \log p_\theta}{\partial \theta_i}\Big] \\ &+ \sigma_x^2 \mathbb{E}_\theta\Big[\frac{\partial}{\partial x} \frac{\partial \log p_\theta}{\partial \theta_i}\Big]\Big)\Big] \circ dU_t\,.\end{aligned} \tag{22}$$

C. Projection on exponential family distributions

Analogous to [18], it is possible to derive explicit filter equations for the natural parameters of a projected exponential family distribution. Consider the following exponential family parametrization:

$$p_\theta(\mathbf{x}) := \exp\left(\boldsymbol{\theta}^\top \mathbf{T}(\mathbf{x}) - \Psi(\boldsymbol{\theta})\right), \tag{23}$$

where θ is the vector of *natural* or *canonical parameters*, $\mathbf{T}(x)$ is the vector of sufficient statistics and $\exp(\Psi(\boldsymbol{\theta}))$ is the normalization.

Corollary 2.—*The projection filter for the filtering problem with observed state increments is given by the following SDE of the natural parameters* $\boldsymbol{\theta}$ *of a projected density* p_θ *belonging to the exponential family:*

$$\begin{aligned} d\theta_j = \sum_i g^{ij} \Big[& \mathbb{E}_\theta\left[\widetilde{\mathscr{L}}[T_i(\mathbf{x})]\right] \\ & - \frac{1}{2}\Big(\mathbb{E}_\theta\left[\left\| \widetilde{\Sigma}_u^{-1/2} C\, \mathbf{f}_t \right\|^2 (T_i(\mathbf{x}) - \eta_i) \right] \\ & + \mathbb{E}_\theta\left[\mathrm{Tr}\left(\widetilde{\Sigma}_u^{-1} C J_f \Sigma_x C^T \right) (T_i(\mathbf{x}) - \eta_i) \right] \Big) \Big] dt \\ & + \sum_i g^{ij} \left(\mathbb{E}_\theta[(T_i(\mathbf{x}) - \eta_i)\mathbf{f}_t] + \Sigma_x \mathbb{E}_\theta[\nabla T_i(\mathbf{x})] \right)^T \\ & \cdot C^T \widetilde{\Sigma}_u^{-1} \circ d\mathbf{U}_t, \end{aligned} \tag{24}$$

where $\eta_i = \mathbb{E}_\theta[T_i(\mathbf{x})]$ is the posterior expectation of the sufficient statistic $T_i(\mathbf{x})$. The parameters $\boldsymbol{\eta}$ are sometimes referred to as dual or expectation parameters.

Proof.: Making use of the duality relation between natural and expectation parameters for exponential families,

$$\frac{\partial}{\partial \theta_i} \log p_\theta(\mathbf{x}) = T_i(\mathbf{x}) - \frac{\partial}{\partial \theta_i} \Psi(\theta) = T_i(\mathbf{x}) - \eta_i, \tag{25}$$

and the fact that $\frac{\partial}{\partial x} \eta_i = 0$, Eq. (24) follows directly from the projection filter with observed state increments (19). □

Example 1—(Generalized Kalman–Bucy filter). In order to demonstrate the general approach, let us consider a model with linear state dynamics

$$dX_t = aX_t dt + \sigma_x dW_t, \tag{26}$$

$$dU_t = c dX_t + \sigma_u dV_t\,. \tag{27}$$

Here, we will show that a projection on a Gaussian manifold with

$$p_\theta(x) = \mathcal{N}\left(x; \mu_t, \sigma_t^2\right) \tag{28}$$

results in dynamics for the parameters μ_t and σ_t that are consistent with the generalized Kalman–Bucy Filter for observed state increments [20, Section 4.2]. In fact, since Eqs. (26) and (27) are both linear, the posterior density is a Gaussian and thus the projection filter becomes exact. For this particular problem, the projection filter reads:

$$\begin{aligned} d\theta_j = \sum_i g^{ij} \Bigg[& \frac{\sigma_u^2}{\tilde{\sigma}_u^2} \mathbb{E}_\theta \left[a X_t \frac{\partial}{\partial x} \frac{\partial \log p_\theta}{\partial \theta_i} + \sigma_x^2 \frac{\partial^2}{\partial x^2} \frac{\partial \log p_\theta}{\partial \theta_i} \right] \\ & - \frac{c^2 a^2}{2\tilde{\sigma}_u^2} \mathbb{E}_\theta \left[X_t^2 \frac{\partial \log p_\theta}{\partial \theta_i} \right] \Bigg] dt \\ & + \sum_i g^{ij} \Bigg[\frac{c}{\tilde{\sigma}_u^2} \Bigg(a \mathbb{E}_\theta \left[X_t \frac{\partial \log p_\theta}{\partial \theta_i} \right] \\ & + \sigma_x^2 \mathbb{E}_\theta \left[\frac{\partial}{\partial x} \frac{\partial \log p_\theta}{\partial \theta_i} \right] \Bigg) \Bigg] \circ dU_t \,. \end{aligned} \tag{29}$$

We will use this to determine SDEs for the parameters μ_t and σ_t^2, with $\frac{\partial \log p_\theta}{\partial \mu_t} = \frac{x - \mu_t}{\sigma_t^2}$ and $\frac{\partial \log p_\theta}{\partial \left(\sigma_t^2\right)} = -\frac{1}{2\sigma_t^2} + \frac{(x-\mu_t)}{2\sigma_t^4}$, respectively, under the Gaussian assumption. First, the components of the Fisher information matrix of a Gaussian parametrized by its expectation parameters are given by

$$g_{\mu\mu} = \frac{1}{\sigma_t^2}, \qquad g_{\sigma^2\sigma^2} = \frac{1}{2\sigma_t^4}, \qquad g_{\sigma^2\mu} = g_{\mu\sigma^2} = 0. \tag{30}$$

Since the matrix is diagonal, the components of its inverse are $g^{\mu\mu} = g_{\mu\mu}^{-1} = \sigma_t^2$ and $g^{\sigma^2\sigma^2} = g_{\sigma^2\sigma^2}^{-1} = 2\sigma_t^2$. This considerably simplifies the projection in Eq. (29) for the time evolution of μ_t and σ_t. Explicitly carrying out the expectations in Eq. (29) under the assumed Gaussian density, the SDEs for these parameters read:

$$d\mu_t = a\mu_t dt + \frac{c}{c\sigma_x^2 + \sigma_u^2} \left(a\sigma_t^2 + \sigma_x^2 \right) \cdot \left(dU_t - ac\mu_t dt \right), \tag{31}$$

$$d\sigma_t^2 = \left[2a\sigma_t^2 + \sigma_x^2 - \frac{c^2}{c\sigma_x^2 + \sigma_u^2} \left(a\sigma_t^2 + \sigma_x^2 \right)^2 \right] dt \,. \tag{32}$$

We found these Ito SDEs from their Strontonovich form by noting for the first line that the quadratic variation between σ_t and the observations process U_t is zero. In other words, no correction term according to the Wong–Zakai theorem [27] is required, such that both Ito and Stratonovich form have the same representation. Note that for a nonlinear generative model, this will in general *not* be the case [18]. Equations (31) and (32) are identical to the

generalized Kalman–Bucy filter [20, Eqs. (62) & (65)], thus demonstrating the validity of our approach.

Despite being able to reproduce existing results, the main purpose of the projection filtering approach is to simplify potentially hard filtering problems such that the parameter SDEs become analytically accessible. This becomes particularly useful if a certain parametric form of the posterior is desired, for instance for computational reasons, and can be very appealing if expectations are easily carried out under the assumed posterior and the Fisher matrix is straightforward to invert or even diagonal.

Example 2—(Multivariate Gaussian with diagonal covariance matrix). From a computational perspective, projection on a Gaussian density with diagonal covariance matrix can be advantageous in certain situations. In particular, such a solution only requires equations for $2N$ parameters, instead of $N^2 + N$ for a general Gaussian with Fisher matrix components

$$g_{\mu_i,\mu_j} = \left(\Sigma_t^{-1}\right)_{ij}, \qquad g_{\mu_i,\sigma^2_{nm}} = 0,$$

$$g_{\sigma_{ij},\sigma nm} = \frac{1}{4}\left[\left(\Sigma_t^{-1}\right)_{in}\left(\Sigma_t^{-1}\right)_{jm} + \left(\Sigma_t^{-1}\right)_{im}\left(\Sigma_t^{-1}\right)_{jn}\right],$$

which is in general hard to invert. For a Gaussian with diagonal covariance matrix, these simplify to

$$g_{\mu_i,\mu_t} = \left(\Sigma_t^{-1}\right)_{ij}, \qquad g_{\sigma^2_{ii},\sigma^2 ii} = \frac{1}{2}\left(\Sigma_t^{-1}\right)^2_{ii} = \frac{1}{2\sigma^2_{ii}},$$

while all other components evaluate to zero, making the Fisher matrix diagonal and straightforward to invert. Since the diagonality of the covariance matrix effectively decouples the dimensions, expectations can be carried out in each dimension separately. Nevertheless, the specific form of the parameter SDEs will crucially depend on the specific form of the nonlinear function $\mathrm{f}_t(\mathbf{x})$. For instance, considering the linear case, i.e. $\mathrm{f}_t(\mathbf{x}) = A\mathbf{x}$, yields

$$d\boldsymbol{\mu}_t = A\boldsymbol{\mu}_t + \left(\Sigma_x + \mathrm{diag}\left(\sigma^2_{ii}\right)A\right)C^T\widetilde{\Sigma}_u^{-1}(d\mathbf{U}_t - CA\boldsymbol{\mu}_t dt),$$

$$\begin{aligned}\frac{d\sigma^2_{ii}}{dt} &= 2\sigma^2_{ii}\left(A - \Sigma_x C^T\widetilde{\Sigma}_u^{-1}CA\right)_{ii} \\ &\quad -2\sigma^4_{ii}\left(A^T C^T\widetilde{\Sigma}_u^{-1}CA\right)_{ii} + \left(\Sigma_x - \Sigma_x C^T\widetilde{\Sigma}_u^{-1}C\Sigma_x\right)_{ii}.\end{aligned}$$

IV. Continuous-time circular filtering

In this section, we will consider continuous-time circular filtering with observed state increments as a concrete application of the framework derived above. We will further extend it to account for quasi continuous-time von Mises-valued observations (formally defined below), to provide a continuous-time generalization of the discrete-time circular filtering problem that is frequently encountered in spatial navigation problems.

