

Non-parametric Mixed-Manifold Products using Multiscale Kernel Densities

Dehann Fourie¹, Pedro Vaz Teixeira², and John Leonard^{1,2}

Abstract—We extend the core operation of non-parametric belief propagation (NBP), also known as multi-scale sequential Gibbs sampling, to approximate products of kernel density estimated beliefs that reside on some manifold. The original algorithm, though multidimensional, implicitly assumes the beliefs to reside on the Euclidean \mathbb{R}^d space only. The proposed extension generalizes to any mixture of Riemannian manifolds, provided the primary operations—addition and subtraction—are defined. Our motivation is primarily focused on state-estimation using non-Gaussian factor graphs for multimodal simultaneous localization and mapping in robotics. The paper presents the method as well as simulation and experimental results for validation. Our implementation is publicly available and allows for expansion with user-defined manifold mixtures.

I. INTRODUCTION

Robust state estimation and sensor fusion are critical to many robotics applications. The ability to combine inertial navigation data (odometry) with ambiguous sensor measurements requires moving away from typical parametric belief models and towards more capable representations. While methods exist that allow for state estimation on certain manifolds, these are often specialized, and extension to a different manifold is likely to require time consuming modifications. Removing the constraints that limit an inference algorithm to operate on a pre-determined manifold allows for much greater flexibility in its application. To the best of the authors knowledge, there is no established generic method for posterior inference over a heterogeneous combination of manifolds using a shared belief representation. The aim is to extend our previous work on multi-modal incremental smoothing and mapping [1], [2] to handle arbitrary mixed-manifold belief representations with minimal restrictions. To this end, we propose a multi-scale Gibbs sampling process to approximate products from incoming, smooth likelihood densities defined across combinations of manifolds. This algorithm can readily be used in applications such as Bayesian filtering, signal processing, and factor graph optimization. As belief representations and inference methods are intimately related, we consider this extension a step towards fully functional methods that can be used in large scale inference problems, such as those found in simultaneous Localization and Mapping (SLAM) applications. We validate the proposed method using a simulated a target tracking

¹D. Fourie and J. Leonard are with the Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139, USA. {fourie}@csail.mit.edu

²P. V. Teixeira, J. Leonard are also with the Mechanical Engineering Department, MIT, Cambridge, MA 02139, USA. {pvt, jleonard}@mit.edu

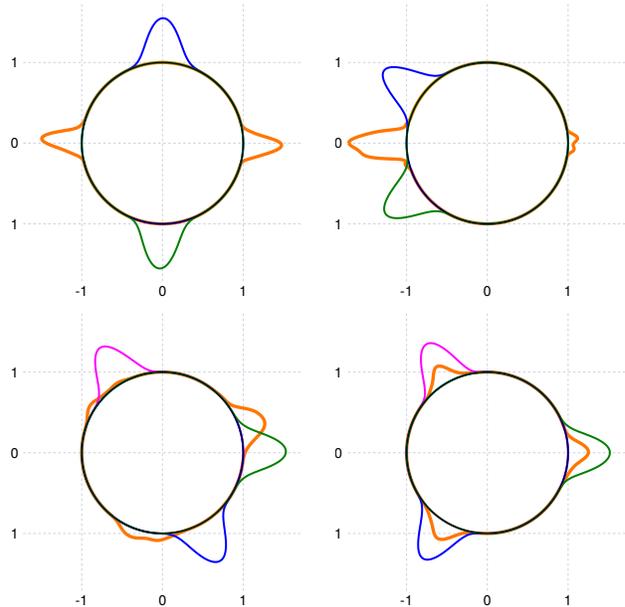


Fig. 1: Four separate examples of approximate products of kernel densities on single circular manifold. In each case, the orange trace represents the approximate density functional product (*posterior*) of the red, blue, or magenta probability traces (*likelihoods*). For illustration, the normalized probability densities are drawn as though wrapped around the unit circle with some scaling factor.

example, and present experimental results on non-Gaussian acoustic triangulation.

While the evaluation presented in this paper focuses on Euclidean, circular, cylindrical, or special Euclidean $SE(2)$ manifolds, the proposed technique is readily expandable to most, or all, mixed-manifold Riemannian spaces for which addition and subtraction operations can be defined and a local Euclidean tangent space expansion exists. Moreover, we aim to provide, by method and example, a guide on how to incorporate other manifolds not highlighted here.

The remainder of this paper is structured as follows: section II provides an overview of related work on non-parametric belief representations on manifolds and the associated inference tasks. The method is presented in section III, with simulation and experimental results in Section IV with canonical application examples that are relevant to many state estimation problems. Finally, we conclude in section V with a critical overview of the proposed method and pointers for future research directions.

II. RELATED WORK

Representation and inference techniques found in the literature can be broadly classified into one or more of the

following five categories: (i) directional statistics (wrapped Euclidean distributions); (ii) parametric (manifold-specific models); (iii) discrete grids; (iv) orthogonal-basis/wavelets (harmonic analysis); (v) kernel approximation (mixtures and embeddings).

Directional statistics generally modifies parametric models from the Euclidean manifold to obtain the desired statistics on the chosen manifold, notably the n -dimensional sphere S^n . As an example, the circular wrapped normal distribution $WN_K(\theta; \mu, \sigma)$ (discussed in greater detail in Section III-B) assumes an on-manifold difference operator $\ominus\theta \oplus \mu_\theta$, equivalent to the Euclidean difference $-x + \mu_x$.

Parametric stochastic processes on common manifolds are well documented in the literature [3]. Examples of manifold-specific parametric models include Bingham ($B(\cdot)$) [4] or Von Mises ($VM(\cdot)$) [5]. Combinations of various manifold distributions, such as a recently proposed $SE(2)$ parametric probability model by Gilitschenski et al. [6], [7], have also been suggested. Still, it should be noted that combinations of arbitrary manifold-specific models might not always capture significant conditional dependence structure between the dimensions, as required by the probabilistic chain rule, for example¹:

$$[X, \Psi] = [X | \Psi][\Psi] \neq [X][\Psi] \quad (1)$$

Beyond bespoke parametric density models—like the Kalman filter—several non-parametric methods exist: For example, the *particle filter* [8] generally builds on the simplicity/effectiveness of *importance sampling*, used to approximate the product between predicted and measured belief functions, as demonstrated in Rypkema et al. [9]. Particle filters are successful with lower dimensional state variables $\theta \in \Xi$, but quickly suffer from the *particle depletion* problem when used with high dimensional systems. This effect is a consequence of the rejection/resampling scheme used by importance sampling.

Histogram filtering methods [10] discretize some bounded region of the state variable domain into intervals [11], and utilize discrete update rules for predictions and measurements on a per-particle basis. Discretized methods are limited in the resolution and bounded domain over which they can operate. Examples include bounded circular or spherical intervals, using fixed grid Fourier series [12] or with spherical harmonics. Further examples include representing belief on a bounded interval with orthogonal basis functions such as polynomials or wavelets [13], [14], [15].

Reproducing Kernel Embeddings can readily be applied to manifold operations, but the techniques developed in machine learning require training data from the joint distribution [16], [17]. Unfortunately, many applications of interest, such as Simultaneous Localization and Mapping (SLAM), require non-Gaussian posterior estimation for which the joint distribution is unknown.

Some of the most significant recent work done on functional representations for probabilities on rigid transform

¹We use the following shorthand notation for probabilities: $\int_{\Xi} p(\Theta|\cdot) d\theta = \int_{\Xi} [\Theta|\cdot] d\theta = 1$

manifolds is by Duits et al. [18], [19]. Ongoing advances in reproducing kernel embeddings, particle flow homotopy [20], [21], [22], and wavelet analysis such as those by Duits et al. [18] allude that fully deterministic approaches to fully-functional inference may be possible.

Possibilities aside, kernel density estimation (KDE) [23] is well known and effective methods to compute multidimensional products already exist, for example work by Sudderth and Ihler et al. [24], [25], [26], [27], or Hanebeck et al. [22], [21]. Although KDE methods are easily modifiable for a number of dimensions, they are commonly restricted to the Euclidean manifold.