A. Assuming observed angular increments only

We assume that the hidden state X_t is parametrized on S^1 by angle $\varphi_t \in [0, 2\pi)$, effectively embedding S^1 in $\mathbb{R}^2$ as a unit circle. We further assume that φ_t follows a diffusion on the circle:

$$d\varphi_t = f(\varphi_t, t)dt + \sigma_\varphi dW_t, \tag{33}$$

where W_t is now an $\mathbb{R}^1$-BM process, and drift and diffusion functions are as defined in Section II. The propagator $\mathscr{L}[\cdot]$ for this process is the same as for the corresponding process in $\mathbb{R}^1$ as given in Eq. (5).

For state processes that evolve on submanifolds of $\mathbb{R}^n$, as the S^1 considered here, Tronarp and Särkkä [19] have shown that projection filter equations are identical to the case where the state variable $\mathbf{X}_t$ evolves in Euclidean space. Since the mathematical operations performed to derive the projection filter with observed state increments in Theorem 1 are essentially the same as in Tronarp and Särkkä [19] for the state diffusion, their result carries over to our problem. Thus, Theorem 1 can straightforwardly be applied to the circular filtering problem by considering a circular projected density $p_\theta(\varphi)$, such as the von Mises or a wrapped normal distribution.

Example 3—(Circular diffusion with observed state increments). In this example, we explicitly model angular path integration as the estimation of a circular diffusion based on observed angular increments. Consider a model where the hidden state evolves according to a Brownian motion on the circle, with noisy observations of its increment

$$d\varphi_t = \frac{1}{\sqrt{\kappa_\varphi}} dW_t, \tag{34}$$

$$dU_t = d\varphi_t + \frac{1}{\sqrt{\kappa_u}} dV_t\,. \tag{35}$$

Here, φ_t could, for instance, correspond to the heading direction of an animal (or a robot) that is navigating in darkness and only has access to self-motion cues dU_t, i.e., measurements of angular increments, but not to direct heading cues such as landmark positions. We chose to parametrize the diffusion constants in terms of precisions, κ_φ and κ_u,

to make units comparable to that of the precision of the projected density, which we will denote κ_t. Thus, the parameter κ_φ governs the speed of the hidden state diffusion, and the parameter κ_u modulates the reliability of the observation process that is governed by the increments. The gKSE for this model's posterior expectation of a test function $\phi_t := \phi(\varphi_t)$ in Stratonovich form reads:

$$d\mathbb{E}[\phi_t] = \frac{1}{2(\kappa_\varphi + \kappa_u)}\mathbb{E}\left[\frac{\partial^2}{\partial\varphi^2}\phi_t\right]dt + \frac{\kappa_u}{\kappa_\varphi + \kappa_u}\mathbb{E}\left[\frac{\partial}{\partial\varphi}\phi_t\right] \circ dU_t\,. \tag{36}$$

As a result, the projection filter for the parameters θ becomes (cf. Eq. (19))

$$\begin{aligned} d\theta_j = \frac{1}{\kappa_\varphi + \kappa_u}\sum_i g^{ij}\Bigg[&\frac{1}{2}\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\varphi^2}\frac{\partial \log p_\theta}{\partial\theta_i}\right]dt \\ &+\kappa_u\mathbb{E}_\theta\left[\frac{\partial}{\partial\varphi}\frac{\partial \log p_\theta}{\partial\theta_i}\right]\circ dU_t\Bigg]. \end{aligned} \tag{37}$$

We now want to solve the circular filtering problem with observed state increments by projecting on the von Mises density

$$p_{\mu,\kappa}(\varphi) = \mathcal{VM}(\varphi;\mu_t,\kappa_t) = \frac{1}{2\pi I_0(\kappa_t)}\exp(\kappa_t\cos(\varphi-\mu_t)), \tag{38}$$

parametrized by mean μ_t and precision κ_t, using Eq. (37). Unlike e.g., the wrapped normal distribution, which is another popular choice for unimodal circular distributions, the von Mises distribution is an exponential family distribution and could alternatively be written in natural parametrization (23). Here, we chose to parametrize it by μ_t and κ_t, as it significantly simplifies the computation of the Fisher metric and its inverse, which appears on the right-hand side of (37). Noting that

$$\frac{\partial}{\partial\mu_t}\log p_{\mu,\kappa}(\varphi) = \kappa_t\sin(\varphi-\mu_t), \tag{39}$$

$$\frac{\partial}{\partial\kappa_t}\log p_{\mu,\kappa}(\varphi) = -F(\kappa_t) + \cos(\varphi-\mu_t), \tag{40}$$

where $F(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}$ denotes a ratio of Bessel functions, the components of the Fisher metric (with respect to the μ_t, κ_t parametrization) are given by:

$$g_{\mu\mu} = \kappa_t^2\mathbb{E}_{\mu,\kappa}\left[\sin^2(\varphi_t-\mu_t)\right] = \kappa_t F(\kappa_t), \tag{41}$$

$$g_{\kappa\kappa} = \mathbb{E}_{\mu,\kappa}\left[\left(F(\kappa_t) + \cos(\varphi_t - \mu_t)\right)^2\right] = 1 - \frac{F(\kappa_t)}{\kappa_t} - F(\kappa_t)^2, \tag{42}$$

$$g_{\mu\kappa} = g_{\kappa\mu} = 0. \tag{43}$$

Since the Fisher metric is diagonal, the components of its inverse are simply $g^{\mu\mu} = g_{\mu\mu}^{-1}$, $g^{\kappa\kappa} = g_{\kappa\kappa}^{-1}$ and $g^{\mu\kappa} = g^{\kappa\mu} = 0$.

Using these g_{ij}'s and explicitly computing the expectations on the right-hand side of Eq. (37) with respect to the von Mises approximation, we find the projection filter equations for this model:

$$d\mu_t = \frac{\kappa_u}{\kappa_\varphi + \kappa_u} \cdot d\mathbf{U}_t, \tag{44}$$

$$d\kappa_t = -\frac{1}{2(\kappa_\varphi + \kappa_u)} \mathscr{F}(\kappa_t) dt, \tag{45}$$

where we defined the strictly positive function

$$\mathscr{F}(\kappa_t) = \frac{F(\kappa_t)}{1 - \frac{F(\kappa_t)}{\kappa_t} - F(\kappa_t)^2}, \tag{46}$$

and found the Ito form of $d\mu_t$ from its Stratonovich form by noting that the noise variance is constant, making this conversion straightforward.

The projection filter defined by Eqs. (44) and (45) for orientation tracking in darkness has an intuitive interpretation: the mean μ_t is updated according to the angular increment observations, weighted by their reliability, as quanitified by κ_u. Even in the presence of such observations, the estimate's precision, κ_t, decays towards zero, since $\mathscr{F}(\kappa_t)$ is strictly positive. This decay reflects the accumulation of noisy observations. Very informative angular velocity observations with large κ_u may slow the decay, but cannot fully prevent it. In other words, without direct angular observations (which we will introduce in Section IV-B below), the estimate will inevitably become less accurate over time.

In Figure 1a, we illustrate in an example simulation that, despite the presence of angular increment observations, the estimate slowly drifts away from the true heading φ_t. As a benchmark we use a particle filter, and further compare mean μ_t and precision $r_t = F(\kappa_t)$ of the projection filter to that estimated by a Gaussian projection filter approximation (see Appendix V-F for details on these benchmarks). Such a filter relies on the assumption that the hidden state φ_t evolves on the real line, and thus leads to a slight deviation in the dynamics of the estimated precision κ_t.

Numerically, our projection filter's performance in this example is indistinguishable from that of the particle filter (Figure 1b, c), and its estimated precision r_t matches exactly the empirical precision evaluated by averaging the estimation error over 5000 simulation runs. The estimated precision of the Gaussian approximation, in contrast, systematically underestimates its precision for large observation reliability, and overestimates it when the angular increment observations become very noisy (Figure 1c).

Example 4—(Higher-order circular distributions). The previous example is one of the simplest examples of a circular filtering problem, and results in an approximated posterior that is always unimodal. This might be insufficient for certain settings in which we would like to consider more sophisticated projected densities. To show that our framework extends beyond such simple models, let us consider a class of circular distributions with exponential-family densities of the form

$$p(\varphi) = \frac{1}{Z(\mathbf{a}, \mathbf{b})} \exp\left[\sum_{k=1}^{K} a_k \cos(k\varphi) + b_k \sin(k\varphi)\right]. \tag{47}$$

The case $K = 1$ recovers the previously used von Mises distribution (38), but with a different parametrization. For $K = 2$, this density is referred to as the generalized von Mises distribution, whose properties have been studied extensively [28, 29]. As a proof of concept, we will now use the projection filter for the natural parameters (Eq. (24)) to project the solution to the generative model in Eqs. (34) and (35) onto a distribution with density (47).