Finally, efficient KDEs representations depend on ball tree representations, which in turn have been extended in areas such as classification [28]. Here follows an extension—to mixed-manifold functional products—related to or extending techniques such as [27], [29], [16]. Note that this paper chooses to use the term *Euclidean manifold* rather than *Cartesian coordinates*.

III. BACKGROUND AND METHOD

This section discusses some background to mixed-manifold inference, followed by the method used to approximate products of infinite functionls on mixed manifolds.

A. Non-Gaussian Factor Graphs & Bayes (Junction) Tree

Factor graphs [30] are bipartite graphs of variables $\Theta_i \in \Xi_i$ and factors, offering a convenient language for describing the data-fusion inference task. Figures 4 and 7 each show small canonical factor graph examples. Each factor represents a noisy measurement (relative observation) $\tilde{z}_i \in \mathcal{Z}_i$ that sparsely interconnects a small set of system variables Θ'_i , and is expressed through probabilistic conditional likelihood models such as $[Z_i = \tilde{z}_i | \Theta'_i]$. Prior probabilities are unary (assumed absolute) and could apply to any variable – i.e. $[\Theta_j | z_j, \cdot]$, which themselves could be based on observations $Z_j = \tilde{z}_j \in \mathcal{Z}_j$. The notation (\cdot) is used to emphasize the dependence on some outside—or potentially iterative—process that was used to produce the prior factors $[\Theta_j | \cdot]$.

The multimodal-iSAM [2] approach to data-fusion is to approximate—through incremental/quasi-fixed-lag & distributed smoothing [1]—the marginal posterior belief $[\hat{\Theta}_m | \tilde{z}]$ of the full joint probability $[\Theta_m | \tilde{z}]$:

$$[\Theta_m | \tilde{z}] = \int_{\backslash\Xi_m} [\Theta | Z = \tilde{z}] d\theta_{\backslash m} \propto \int_{\backslash\Xi_m} \prod_i [z_i | \Theta'_i] \prod_j [\Theta_j | \cdot] d\theta_{\backslash m}, \quad (2)$$

with events $\theta_m \in \Xi_m$ (stochastic variables) and observations $z_i \in \mathcal{Z}_i$ (measurements). Eq. (2)—an algebraic refactored joint probability product—assumes that all measurements are uncorrelated $\tilde{z}_k \perp \tilde{z}_l | \Theta_i, \forall k \neq l$, since the *structure* between measurement observations is systematically modelled by the factor graph.

Furthermore, the mm-iSAM method also relies on the construction of a specialized *Bayes (Junction) tree* [31], [32],

[33], [1] of cliques. The Bayes tree specialization allows advanced features, such as incremental (recycled computation) / quasi-fixed-lag / distributed (federated) inference. A clique c is an algorithmic refactoring—Bayes network [33]—of several factors (likelihoods & priors) into groups $\{C\}$; which is represented by frontal $\Theta_{F,c}$ and separator variables $\Theta_{S,c}$ that describe an acyclic statistical (and exact) independence structure within the original factor graph – also known as the chordal property:

$$[\Theta | \mathbf{Z}] \propto \prod_{c \in \{C\}} [\Theta_{F,c} | \Theta_{S,c}]. \quad (3)$$

The particular nature of $\Theta_{F,c}, \Theta_{S,c}$ depends on each specific factor graph and tree. The theoretical product of all cliques would produce the same original factor graph joint distribution, but is explicitly never calculated due to its significant, high-dimensional complexity:

The mm-iSAM [1], [2] inference algorithm proceeds to find/estimate the marginal beliefs of each clique—also known as the Chapman-Kolmogorov transit integral—through a message passing (belief propagation/sum-product) *upward* process that assembles information from the leaves to root of the tree. A *downward* pass follows to spread all measurement information back to all variables back from the root until numerical changes become negligible.

$$[\Theta_{F,c} | \Theta_{S,c}] \propto \prod_{c \in \{C\}} [\tilde{\mathbf{z}}_c | \Theta_c] \prod_{d \in \{C\}} [\Theta_d | \tilde{\mathbf{z}}_d, \cdot]. \quad (4)$$

For the upward pass, at each clique the Chapman-Kolmogorov integral is computed (estimated) and the frontal variables are marginalized out and the separator variables are sent as a message $[\Theta_{S,c} | \mathbf{Y}_c]$, where \mathbf{Y}_c represents all measurements \mathbf{z}_i lower down on that branch of the tree. At each point during the inference process, partial & marginal posterior beliefs must be approximated, specific to the particular mixture of manifolds present in each state variable, where the product of several proposal distributions is a core operation.

B. On-Manifold, Arbitrary Function Estimation

The inference process requires a method of approximating the marginal beliefs for each of the variables during different stages of the inference processes. State variables must exist in some smooth metric space $\theta_i \in \Xi_i$ (containing an identity element θ_0) where on-manifold addition and subtraction are defined $(\Xi, \theta_0, \oplus, \ominus)$, as well as a metric distance $d(\theta, \theta') = \ominus \theta' \oplus \theta$ – i.e. a Riemannian manifold.

Kernel representations $k(\mu, \theta)$ utilize the local tangent space around $\mu \in \Xi_i$ and manifold smoothness conditions to approximate probabilistic events $\mathcal{P}(\Xi_i)$ on the local manifold tangent space. The kernel k is located by parameter μ , with assumed local support (spread) defined by bandwidth parameter Λ . By design, the representation is limited to a Hilbert space subset of all probability densities $\mathbf{P}(\Xi) \subset$

$\mathcal{P}(\Xi)$:

$$[\Theta | \cdot] \approx [\hat{\Theta}_i | \cdot], \quad [\Theta_i | \cdot] \in \mathcal{P}, \quad [\hat{\Theta}_i | \cdot] \in \mathbf{P} \\ \int_{\Xi_i} [\hat{\Theta}_i = \theta | \cdot]^2 d\theta < \infty, \quad \forall i \quad (5)$$

with multivariate density

$$\theta \in \Xi_i : [\Theta_i = \theta | \cdot] \rightarrow \mathbb{R}^+. \quad (6)$$

Finite dimensional—i.e. N' degrees of freedom—approximations of probability densities are taken as the sum of smooth normal (or radial basis) kernels, which are specifically chosen for their congruent property:

$$[\hat{\Theta} | \cdot] = \frac{1}{N} \sum_{n=1}^N k(d(\mu_n, \theta)). \quad (7)$$

Congruency is required so that the product of normal kernels produce yet another normal. This in turn allows all belief representations required during inference to be constructed in similar Hilbert spaces.

Additional numerical benefits might be attained on certain bounded domains—such as circular—by using a wrapped normal kernel:

$$k(\mu_i, \theta) = WN_K(d(\mu_i, \theta); \Lambda) \quad (8)$$

where theoretically $K \rightarrow \infty$ recovers infinite periodicity. In this work, however, we will restrict to $K = 1$ during practical computations that can use wrapped kernels. As an example, the circular manifold has distance $d(\mu, \theta) = \Delta\theta \in \Xi_i$ and $\Lambda' = \Lambda\sqrt{2\pi}$:

$$WN_K(\Delta\theta; \Lambda) = \frac{1}{\Lambda'} \sum_{\kappa=-K}^K \exp\left(-\frac{\|\Delta\theta \oplus 2\pi\kappa\|^2}{2\Lambda^2}\right) \quad (9)$$

The generic normal (or wrapped) kernel density definition allows introduction of new manifolds, on which to apply the method, after the inference algorithm development has completed; this is in stark contrast to many previous non-Gaussian (*functional*) methods that require prior knowledge of the domains before the algorithm is developed.

To summarize, the (as yet undefined) on-manifold probability density domain is approximated by eq. (7) with operations relating to $d(\mu_{i,n}, \theta_i)$ and Λ_i depending on manifold operations \oplus and \ominus .