The circular distributions with densities (47) belong to the exponential family distributions, with natural parameter vector $\theta = (\mathbf{a}, \mathbf{b})$ with $\mathbf{a} = \{a_1, \ldots, a_K\}$ and $\mathbf{b} = \{b_1, \ldots, b_K\}$, and sufficient statistics given by

$$T_k^{cos}(\varphi) = \cos(k\varphi), \qquad T_k^{sin}(\varphi) = \sin(k\varphi)\,. \tag{48}$$

The corresponding expectation parameters are defined by

$$\eta_k^{cos} = \mathbb{E}_\theta\left[T_k^{cos}\right], \qquad \eta_k^{sin} = \mathbb{E}_\theta\left[T_k^{sin}\right]. \tag{49}$$

According to Corollary 2, the projection filter for the natural parameters of an exponential family density requires us to apply the right-hand side of the gKSE (36) to the sufficient statistics. For this, we need to compute

$$\mathbb{E}_\theta\left[\frac{\partial}{\partial\varphi} T_k^{cos}(\varphi)\right] = -k\mathbb{E}_\theta[\sin(kx)] = -k\eta_k^{sin}, \tag{50}$$

$$\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\varphi^2} T_k^{cos}(\varphi)\right] = -k^2\mathbb{E}_\theta[\cos(kx)] = -k^2\eta_k^{cos}, \tag{51}$$

$$\mathbb{E}_\theta\left[\frac{\partial}{\partial\varphi}T_k^{sin}(\varphi)\right] = k\eta_k^{cos}, \tag{52}$$

$$\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\varphi^2}T_k^{sin}(\varphi)\right] = -k^2\eta_k^{sin}. \tag{53}$$

Furthermore, we note that the components of the Fisher matrix g_{ij} are given by

$$g_{ij} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\log Z(\mathbf{a},\mathbf{b}), \tag{54}$$

where the θ_i refer to the ith element in the parameter vector $\boldsymbol{\theta}$ that contains the elements of both a and b. Inverting this matrix to get the inverse components g^{ij} is in general not straightforward. In fact, already the Fisher metric needs to be computed numerically, as the normalization $Z(\mathbf{a}, \mathbf{b})$ is inaccessible in closed form. We will thus treat G^{-1} symbolically and sort the parameters such that G^{-1} is composed of the blocks

$$G^{-1} = \begin{pmatrix} \widetilde{G}^{c,c} & \widetilde{G}^{s,c} \\ \widetilde{G}^{s,c} & \widetilde{G}^{s,s} \end{pmatrix}, \tag{55}$$

Then, the projection filter for the natural parameters can be formally written as

$$\begin{aligned} d\mathbf{a} = \widetilde{G}^{c,c}\left(-\frac{1}{\kappa_\varphi+\kappa_u}\mathbf{k}^2\odot\boldsymbol{\eta}^{cos}dt - \frac{\kappa_u}{\kappa_\varphi+\kappa_u}\mathbf{k}\odot\boldsymbol{\eta}^{sin}\circ d\mathbf{U}_t\right) \\ +\widetilde{G}^{s,c}\left(-\frac{1}{\kappa_\varphi+\kappa_u}\mathbf{k}^2\odot\boldsymbol{\eta}^{sin}dt + \frac{\kappa_u}{\kappa_\varphi+\kappa_u}\mathbf{k}\odot\boldsymbol{\eta}^{cos}\circ d\mathbf{U}_t\right), \end{aligned} \tag{56}$$

$$\begin{aligned} d\mathbf{b} = \widetilde{G}^{s,c}\left(-\frac{1}{\kappa_\varphi+\kappa_u}\mathbf{k}^2\odot\boldsymbol{\eta}^{cos}dt - \frac{\kappa_u}{\kappa_\varphi+\kappa_u}\mathbf{k}\odot\boldsymbol{\eta}^{sin}\circ d\mathbf{U}_t\right) \\ +\widetilde{G}^{s,s}\left(-\frac{1}{\kappa_\varphi+\kappa_u}\mathbf{k}^2\odot\boldsymbol{\eta}^{sin}dt + \frac{\kappa_u}{\kappa_\varphi+\kappa_u}\mathbf{k}\odot\boldsymbol{\eta}^{cos}\circ d\mathbf{U}_t\right), \end{aligned} \tag{57}$$

where we denote with $\mathbf{k} = (1, \ldots, k)^\top$ the vector of values k, k^2 results from the element-wise squaring of $\mathbf{k}$, $\boldsymbol{\eta}$ is the vector of expectation parameters, and $\odot$ the element-wise (Hadamard) product. This example demonstrates that our framework could, in principle, be applied to project the posterior to more general and inevitably more complicated densities, should the need arise. Although we do not show this here explicitly, an instance where this might increase filtering accuracy is one where the initial density $p_0(\varphi)$ is multimodal. The example also shows that, in general, a projection filtering approach might not be the most practical approach. Here in particular, computation of the Fisher matrix components might only be possible numerically, and in that case is computationally expensive. This highlights the need

for a careful choice of the posterior density and its corresponding parametrization, where (ideally) expectations of sufficient statistics are available in closed-form.

B. Quasi continuous-time von Mises observations

So far we have focused on filtering algorithms that rely exclusively on observed angular increments. Such algorithms are bound to accumulate noise, such that their precision will decay to zero in the long run. To counteract this effect, let us now consider how to additionally include observations that are generated directly from the hidden state, rather than only its increments. Specifically, we will in this section propose an observation model for (quasi-)continuous time von Mises valued observations with $Z_t \in S^1$, which we will refer to as direct angular observations, because they are governed by the hidden state φ_t directly. This model will allow us to formulate a von Mises projection filter for both angular observations, and angular increment observations in the continuous-time circular filtering problem.

In classical continuous-time filtering settings, continuous-time observations Y_t are usually considered to follow a Gaussian diffusion process, whose drift component is governed by the hidden state X_t [23]. Equivalently, one could consider 'time-discretized' (or 'quasi-continuous') observations $\widetilde{Z}_t := \frac{d\widetilde{Y}_t}{\Delta t}$ with sampling time step Δt, according to

$$\widetilde{Z}_t \sim \mathcal{N}\left(h(X_t), \sigma_z^2 \Delta t^{-1}\right), \tag{58}$$

which is the usual setting of discrete-time filtering (with fixed Δt), with $h(x)$ being a potentially nonlinear function. Notably, the Fisher information $\mathcal{I}(X_t)$ about the state of the hidden variable X_t that is conveyed by these quasi-continuous observations $\widetilde{Z}_t$ grows linearly with sampling time step Δt (seel Proposition 2 in Appendix V-D). The consequence of this scaling is rather intuitive: decreasing the sampling time step Δt will result in overall more observations per unit time, which, in turn, are individually less informative about the state X_t. This renders the information rate (information per unit time) independent of the chosen time step.

Analogously, we now consider observations that are drawn from a von Mises distribution centered around a nonlinear S^1 valued transformation $h: S^1 \to S^1$ of the true hidden state φ_t,

$$Z_t \sim \mathcal{VM}(h(\varphi_t), \alpha(\kappa_z, \Delta t)). \tag{59}$$

We would like this observation model to have the same linear information scaling properties as the Gaussian observations encountered in the classical filtering problems, i.e. when hidden state noise and observation noise are uncorrelated. Thus, we need to choose the function $\alpha(\kappa_z, \Delta t)$ such that the information content about the state φ_t scales linearly with step size Δt and observation precision κ_z.

Theorem 2.—*If $\alpha(\kappa_z, \Delta t)$ is chosen such that*

$$\alpha(\kappa_z, \Delta t) = \xi^{-1}(\kappa_z \Delta t), \tag{60}$$

where ξ^{-1} *is the inverse of* $\xi(x) = xF(x)$ *(and* $F(x) = \frac{I_1(x)}{I_0(x)}$ *as defined earlier), then the information about the state of the random variable* φ_t *scales linearly with sampling time step and observation precision, i.e.,* $\mathcal{I}(\varphi_t) \propto \kappa_z \Delta t$.

Proof.: The information content about the random variable φ_t is given by the Fisher information

$$\mathcal{I}(\varphi_t) = \mathbb{E}_{Z_t}\left[\left(\frac{\partial}{\partial \phi} \log \mathcal{VM}(Z_t; h(\phi), \alpha)\right)^2 \mid \phi = \varphi_t\right] \tag{61}$$

$$= h'(\varphi_t)^2 \alpha \frac{I_1(\alpha)}{I_0(\alpha)} \propto \alpha F(\alpha)\,. \tag{62}$$

We require that the information content per time step Δt is constant and proportional in κ_z, which can be achieved if α varies with Δt and κ_z according to

$$\alpha(\kappa_z, \Delta t) F(\alpha(\kappa_z, \Delta t)) \propto \kappa_z \Delta t\,. \tag{63}$$

This can be achieved if $\alpha(\kappa_z, \Delta t) = \xi^{-1}(\kappa_z \Delta t)$, with

$$\xi(x) := x \cdot \frac{I_1(x)}{I_0(x)}\,. \tag{64}$$

□

Once a step size Δt is chosen, $\xi^{-1}(\kappa_z \Delta t)$ can be computed numerically. For sufficiently small $\kappa_z \Delta t$, e.g., in the continuum limit, this function can be approximated by $\xi^{-1}(\kappa_z \Delta t) \approx \sqrt{2\kappa_z \Delta t}$, while it becomes $\xi^{-1}(\kappa_z \Delta t) \approx \kappa_z \Delta t$ for large $\kappa_z \Delta t$ (Figure 2). The latter is consistent with the intuition that, for highly informative observations, the single observation likelihood is well approximated by a Gaussian (which breaks down in the limit $\Delta t \to 0$).

What we have considered here is in essence a modified discrete time observation model, which implies that we can take advantage of filtering methods available for circular filtering with discrete-time observations [13, 12]. However, by allowing the precision $\alpha(\kappa_z, \Delta t)$ to vary with time step, we additionally ensure that the information rate stays constant: increasing the time step will result in *more* observations per unit time, which is accounted for by less informative individual observations. Thus, the observation model defined in (59) and (60) constitutes a *quasi* continuous-time observation model.