Leave-one-out-cross-validation [23], [34] is used to select the bandwidth parameter Λ^* that has the minimum entropy,

$$\Lambda^* = \operatorname{argmin}_{\Lambda} -\hat{H}_{CV}(\{\mu\}_{1:N}, \Lambda) \quad (10)$$

with equal weighting and bandwidth parameter for all kernels, and \hat{H}_{CV} as the cross validation version of estimated entropy. Beirlant et al. [35] provides an overview of several entropy estimation techniques, as well as optimizations by Schwartz et al. [36].

1) *Utilizing On-Manifold (Metric) Ball Trees*: Kernel density estimates can be constructed with a binary ball tree algorithm [37], [38], such that multiple scales of the belief is captured at each of the depths of the ball tree. The root of the tree represents the on-manifold mean of all kernel centers combined, while the leaves of the ball tree represent the actual kernel center value across each of the manifolds. We modify the ball tree construction simply by ensuring that addition and subtraction manifold that meets the conditions discussed in III-B.

Ball tree construction starts at the root by finding the dimension along which the largest spread of points occur, and proceeds by splitting along that dimension at the equidistant mean. The measure of spread is proportional to the variance from the on-manifold mean for a given set of points. The mean is calculated as the on-manifold center of all associated points:

For example, if the split occurs on a circular dimension, then a relative exponential parameterization $[\hat{\psi}_\times] \in so(n)$ (Lie algebra) [3] of the equivalent rotational (Lie) group ${}^a_b\mathbf{R} = \exp[{}^a_b\psi_\times] \in SO(n)$ can be used to define the signed distance between elements;

$${}^\mu_i\psi = d({}^0_i\mathbf{R}, {}^0_\mu\mathbf{R}) = \log_{SO(n)}({}^0_\mu\mathbf{R}^T {}^0_i\mathbf{R}) \quad (11)$$

The on-manifold mean element is found when all distances from the mean to each kernel center accumulate to zero

$$\sum_{i=1}^d \Lambda_i^{-1} {}^\mu_i\psi = 0. \quad (12)$$

Each dimension of the ball tree could potentially be on a different manifold, where we assume the scale on each dimension is comparable; and while this is not strictly true in general, our empirical study shows that this does not seem like a significant limitation. Methods such as prescaling dimensions for improved balance may improve performance, but are not addressed in this paper. The kernel centers are then split in the direction according greater or smaller signed distance from the mean – i.e. elements landing “before” or “after” the local tangent plane expansion, see [38]. Figure 2 illustrates the ball tree construction process with a canonical example on the $SO(n=2)$ (circular) manifold.

C. Proposed Multiscale Gibbs Sampling Extension

The approximate product of infinite functionals through multi-scale Gibbs sampling [27], [25]—also known as *non-parametric belief propagation* (NBP)—leverages the ball tree representation to cascade the sampling procedure from coarse to fine as computation nests from root to leaves and promotes exposure of multiple modes. Incoming (likelihood) densities $[\hat{\Theta}_i | \cdot]$ are kernel density estimates, each with N wrapped normal kernel centers, bandwidths, and weights as in eq. (7).

The product is approximated by drawing a number of samples from the product of s incoming kernels and reconstituting a new kernel density estimate from those samples:

$$[\hat{\Theta}_i | \cdot] \propto \prod_{q=1}^s [\hat{\Theta}_i | \cdot]_q. \quad (13)$$