C. Adding quasi continuous-time observations to the circular projection filter

If the measurement function $h(\varphi)$ in (59) is the identity, we can add the direct observations to our filter by straightforwardly making use of Bayes' theorem at every time step. Specifically, since we assumed our approximated (projected) density to be von Mises at all times, the measurement likelihood

$$p(Z_t \mid \varphi_t) = \mathscr{VM}(Z_t; \varphi_t, \alpha(\kappa_z, \Delta t)) \tag{65}$$

is conjugate to the density before the update $p_t^-(\varphi) := p(\varphi_t = \varphi \mid Z_{0:t-\Delta T}) = \mathscr{VM}(\varphi; \mu_{t-}, \kappa_{t-})$. In other words, the posterior $p_t(\varphi) := p(\varphi_t = \varphi \mid Z_{0:t})$ is guaranteed to be a von Mises density as well:

$$p_t(\varphi) \propto p(Z_t \mid \varphi_t) p_t^-(\varphi) \tag{66}$$

$$= \exp\left[\begin{pmatrix}\cos\phi_t \\ \sin\phi_t\end{pmatrix}^\top \left(\alpha(\kappa_z, dt)\begin{pmatrix}\cos Z_t \\ \sin Z_t\end{pmatrix} + \kappa_t - \begin{pmatrix}\cos\mu_{t-} \\ \sin\mu_{t-}\end{pmatrix}\right)\right]. \tag{67}$$

As expected from an exponential family distribution, the natural parameters $\boldsymbol{\theta}_t = \kappa_t(\cos\mu_t, \sin\mu_t)^\top$ are updated according to

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-} + \alpha(\kappa_z, \Delta t)\begin{pmatrix}\cos Z_t \\ \sin Z_t\end{pmatrix}. \tag{68}$$

This operation is equivalent to a summation of vectors in $\mathbb{R}^2$, where the natural parameters refer to Eucledian coordinates, and (μ, κ) are the corresponding polar coordinates (Figure 3a). In the continuum limit, we write

$$d\boldsymbol{\theta}_t = \sqrt{2\kappa_z dt}\begin{pmatrix}\cos Z_t \\ \sin Z_t\end{pmatrix}, \tag{69}$$

where we used that $\alpha(\kappa_z, dt) \to \sqrt{2\kappa_z dt}$ for $dt \to 0$. A coordinate transform from $\boldsymbol{\theta}$ to (μ, κ) recovers the update equations for mean μ_t and κ_t that result from quasi-continuous time observations

$$d\mu_t = d\,\arctan(\theta_2, \theta_1) = \frac{\sqrt{2\kappa_z dt}}{\kappa_t}\sin(Z_t - \mu_t), \tag{70}$$

$$d\kappa_t = d\sqrt{\theta_1^2 + \theta_2^2} = \sqrt{2\kappa_z dt}\cos(Z_t - \mu_t)\,. \tag{71}$$

The recovered update equations are appealingly simply for identity (or linear) observation functions, as such functions allow us to leverage Bayesian conjugacy properties. For nonlinear observation functions, in contrast, the posterior after the update is in general *not* in

the same class of densities (and in this particular case not a von Mises distribution anymore). To apply the projection filtering framework to project such a nonlinear observation model back onto the desired manifold of densities, one would need to know the vector field for the density p_t that incorporates the observation-induced update, a derivation that is beyond the scope of this work.[1]

Example 5—(The circular Kalman filter). Let us revisit angular path integration, which we introduced in Example 3 as a circular diffusion with observed angular increments, and extend it to include direct angular observations Z_t with likelihood (65). Such angular observations could, for instance, correspond to directly accessible angular cues, such as visual landmarks. Combining Eqs. (44) and (45) with Eqs. (70) and (71) to include the quasi-continuous updates results in

$$d\mu_t = \frac{\kappa_u}{\kappa_\varphi + \kappa_u} \cdot dU_t + \frac{\sqrt{2\kappa_z dt}}{\kappa_t} \sin(Z_t - \mu_t), \tag{72}$$

$$d\kappa_t = -\frac{1}{2(\kappa_\varphi + \kappa_u)} \mathcal{F}(\kappa_t) dt + \sqrt{2\kappa_z dt} \cos(\mu_t - Z_t)\,. \tag{73}$$

Due to the simplicity of these equations, as well as their structural similarity to the generalized Kalman filter, while taking full account of the circular state and measurement space, we coin the SDEs (72) and (73) the *circular Kalman filter (circKF)*.

Considering that $\sqrt{2\kappa_z dt}$ is modulated by the direct observation's reliability κ_z, the final terms in the filter's update equations nicely reflect the reliability-weighting that is already present in classical filtering problems: more reliable direct observations, relative to the current certainty κ_t, have a stronger impact on the mean μ_t. Furthermore, if the current observation Z_t is similar to the current estimate μ_t, this estimate is hardly updated, while the estimate's certainty κ_t increases. In contrast, if direct observations and current estimate are in conflict, the certainty κ_t might temporarily decrease (Figure 3a). What makes this last feature particularly interesting is that it only occurs in the circKF, but not in its Euclidean counterpart, the standard Kalman filter. In the latter an update induced by direct observations *always* leads to an increase in certainty. Numerically, the circKF features performance close to that of a particle filter over a wide range of parameters, while outperforming a Gaussian projection filter (Figure 3b, c). The reason for the deviation of the circKF from the optimal solution is that the von Mises distribution is still an *approximation* of the true posterior, which leads to slight deviations in the updates when the direct observations are integrated. These deviations are offset by the more than 10-fold decrease in computation time of the circKF that we observed in simulations: a single run in Figure 3c with the particle filter took 3.14±0.11s, while it only took 0.113±0.005s with the circKF on a MacBook Pro (Mid 2019) running 2.3 GHz 8-core Intel Core i9 using NumPy 1.19.2 on Python 3.9.1.

[1]As outlined earlier, for a Gaussian setting this would be the update part of the Kushner–Stratonovich equation. To the best of our knowledge, no such equation exists for von Mises-valued observations.

V. Conclusion

In this paper, we derived a continuous-time nonlinear filter with observed state increments, based on a projection filtering approach. Using this framework, we revisited the problem of probabilistic angular path integration. By additionally proposing a quasi continuous-time model for von Mises-valued direct observations, we were able to formulate a circular filtering algorithm that accounts for both increment and direct angular observations. Notably, this algorithm fulfills the following four conditions: (i) it operates on a circular state-space and (ii) in continuous time, (iii) it maintains a consistent representation during state propagation and observation update, as ensured by the projection filtering method, and (iv) it performs the proper integration of both increment and direct observations. Even though we have only fully worked out the algorithm for univariate circular filtering problems, we have formulated the overall projection filtering framework for more general multivariate problems. As by the results in [19], we expect our framework to carry over to multivariate circular filtering problems, such as reference vector tracking on the unit sphere. A possible shortcoming of this approach is that the class of generative models it can deal with is fairly limited. The generalized Kushner–Stratonovich equation is only valid if the error covariance of the observation process does not explicitly depend on the value of the hidden state. This constrains the generative model in the following way: first, we can only allow additive noise in the state process, as any multiplicative noise would enter the increment observation process U_t as state-dependent noise. Second, only *linear* transformations of state increments can be considered. As demonstrated in Example 4, another shortcoming of the projection-filtering method in general is that this approach is only computationally feasible for problems that are analytically accessible, i.e., where expectations under the projected density can be computed rapidly, or in closed form. In this paper, we thus focused on projected densities where these expectations could be efficiently computed.

Despite these limitations, the analytical accessibility and interpretability of our main result, in particular the circular Kalman filter, make it an attractive algorithm for unimodal circular filtering problems. First, it is straightforward to implement in software. Since its representation stays fixed through all times, it relies on only two equations which can be integrated straightforwardly, e.g. with an Euler–Maruyama scheme [30]. As we have seen in our numerical experiments, this makes this filter much faster than established methods, such as particle filters. Second, the interpretation of the dynamics are intuitively comprehensible. Third, since it is a continuous-time formulation, it automatically scales with respect to chosen sampling step size (as long as it is sufficiently small). This is an advantage over continuous-discrete filtering problems, which usually consider a fixed sampling step size, and need to be reformulated should the sampling rate in the observations change or vary across time. Lastly, animals navigate the world based on a continuous stream of sensory information, which motivates the use of continuous-time models when trying to understand how the brain operates under uncertainty. Thus, one possible application could be a conceptual description of how the brain performs angular path integration [31].

Acknowledgements

We would like to thank Melanie Basnak and Rachel Wilson (Harvard Medical School) for active and ongoing discussions that motivated the question of probabilistic path integration, and who helped us formulate the questions that we address in this manuscript. We further would like to thank Johannes Bill (Harvard Medical School) and Simone Surace (University of Bern), as well as the anonymous reviewers, for helpful feedback and suggestions on the manuscript.

FUNDING

A.K. was funded by the Swiss National Science Foundation (grant numbers P2ZHP2 184213 and P400PB 199242). J.D. and L.R. were supported by the Harvard Medical School Dean's Initiative Award Program for Innovative Grants in the Basic and Social Sciences, and a grant from the National Institutes of Health (NIH/NINDS, 1R34NS123819).

APPENDIX

A. Nonlinear filtering in a nutshell

Let us briefly review the classical nonlinear filtering setting, i.e., filtering with observations that follow a diffusion process with noise uncorrelated to that of the hidden state process, in order to compare this with filtering with observed state increments. In line with standard literature [23], let $\mathbf{Y}_t$ denote the process of $\mathbb{R}^M$-valued direct observations. Then the generative model commonly referred to in classical nonlinear filtering is given by

$$d\mathbf{X}_t = \mathrm{f}(\mathbf{X}_t, t)dt + \Sigma_x^{1/2} d\mathbf{W}_t, \tag{74}$$

$$d\mathbf{Y}_t = \mathbf{h}(\mathbf{X}_t, t)dt + \Sigma_y^{1/2} d\widetilde{\mathbf{V}}_t, \tag{75}$$

with $\widetilde{\mathbf{V}}_t$ and $\mathbf{W}_t$ independent standard Brownian motion processes, and $\mathbf{h}: \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^M$ a potentially nonlinear, vector-valued observation function. All other quantities have been defined below (1). Note that the observations process is a diffusion with error covariance $\Sigma_y dt$. The goal of nonlinear filtering is to compute dynamics of the posterior expectations $\mathbb{E}[\phi_t] := \mathbb{E}[\phi(\mathbf{X}_t) \mid Y_{0:t}]$, where $Y_{0:t} = \{\mathbf{Y}_\tau : \tau < t\}$ denotes the filtration generated by the process $\mathbf{Y}_t$. Formally, this is solved by the Kushner–Stratonovich equation (KSE, [23, Theorem 3.30]):

$$d\mathbb{E}[\phi_t] = \mathbb{E}[\mathscr{L}[\phi_t]]dt + \mathrm{cov}(\phi_t, \mathbf{h}_t)^T \Sigma_y^{-1} (d\mathbf{Y}_t - \mathbb{E}[\mathbf{h}_t]dt), \tag{76}$$

with $\mathrm{cov}(\phi_t, \mathbf{h}_t) = \mathbb{E}[\phi_t, \mathbf{h}_t] - \mathbb{E}[\phi_t]\mathbb{E}[\mathbf{h}_t]$.