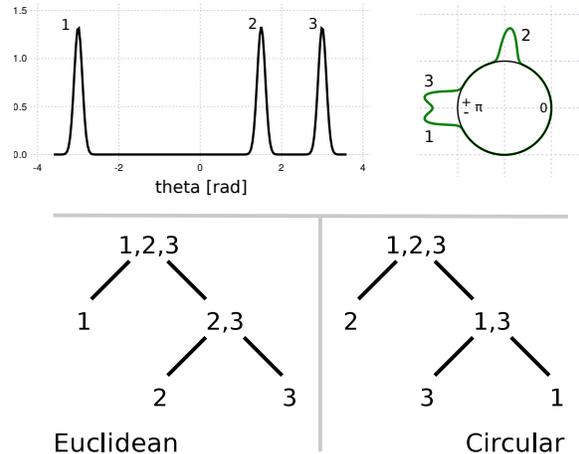


Fig. 2: Illustration of two contrasting ball trees. *Top*: shows a mixture probability density on the Euclidean or Circular manifolds. *Bottom*: shows the respective ball trees with (subset mean) binary divisions. The three kernel locations are $(1, 2, 3) \equiv (-3.0, 1.5, 3.0)$, with equal bandwidth 0.1 and equal weight 0.333. On the circular manifold, the first division (mean for points $(1, 2, 3)$) occurs at 2.594 rad, contrasted to an assumed Euclidean mean of 0.5.

For comparison, if the exact product were to be computed, then N^S new kernel parameters would need to be computed from all product combinations of incoming kernel components. Every one of M kernels representing the product density originates as the product of S separate kernels. One of N kernels is selected for each of S incoming kernel densities.

A Gibbs sampling strategy over the indices of incoming mixture components is used to stochastically select a subset (of size $M \approx N$) kernels in the product density from all possible kernels in the theoretical exact product of N^S kernels. The selection starts with an initial random label $l_{s,n}^{(0)}$ selection for each of the S incoming densities $[\hat{\Theta}_i | \cdot]_d$. A new label $l_{d,n}^{(0)}$ is selected for one of the incoming densities, say $s = 1$. The selection must only depend on information from the other $S - 1$ densities – i.e. $d \neq 1$:

$$\begin{aligned} l_1^{(1)} &\sim [L_1 | l_2^{(0)}, l_3^{(0)}, \dots, l_S^{(0)}] = \text{Cat}([\rho_1, \rho_2, \dots, \rho_N]) \\ l_2^{(1)} &\sim [L_2 | l_1^{(1)}, l_3^{(0)}, \dots, l_S^{(0)}] = \text{Cat}([\rho_1, \rho_2, \dots, \rho_N]) \\ &\dots \end{aligned} \quad (14)$$

The categorical label selection is probabilistically weighted by $\{\rho_n\}_{n=1}^N$. The weights ρ_n are computed by evaluating each kernel center $\mu_{1,n}$ of the $s = 1$ incoming density against the freshly computed temporary *near-product* kernel $\bar{k}(d(\bar{\mu}, \theta); \bar{\Lambda})$:

$$\rho_n = \bar{k}(d(\bar{\mu}, \mu_{1,n}); \bar{\Lambda}), \quad \forall n \in [1, N]. \quad (15)$$

The temporary near-product kernel is analytically computed from the remaining $S - 1$ kernels selected from the incoming

densities, as defined by the labels $l_2^{(0)}, l_3^{(0)}, \dots, l_S^{(0)}$:

$$\bar{k}(d(\bar{\mu}, \cdot); \bar{\Lambda}) = \prod_{s \neq 1}^S k(d(\mu_{s'}, \cdot); \Lambda_{s'}). \quad (16)$$

with $s' = l_s^{(0)}$. Note that the congruent property has been invoked on the kernel product. The new weights $\{\rho_n\}$ are used to probabilistically sample a new label $l_1^{(1)}$ index to replace the now stale label $l_1^{(0)}$. The process is repeated to sample a new $l_2^{(1)}$, as shown in eq. (14). After a given number of iterations, the process is frozen and a new product is calculated from all S densities similar to eq. (16) which is used as an outgoing product kernel. At this point the process drops one level down in the ball tree, and continues until a component product is obtained at the lowest level. The process is repeated until M outgoing samples have been generated.

The method for analytically calculating the on-manifold near-product kernel in eq. (16) is as follows: each selected kernel component $k(d(\mu_{s'}, \cdot); \Lambda_{s'})$ is a normal distribution about $\mu_{s'}$ with bandwidth $\Lambda_{s'}$. Figure 3 illustrates the how two normal kernels (the blue and red traces) are multiplied on a circular ($SO(2)$) manifold. The resulting density (magenta trace in figure 3) is obtained by calculating the on-manifold mean $\bar{\mu}$ and bandwidth $\bar{\Lambda}$ parameters.