B. Derivation of the generalized Kushner–Stratonovich equation for the posterior expectation (Eqs. (7) and (21))

Let us revisit the generative model in Eqs. (1) and (6):

$$d\mathbf{X}_t = \mathbf{f}(\mathbf{X}_t, t)dt + \Sigma_x^{1/2} d\mathbf{W}_t, \tag{77}$$

$$d\mathbf{U}_t = C\,\mathbf{f}(\mathbf{X}_t, t)dt + C\Sigma_x^{1/2} d\mathbf{W}_t + \Sigma_u^{1/2} d\mathbf{V}_t, \tag{78}$$

with $\mathbf{W}_t$ and $\mathbf{V}_t$ independent vector-valued Brownian motion processes, as defined earlier. Here, we derive the gKSE (7) by treating this model as a correlated noise filtering problem, which allows us to directly apply established results [23, Corollary 3.38].

Lemma 1.

The generalized Kushner–Stratonovich equation (gKSE) for the evolution of the conditional expectation of a test function $\mathbb{E}[\phi_t] = \mathbb{E}[\phi(X_t) \mid Y_{0:t}]$ *in the presence of state increment observations* $d\mathbf{U}_t$ *is given by (Itô form)*

$$\begin{aligned} d\mathbb{E}[\phi_t] = \mathbb{E}[\mathscr{L}[\phi_t]]dt + (\mathrm{cov}(\phi_t, \mathbf{f}_t) + \Sigma_x \mathbb{E}[\nabla \phi_t])^T \\ \cdot C^T \widetilde{\Sigma}_u^{-1} (d\mathbf{U}_t - C\mathbb{E}[\mathbf{f}_t]dt)\,. \end{aligned} \tag{79}$$

Proof.—To simplify calculations, we first require that the observations process has an error covariance that equals the identity, and thus we rescale the process $d\mathbf{U}_t$ by $\widetilde{\Sigma}_u^{1/2} = \left(C\Sigma_x C^T + \Sigma_u\right)^{1/2}$,

$$d\widetilde{\mathbf{U}}_t = \widetilde{\Sigma}_u^{-1/2} d\mathbf{U}_t\,. \tag{80}$$

Eqs. (77) and (80) define a filtering problem where the noise in hidden process $\mathbf{X}_t$ and observation process $\widetilde{\mathbf{U}}_t$ are correlated, with quadratic covariation

$$\left[\mathbf{X}, \widetilde{\mathbf{U}}^T\right]_t = \Sigma_x C^T \widetilde{\Sigma}_u^{-1/2}\,. \tag{81}$$

Thus, the result of [23, Corollary 3.38] (KSE for correlated noise filtering problems) is directly applicable to our problem. For more details on [23, Corollary 3.38], we kindly refer to the proof based on the innovations method provided therein. An alternative proof based on the change of measure approach is presented in [20].

Note that the KSE for correlated noise [23, Eq. 3.72] is similar to the classical KSE for uncorrelated noise problems (76), except for a correction term, given by a vector field $\mathscr{B}[\phi]$,

$$\begin{aligned} d\mathbb{E}[\phi_t] = \mathbb{E}[\mathscr{L}[\phi_t]]dt + \left(\mathrm{cov}\left(\phi_t, \widetilde{\Sigma}_u^{-1/2} C\mathbf{f}_t\right) + \mathbb{E}[\mathscr{B}[\phi_t]]\right)^T \\ \cdot \left(d\widetilde{\mathbf{U}} - \widetilde{\Sigma}_u C\mathbb{E}[\mathbf{f}_t]dt\right). \end{aligned} \tag{82}$$

The vector field $\mathscr{B}[\phi]$ can be read out from the dynamics of the quadratic covariation between the process ϕ_t and the rescaled observations process $\widetilde{\mathbf{U}}_t$. Using Ito's lemma, we can write:

$$d\phi_t = (\nabla\phi_t)^T \cdot d\mathbf{X}_t + \frac{1}{2}\mathrm{Tr}\big(\Sigma_x H_\phi\big)dt, \tag{83}$$

and thus

$$d\Big[\phi, \widetilde{\mathbf{U}}^T\Big]_t = (\nabla\phi_t)^T d\Big[\mathbf{X}, \widetilde{\mathbf{U}}^T\Big]_t \tag{84}$$

$$= (\nabla\phi_t)^T \Sigma_x C^T \widetilde{\Sigma}_u^{-1/2} dt =: \mathscr{B}[\phi_t]^T dt\,. \tag{85}$$

Further identifying $\mathbf{h}_t = \widetilde{\Sigma}_u^{-1/2} C\,\mathbf{f}_t$, we find

$$\begin{aligned} d\mathbb{E}[\phi_t] = \mathbb{E}[\mathscr{L}[\phi_t]]dt + (\mathrm{cov}(\phi_t, \mathbf{f}_t) + \Sigma_x \mathbb{E}[\nabla\phi_t])^T \\ \cdot C^T \widetilde{\Sigma}_u^{-1/2}\Big(d\widetilde{\mathbf{U}}_t - \widetilde{\Sigma}_u^{-1/2} C\mathbb{E}[\mathbf{f}_t]dt\Big). \end{aligned} \tag{86}$$

Rescaling $d\widetilde{\mathbf{U}}_t = \widetilde{\Sigma}_u^{-1/2} d\mathbf{U}_t$ yields (79).

Remark 3.—By combining (76) and (79) it is possible to derive a generalized Kushner equation when both types of observations are present, i.e., when we consider both the process $\mathbf{Y}_t$ (Eq. 75) and the process $\mathbf{U}_t$ (Eq. 78) as the observations:

$$\begin{aligned} d\mathbb{E}[\phi_t] = \mathbb{E}[\mathscr{L}[\phi_t]]dt + \mathrm{cov}(\phi_t, \mathbf{h}_t)^T \Sigma_y^{-1}(d\mathbf{Y}_t - \mathbb{E}[\mathbf{h}_t]dt) \\ +(\mathrm{cov}(\phi_t, \mathbf{f}_t) + \Sigma_x \mathbb{E}[\nabla\phi_t])^T C^T \widetilde{\Sigma}_u^{-1/2} \\ \cdot \Big(d\widetilde{\mathbf{U}}_t - \widetilde{\Sigma}_u^{-1/2} C\mathbb{E}[\mathbf{f}_t]dt\Big). \end{aligned} \tag{87}$$

In this case, the expectations $\mathbb{E}[\,\cdot\,]$ are with respect to the filtration $Y_{0:t}$ and $U_{0:t}$, i.e., $\mathbb{E}[\phi_t] = \mathbb{E}[\phi(\mathbf{X}_t) \mid Y_{0:t}, U_{0:t}]$.

Corollary 3.

In Stratonovich form, the generalized Kushner–Stratonovich equation (gKSE) for the evolution of the conditional expectation of a test function $\mathbb{E}[\phi_t] = \mathbb{E}[\phi(X_t) \mid Y_{0:t}]$ *in the presence of increment observations* $d\mathbf{U}_t$ *reads:*

$$\begin{aligned} d\mathbb{E}[\phi_t] = \Big[\mathbb{E}\big[\widetilde{\mathscr{L}}[\phi_t]\big] - \frac{1}{2}\Big(\mathrm{cov}\Big(\phi_t, \big\|\widetilde{\Sigma}_u^{-1/2} C\mathrm{f}\big\|_t^2\Big) \\ +\mathrm{Tr}\Big(\widetilde{\Sigma}_u^{-1} C \cdot \mathrm{cov}\big(\phi_t, J_f\big)\Sigma_x C^T\Big)\Big)\Big]dt \\ +(\mathrm{cov}(\phi_t, \mathbf{f}_t) + \Sigma_x \mathbb{E}[\nabla\phi_t])^T C^T \widetilde{\Sigma}_u^{-1} \circ d\mathbf{U}_t, \end{aligned} \tag{88}$$

with

$$\mathbb{E}\left[\widetilde{\mathscr{L}}[\phi_t]\right] = \mathbb{E}[\mathscr{L}[\phi_t]] - \mathbb{E}\left[(\nabla\phi_t)^T\Sigma_x C^T\widetilde{\Sigma}_u^{-1}C\mathrm{f}_t\right] - \frac{1}{2}\mathrm{Tr}\left(\Sigma_x\mathbb{E}\left[H_\phi\right]\Sigma_x C^T\widetilde{\Sigma}_u^{-1}C\right) \tag{89}$$

For one-dimensional problems with $N = M = 1$, $C = c$, $\Sigma_x = \sigma_x^2$, $\Sigma_u = \sigma_u^2$ *and* $\widetilde{\Sigma}_u = \tilde{\sigma}_u^2 = c^2\sigma_x^2 + \sigma_u^2$, (88) *simplifies to*

$$d\mathbb{E}[\phi_t] = \left[\frac{\sigma_u^2}{\tilde{\sigma}_u^2}\mathbb{E}[\mathscr{L}[\phi_t]] - \frac{c^2}{2\tilde{\sigma}_u^2}\left(\mathrm{cov}\left(\phi_t, f_t^2\right) + \sigma_x^2\mathrm{cov}\left(\phi_t, \frac{\partial}{\partial x}f_t\right)\right)\right]dt + \frac{c}{\tilde{\sigma}_u^2}\left[\mathrm{cov}(\phi_t, f_t) + \sigma_x^2\mathbb{E}\left[\frac{\partial}{\partial x}\phi_t\right]\right]\circ d\mathrm{U}_t\,. \tag{90}$$

Proof.—We convert between Itô and Stratonovich calculus using the Wong–Zakai theorem [27]:

$$\mathbf{B}_t^T\cdot d\mathbf{U}_t = \mathbf{B}_t^T\circ d\mathbf{U}_t - \frac{1}{2}d\left[\mathbf{B}^T, \mathbf{U}\right]_t, \tag{91}$$

where the symbol ∘ denotes Stratonovich calculus, and $\left[B^T, Y\right]_t$ is the quadratic covariation between the processes Y_t and $\mathbf{B}_t$. We identify (cf. Eq. (79))

$$\mathbf{B}_t^T = (\mathrm{cov}(\phi_t, \mathbf{f}_t) + \Sigma_x\mathbb{E}[\nabla\phi_t])^T C^T\widetilde{\Sigma}_u^{-1}, \tag{92}$$

and write

$$d\mathbb{E}[\phi_t] = \mathbb{E}[\mathscr{L}[\phi_t]]dt - \mathbf{B}_t^T C\mathbb{E}[\mathbf{f}_t]dt + \mathbf{B}_t^T\cdot d\mathbf{U}_t \tag{93}$$

$$= \mathbb{E}[\mathscr{L}[\phi_t]]dt - \mathbf{B}_t^T C\mathbb{E}[\mathbf{f}_t]dt + \mathbf{B}_t^T\circ d\mathbf{U}_t - \frac{1}{2}d\left[\mathbf{B}^T, \mathbf{U}\right]_t. \tag{94}$$

To obtain the change in quadratic covariation $d\left[\mathbf{B}^T, \mathbf{U}\right]_t$, it is helpful to find $d\mathbf{B}_t$ by Itô's lemma

$$d\mathbf{B}_t^T = (d\mathbb{E}[\phi_t\mathbf{f}_t] - \mathbb{E}[\phi_t]d\mathbb{E}[\mathbf{f}_t] - d\mathbb{E}[\phi_t]\mathbb{E}[\mathbf{f}_t] - d\mathbb{E}[\phi_t]d\mathbb{E}[\mathbf{f}_t] + d\mathbb{E}[\nabla\phi_t]\Sigma_x)C^T\widetilde{\Sigma}_u^{-1}\,. \tag{95}$$

The evolution of the expectations is obtained by straightforward application of the gKSE (79), substituting ϕ_t with the functions $\phi_t\mathbf{f}_t$, $\mathbf{f}_t$ and $\frac{\partial}{\partial x}\phi_t$. This will result in terms that

multiply $d\mathbf{U}_t$, which are those relevant for computing the change of the covariation process, $d\left[\mathbf{B}^T, \mathbf{U}\right]_t$. Further note that the quadratic covariation of the observation process evolves according to $d\left[\mathbf{U}^T, \mathbf{U}\right]_t = \mathrm{Tr}(\Sigma_u)dt$. Some tedious but straightforward algebra and term rearrangements then result for the quadratic covariation in

$$\begin{aligned} d\left[\mathbf{B}^T, \mathbf{U}\right]_t = \Big[& \mathrm{cov}\Big(\phi_t, \left\|\widetilde{\Sigma}_u^{-1/2} C\,\mathbf{f}_t\right\|^2\Big) \\ & + \mathrm{Tr}\Big(C^T\widetilde{\Sigma}_u^{-1} C \cdot \mathrm{cov}(\phi_t, J_f)\Sigma_x\Big) \\ & + 2\mathrm{cov}\Big(\Big(\Sigma_x C^T\widetilde{\Sigma}_u^{-1} C\,\mathbf{f}_t\Big)^T, \nabla\phi_t\Big) \\ & - 2\mathrm{cov}(\phi_t, \mathbf{f}_t) C^T\widetilde{\Sigma}_u^{-1} C\mathbb{E}[\mathbf{f}_t]^T \\ & + \mathrm{Tr}\Big(\Sigma_x \mathbb{E}\big[H_\phi\big]\Sigma_x C^T\widetilde{\Sigma}_u^{-1} C\Big)\Big] dt\,. \end{aligned} \tag{96}$$

Plugging this into (94), and again some algebra, yields

$$\begin{aligned} d\mathbb{E}[\phi_t] = \Big[& \mathbb{E}[\mathscr{L}[\phi_t]] - \mathbb{E}\Big[(\nabla\phi_t)^T\Sigma_x C^T\widetilde{\Sigma}_u^{-1} C\mathbf{f}_t\Big] \\ & - \frac{1}{2}\mathrm{Tr}\Big(\Sigma_x\mathbb{E}\big[H_\phi\big]\Sigma_x C^T\widetilde{\Sigma}_u^{-1} C\Big)\Big] dt \\ & - \frac{1}{2}\Big[\mathrm{cov}\Big(\phi_t, \left\|\widetilde{\Sigma}_u^{-1/2} C\mathbf{f}_t\right\|^2\Big) \\ & + \mathrm{Tr}\Big(\widetilde{\Sigma}_u^{-1} C \cdot \mathrm{cov}(\phi_t, J_f)\Sigma_x C^T\Big)\Big] dt \\ & (\mathrm{cov}(\phi_t, \mathbf{f}_t) + \Sigma_x\mathbb{E}[\nabla\phi_t])^T C^T\widetilde{\Sigma}_u^{-1} \circ d\mathbf{U}_t\,. \end{aligned} \tag{97}$$

This equation can be further simplified by noting that $\mathbb{I} - \Sigma_x C^T\Sigma_u^{-1} C = C^{-1}\Sigma_u\widetilde{\Sigma}_u^{-1} C$. Further, we can substitute

$$\begin{aligned} \mathbb{E}\big[\widetilde{\mathscr{L}}[\phi_t]\big] = \mathbb{E}[\mathscr{L}[\phi_t]] - \mathbb{E}\Big[(\nabla\phi_t)^T\Sigma_x C^T\widetilde{\Sigma}_u^{-1} C\mathbf{f}_t\Big] \\ - \frac{1}{2}\mathrm{Tr}\Big(\Sigma_x\mathbb{E}\big[H_\phi\big]\Sigma_x C^T\widetilde{\Sigma}_u^{-1} C\Big) \end{aligned} \tag{98}$$

in the first line, which yields Eq. (88). Equation (90) follows from (88) as the 1D special case.

C. Fisher metric and scalar product

Proposition 1.

The scalar product defined in Eq. (13),

$$\langle Z_1, Z_2\rangle_\theta := \int_\Omega d\mathrm{x}\frac{Z_1(\mathbf{x})Z_2(\mathbf{x})}{p_\theta(\mathbf{x})}, \quad Z_1, Z_2 \in T_\theta\mathcal{S}, \tag{99}$$

is the scalar product on $T_\theta\mathcal{S}$ *that is associated with the Fisher metric* [26]

$$g_{ij} = \mathbb{E}_{p_\theta}\left[\frac{\partial \log p_\theta(\mathbf{x})}{\partial \theta_i}\frac{\partial \log p_\theta(\mathbf{x})}{\partial \theta_j}\right]. \tag{100}$$

Proof.—Consider Z_1 and Z_2 to correspond to two of the basis vectors of the tangent space $T_\theta \mathcal{S}$, i.e., $Z_1 = \frac{\partial p_\theta(\mathbf{x})}{\partial \theta_i}$ and $Z_1 = \frac{\partial p_\theta(\mathbf{x})}{\partial \theta_j}$. Then

$$\left\langle \frac{\partial p_\theta}{\partial \theta_i}, \frac{\partial p_\theta}{\partial \theta_j} \right\rangle_\theta = \int_\Omega d\mathbf{x} \frac{1}{p_\theta(\mathbf{x})}\frac{\partial p_\theta(\mathbf{x})}{\partial \theta_i}\frac{\partial p_\theta(\mathbf{x})}{\partial \theta_j} \tag{101}$$

$$= \int_\Omega d\mathbf{x} \frac{\partial \log p_\theta(\mathbf{x})}{\partial \theta_i}\frac{\partial \log p_\theta(\mathbf{x})}{\partial \theta_j} p_\theta(\mathbf{x}) \tag{102}$$

$$= \mathbb{E}\left[\frac{\partial \log p_\theta(\mathbf{x})}{\partial \theta_i}\frac{\partial \log p_\theta(\mathbf{x})}{\partial \theta_j}\right] = g_{ij}\,. \tag{103}$$

This concludes the proof. □

D. Information scaling

Proposition 2.

The Fisher information $\mathcal{I}(X_t)$ about the hidden state variable X_t that is conveyed by Gaussian-type discrete-time observations,

$$\widetilde{Z}_t \sim \mathcal{N}\left(\widetilde{Z}_t; g(X_t), \Delta t^{-1}\right), \tag{104}$$

grows linearly with the time step Δt.

Proof.—The information content about the variable φ_t that is conveyed by the observation $\widetilde{Z}_t$ is given by the Fisher information

$$\begin{aligned}\mathcal{I}(X_t) &= \mathbb{E}_{\widetilde{Z}_t}\left[\left(\frac{\partial}{\partial x}\log \mathcal{N}\left(\widetilde{Z}_t; g(x), \frac{\sigma_z^2}{\Delta t}\right)\right)^2 \mid x = X_t\right] \\ &= \frac{g'(x)^2}{\sigma_z^4}(\Delta t)^2 \mathbb{E}_{\widetilde{Z}_t}\left[\left(\widetilde{Z}_t - g(x)\right)^2\right] \\ &= \frac{g'(x)^2}{\sigma_z^2}\Delta t \propto \Delta t\,.\end{aligned}$$

□

E. Details on numerical simulations

Our numerical simulations in Figures 1 and 3, corresponding to Examples 3 and 5, were based on artificial data generated from the true model equations. In particular, the "true" state φ_t is a single trajectory from Eq. (34), and observations are drawn at each time point from Eq. (35) and, in the case of Example 5, additionally from Eq. (65). To simulate trajectories and observations, we use the Euler–Maruyama approximation [30]. In this approximation, the time-discretized generative model with fixed time step size Δt in Examples 3 and 5 reads:

$$\varphi_t \sim \mathcal{N}\left(\varphi_{t-\Delta t}, \frac{1}{\kappa_\varphi}\Delta t\right) \mod 2\pi \tag{105}$$

$$\Delta U_t \sim \mathcal{N}\left(\varphi_t - \varphi_{t-\Delta t}, \frac{1}{\kappa_u}\Delta t\right) \tag{106}$$

$$Z_t \sim \mathcal{VM}\left(\varphi_t, \xi^{-1}(\kappa_z \Delta t)\right). \tag{107}$$

The same time-discretization scheme was used to numerically integrate the SDEs (44), (45), (72) and (73) for the von Mises parameters μ_t and κ_t. Unless stated otherwise, we used $\Delta t = 0.01$ in all our numerical simulations, and give times in units of κ_ϕ.

F. Benchmarks for numerical simulations

For our numerical simulations in Figures 1 and 3, corresponding to Examples 3 and 5, we used the following filtering algorithms to compare the circKF against.

1) Particle Filter:

1. As benchmark, we used a Sequential Importance Sampling/Resampling particle filter (SIS-PF, [32]), that was modified to account for state increment observations ΔU_t.

The N particles in the SIS-PF where propagated according to

$$\begin{aligned}&\pi\left(\varphi_t^{(j)} \mid \varphi_{t-\Delta t}^{(j)}, \Delta U_t\right)\\ &\quad = \mathcal{N}\left(\varphi_{t-\Delta t}^{(j)} + \frac{\kappa_u}{\kappa_u + \kappa_\varphi}\Delta U_t, \frac{1}{\kappa_\varphi + \kappa_u}\Delta t\right) \mod 2\pi,\end{aligned} \tag{108}$$

and each particle j was weighted at each time step according to

$$w_t^{(j)} = w_{t-\Delta t}^{(j)} \cdot \mathcal{VM}\left(Z_t; \varphi_t^{(j)}, \xi^{-1}(\kappa_z \Delta t)\right), \tag{109}$$

yielding an SIS for this model that is asymptotically exact in the $N \to \infty$ limit. Mean μ_t and precision r_t of the filtering distribution were determined at each time step according to a weighted average on the circle, i.e. the first circular moment:

$$r_t \exp(i\mu_t) = \sum_{j=1}^{N} w_t^{(i)} \exp\left(i\varphi_t^{(i)}\right). \tag{110}$$

For a von Mises distribution, the radius r of the first circular moment and the precision parameter κ are related via $r = \frac{I_1(\kappa)}{I_0(\kappa)}$, which is why we use r rather than κ in our plots.

In our simulations, we used $N = 10^3$ if direct angular observations Z_t were present, and $N = 10^4$ if only state increment observations were present. We re-sampled the particles whenever the effective number of particles, $N_{\text{eff}} = \sum_j \left(w^{(i)}\right)^{-2}$, was lower than $N/2$.

2) Gaussian approximation:

The reference filter, which we refer to as "Gauss filter", is a heuristic method that assumes posterior mean μ_t and variance σ_t to evolve according to a generalized Kalman–Bucy filter (Eqs. (31) and (32)). Such a filter is often referred to as "assumed density filter" (ADF) in the literature, which under certain conditions, such as for the circular filtering problem we consider here, becomes fully equivalent to a Gaussian projection filter (see [18, Section 7] for in-depth discussion). In order to make the resulting distribution circular, this Gaussian is consecutively approximated by a von Mises distribution via $\kappa_t \approx \sigma_t^{-2}$, resulting in the following update equations for the model in Example 3:

$$d\mu_t = \frac{\kappa_u}{\kappa_\varphi + \kappa_u} dU_t, \tag{111}$$

$$d\kappa_t = d\left(\frac{1}{\sigma_t^2}\right) = -\frac{1}{\kappa_\varphi + \kappa_u} \cdot \frac{1}{\kappa_t^2} dt\,. \tag{112}$$

For the model used in Example 5, this is combined with the observations in the same way as for the circKF. Note that in absence of direct angular observations z_t, the mean dynamics are the same as for the circKF, while κ_t deviates (as shown in Figure 1). When direct angular observations are present, this, in turn, affects the computation of the update in the mean dynamics, which leads to a worse numerical performance than the circKF.

G. Code availability

Jupyter notebooks to generate Figs. 1–3, the underlying simulation data as well as Python scripts to generate this data has been deposited at Zenodo, and is publicly available at https://doi.org/10.5281/zenodo.5820406.

References

[1]. Heinze S, Narendra A, and Cheung A, "Principles of insect path integration," Current biology : CB, vol. 28, no. 17, pp. R1043–R1058, Sep 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6462409/ [PubMed: 30205054]

[2]. Kreiser R, Cartiglia M, Martel JNP, Conradt J, and Sandamirskaya Y, “A Neuromorphic Approach to Path Integration: A Head-Direction Spiking Neural Network with Vision-driven Reset,” in 2018 IEEE International Symposium on Circuits and Systems (ISCAS), May 2018, pp. 1–5, iSSN: 2379–447X.

[3]. Taube JS, Muller RU, and Ranck JB, “Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis.” The Journal of Neuroscience, vol. 10, no. 2, pp. 420–35, 1990. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2303851 [PubMed: 2303851]

[4]. Seelig JD and Jayaraman V, “Neural dynamics for landmark orientation and angular path integration,” Nature, vol. 521, no. 7551, pp. 186–191, 2015. [PubMed: 25971509]

[5]. Skaggs WE, Knierim JJ, Kudrimoti HS, and McNaughton BL, “A Model of the Neural Basis of the Rat’s Sense of Direction,” Advances in neural information processing systems, p. 10, 1995.

[6]. Turner-Evans D, Wegener S, Rouault H, Franconville R, Wolff T, Seelig JD, Druckmann S, and Jayaraman V, “Angular velocity integration in a fly heading circuit,” eLife, vol. 6, p. e23496, May 2017. [Online]. Available: https://elifesciences.org/articles/23496

[7]. Xie X, Hahnloser RH, and Seung HS, “Double-ring network model of the head-direction system,” Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, vol. 66, no. 4, pp. 9–9, 2002.

[8]. Ma WJ, Beck JM, Latham PE, and Pouget A, “Bayesian inference with probabilistic population codes.” Nature Neuroscience, vol. 9, no. 11, pp. 1432–8, Nov 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/17057707 [PubMed: 17057707]

[9]. Fiser J, Berkes P, Orban G, and Lengyel M, “Statistically optimal perception and learning: from behavior to neural representations,” Trends in Cognitive Sciences, vol. 14, no. 3, pp. 119–130, Mar 2010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1364661310000045 [PubMed: 20153683]

[10]. Kalman RE, “A New Approach to Linear Filtering and Prediction Problems,” Transactions of the ASME Journal of Basic Engineering, vol. 82, no. Series D, pp. 35–45, 1960. [Online]. Available: http://fluidsengineering.asmedigitalcollection.asme.org/article.aspx?articleid=1430402

[11]. Kalman RE and Bucy RS, “New Results in Linear Filtering and Prediction Theory,” Journal of Basic Engineering, vol. 83, no. 1, pp. 95–108, Mar 1961. [Online]. Available: https://asmedigitalcollection.asme.org/fluidsengineering/article/83/1/95/426820/New-Results-in-Linear-Filtering-and-Prediction

[12]. Kurz G, Gilitschenski I, and Hanebeck UD, “Recursive Bayesian filtering in circular state spaces,” IEEE Aerospace and Electronic Systems Magazine, vol. 31, no. 3, pp. 70–87, Mar 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7475421

[13]. Azmani M, Reboul S, Choquel JB, and Benjelloun M, “A recursive fusion filter for angular data,” 2009 IEEE International Conference on Robotics and Biomimetics, ROBIO 2009, pp. 882–887, 2009, publisher: IEEE.

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

[15]. Kurz G, Gilitschenski I, and Hanebeck UD, “Recursive nonlinear filtering for angular data based on circular distributions,” 2013 American Control Conference, pp. 5439–5445, 2014.

[16]. Tronarp F, Hostettler R, and Sarkk a S, “Continuous-Discrete von Mises-Fisher Filtering on S2 for Reference Vector Tracking,” 2018 21st International Conference on Information Fusion, FUSION 2018, no. July, pp. 1345–1352, 2018.

[17]. Hanzon B. and Hut R, “New Results on the Projection Filter,” Research Memorandum, vol. 105, no. March, pp. 9457–9475, 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377221796004043

[18]. Brigo D, Hanzon B, and Le Gland F, “Approximate nonlinear filtering by projection on exponential manifolds of densities,” Bernoulli, vol. 5, no. 3, pp. 495–534, 1999. [Online]. Available: http://projecteuclid.org/euclid.bj/1172617201

[19]. Tronarp F. and Sarkk a S, “Continuous-Discrete Filtering and Smoothing on Submanifolds of Euclidean Space,” arXiv:2004.09335 [math, stat], Apr. 2020, arXiv: 2004.09335. [Online]. Available: http://arxiv.org/abs/2004.09335

[20]. Nusken N, Reich S, and Rozdeba PJ, “State and Parameter Estimation from Observed Signal Increments,” Entropy, vol. 21, no. 5, p. 505, May 2019. [Online]. Available: https://www.mdpi.com/1099-4300/21/5/505

[21]. Stratonovich RL, “Conditional Markov Processes,” Theory of Probability & Its Applications, vol. 5, no. 2, pp. 156–178, Jan 1960. [Online]. Available: http://epubs.siam.org/doi/10.1137/1105015

[22]. Kushner HJ, “On the Differential Equations Satisfied by Conditional Probablitity Densities of Markov Processes, with Applications,” Journal of the Society for Industrial and Applied Mathematics Series A Control, vol. 2, no. 1, pp. 106–119, Jan 1964. [Online]. Available: http://epubs.siam.org/doi/10.1137/0302009

[23]. Bain A. and Crisan D, Fundamentals of stochastic filtering, ser. Stochastic modelling and applied probability. New York: Springer, 2009, no. 60, oCLC: ocn213479403.

[24]. Brigo D, Hanzon B, and LeGland F, “A differential geometric approach to nonlinear filtering: the projection filter,” IEEE Transactions on Automatic Control, vol. 43, no. 2, pp. 247–252, 1998. [Online]. Available: http://ieeexplore.ieee.org/document/661075/

[25]. van Handel R. and Mabuchi H, “Quantum projection filter for a highly nonlinear model in cavity QED,” Journal of Optics B: Quantum and Semiclassical Optics, vol. 7, no. 10, pp. S226–S236, Oct 2005. [Online]. Available: https://iopscience.iop.org/article/10.1088/1464-4266/7/10/005

[26]. Amari S, Differential-geometrical methods in statistics, 2nd ed., ser. Lecture notes in statistics. Berlin; New York: SpringerVerlag, 1990, no. 28.

[27]. Wong E. and Zakai M, “On the relation between ordinary and stochastic differential equations,” International Journal of Engineering Science, vol. 3, no. 2, pp. 213–229, Jul 1965. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0020722565900455

[28]. Gatto R. and Jammalamadaka SR, “The generalized von Mises distribution,” Statistical Methodology, vol. 4, no. 3, pp. 341–353, Jul 2007. [Online]. Available: https://escholarship.org/uc/item/8m91b0p7

[29]. Gatto R, “Some computational aspects of the generalized von Mises distribution,” Statistics and Computing, vol. 18, no. 3, pp. 321–331, Sep 2008.

[30]. Kloeden PE and Platen E, Numerical solution of stochastic differential equations, corr. 3. print ed., ser. Applications of mathematics. Berlin: Springer, 2010, no. 23.

[31]. Kutschireiter A, Basnak MA, Wilson RI, and Drugowitsch J, “A Bayesian perspective on the ring attractor for heading-direction tracking in the Drosophila central complex,” Tech. Rep., Dec. 2021, bioRxiv, Cold Spring Harbor Laboratory, Section: New Results Type: article. [Online]. Available: https://www.biorxiv.org/content/10.1101/2021.12.17.473253v1

[32]. Doucet A, Godsill S, and Andrieu C, “On sequential Monte Carlo sampling methods for Bayesian filtering,” Statistics and Computing, p. 12, 2010.

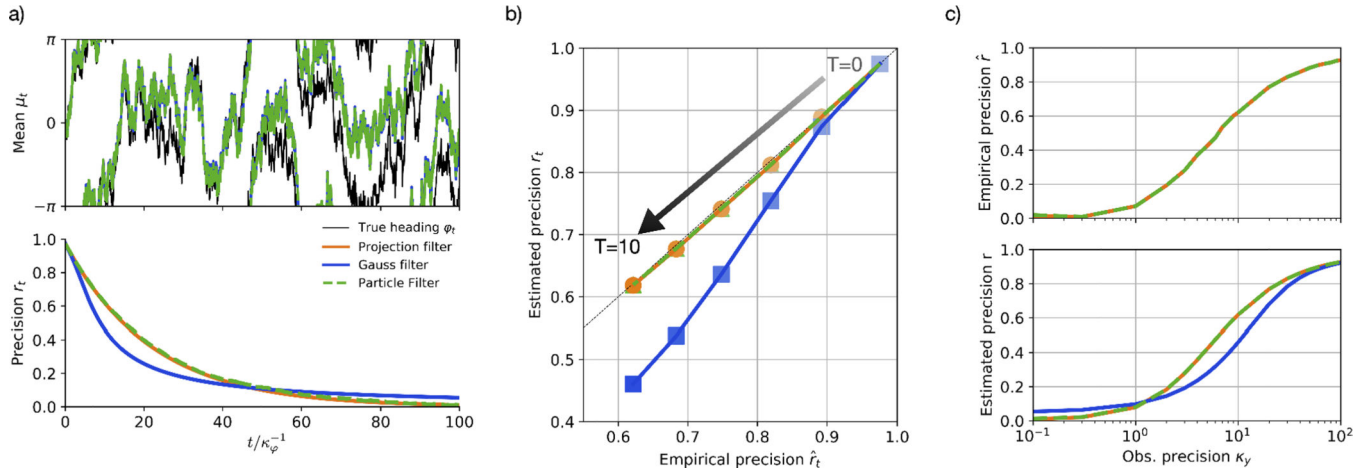

Fig. 1. Circular filtering with observed angular increments.

a) The von Mises projection filter in Eqs. (44) and (45) is able to track a diffusion process on the circle based on observed increments, with mean μ_t and precision r_t (given by $r_t = \frac{I_1(\kappa_t)}{I_0(\kappa_t)}$) matching that of a particle filter (curves are on top of each other). b) While the precision $\hat{r}_t$ estimated from the deviation between mean μ_t and true trajectory φ_t and the precision r_t estimated by the filter coincide for both the von Mises projection filter and the particle filter, the Gaussian projection filter ("Gauss filter") tends to underestimate its precision. c) Empirical (upper panel) and estimated (lower panel) precision for different values of the observations precision κ_u at time $T = 10\ \kappa_\varphi^{-1}$. Note that in the upper panel, the empirical precision $\hat{r}_t$ of the different filters is identical. Parameters for a) and b) are $\kappa_\varphi = 1$, $\kappa_u = 10$, times are in units of κ_φ^{-1}. Simulations in b) and c) were averaged over 5000 runs.

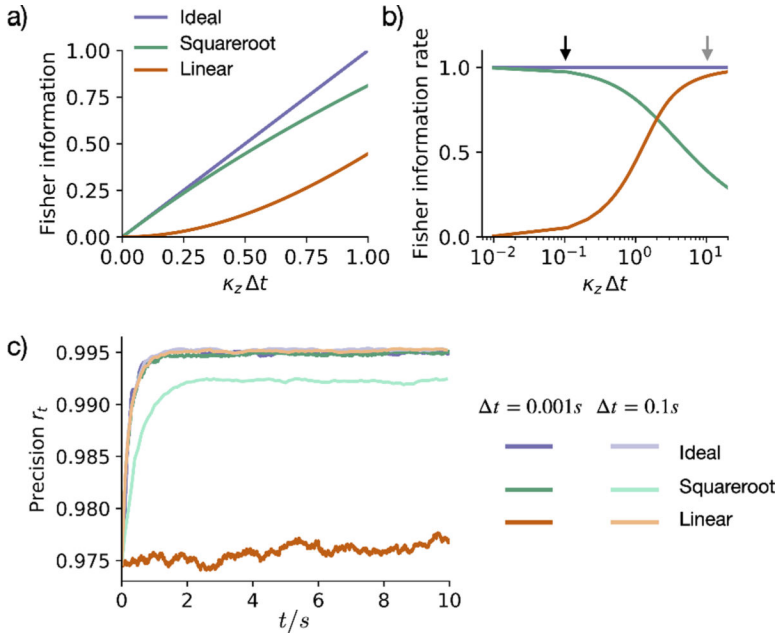

Fig. 2. Time scaling of quasi-continuous angular observations.

Both a) the Fisher information per observation and b) the Fisher information rate are linear and constant, respectively, in the sampling time step Δt when using $\alpha(\kappa_z, \Delta t) = \xi^{-1}(\kappa_z \Delta t)$ (”Ideal”). For comparison, we also plot $\alpha(\kappa_z, \Delta t) = \sqrt{\kappa_z \Delta t}$ (small $\kappa_z \Delta t$ approximation,”Squareroot”) and $\alpha(\kappa_z, \Delta t) = \kappa_z \Delta t$ (Gaussian approximation,”Linear”). c) Sample simulation with the constant position of the state $\varphi_t = \varphi$ estimated from quasi-continuous observations with different α functions and sampling time steps Δt. Ideally, the estimated precision r_t should be independent of the chosen simulation time step Δt. This is satisfied by all simulations except for the linear approximation for small $\kappa_z \Delta t$ (dark orange), and the square root approximation for large $\kappa_z \Delta t$ (light green). In these simulations, we used time units of seconds (s), and set $\kappa_\varphi = 100/s$ and $\kappa_z = 100/s$ (by design, κ_z has units of Fisher information per unit time), without loss of generality. Precision estimates were averaged over 10 simulation runs. Black and grey arrow in panel b) correspond to the two time step sizes shown in panel c).

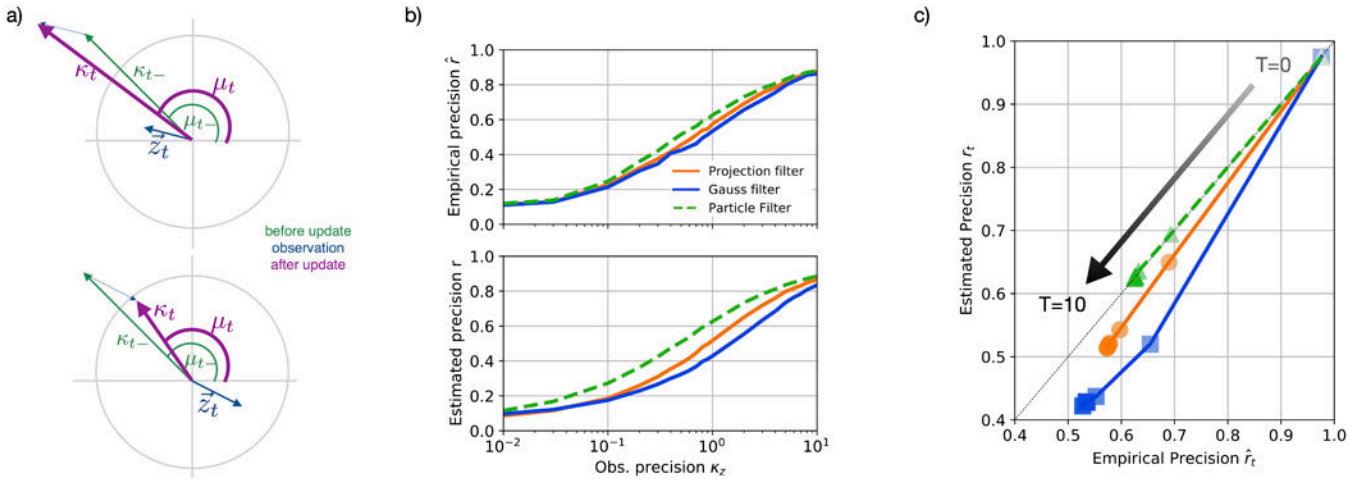

Fig. 3. Filtering with quasi-continuous time angular observations.

a) Single quasi-continuous time update step (68) with angular observation z_t, where the length of the vector indicates observation reliability $\alpha(\kappa_z dt)$. The update step for Bayesian inference on the circle is equivalent to a vector addition in the 2D plane. The lower panel demonstrates that a conflicting observation leads to a decreased certainty of the estimate directly after the update, corresponding to a shorter vector. b) Empirical precision $\hat{r}_t$ (upper panel) and estimated precision r_t of the circular Kalman filter (circKF) and a Gaussian projection filter (Gauss filter), when compared to a particle filter, for different values of the observation precision κ_z at time $T = 10\ \kappa_\varphi^{-1}$. Parameters were $\kappa_\varphi = 1$ and $\kappa_u = 1$, times are in units of κ_φ^{-1}. c) Estimated versus empirical precision up to $T = 10\ \kappa_\varphi^{-1}$ for the different filters at $\kappa_z = 10$. The precisions shown in b) and c) are averages across 5000 simulation runs.