The product between normal distributions on the Euclidean manifold can be computed by covariance/bandwidth according to $\bar{\Lambda} = \sum_{s=1}^S \Lambda_s$ while the mean is computed with $\bar{\Lambda} \bar{\mu} = \sum_{s=1}^S \Lambda'_s \mu_{s'}$. Through differential analysis we can see that the following expression also holds:

$$\bar{\Lambda} \Delta \bar{\mu} = \sum_{s=1}^S \Lambda'_s \Delta \mu_{s'} \quad (17)$$

where μ_L, Λ_L^{-1} are the mean and covariances for each term s of the product. Since near-product is the origin of the differential expansion, $\Delta \bar{\mu} = 0$ and implies the terms on the right hand side should sum to zero, which yields eq. (12). This can be solved for in any manner – either numerically, through nullspace methods, or more advanced parametric optimization such as Smith [39] or Dedieu et al. [40]. The results presented below used a numerical Newton optimization solution.

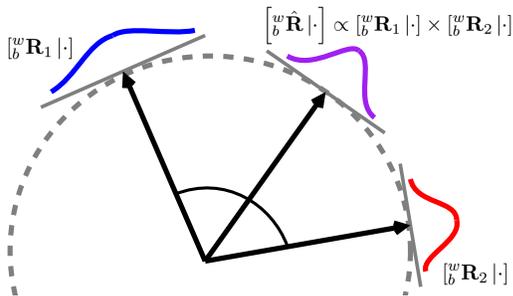


Fig. 3: Illustration of a normal kernel density product on a circular manifold. Red and blue density traces are multiplied together (as congruent kernels) to produce another normal density – i.e. the magenta curve.

IV. RESULTS

A. Simulation: Cylindrical Data (Tracking Problem)

This simulation example is based on the radar tracking problem, mechanized in polar coordinates. Polar coordinates are used as an example of a cylindrical manifold, where range encompasses the positive reals $[+0, +\infty)$ and angle is on the bounded interval $[-\pi, \pi)$. A simulation of a falling object with initial position and velocity is conducted, from where range and bearing measurements are constructed for the tracking problem. Figure 4 shows the simulated, and example estimated trajectories. The trajectory is offset about point A to induce cylindrical wrap around behavior that is seen in Figure 5.

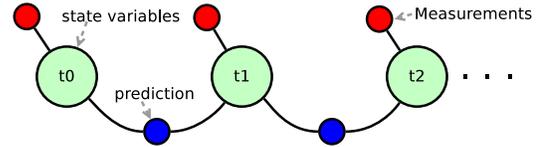
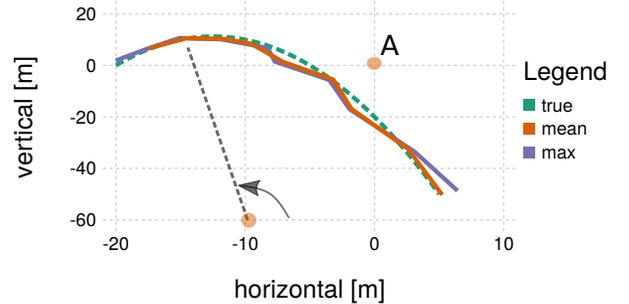


Fig. 4: *Top*: Illustration of classic target tracking example, with mean and max belief estimates from the inference. The dashed line and arc represent the polar coordinate tracking station measurements, as well as the coordinate system used in this example. For demonstrating wrap around behavior, the tracking and representation origin is offset to point A . *Bottom*: Factor graph representation of the target tracking task, with variable states as larger nodes, and factors as algebraic/probabilistic relationships between variables.

Figure 5 shows the unwrapped cylindrical manifold used for this example. The dashed red line shows the true trajectory of the object in range along the horizontal and angle along the vertical axes. Note the $t = 0$ s start and $t = 5$ s end point of the simulated trajectory. The density estimates presented as the blue regions along the true trajectory in Figure 5, and were calculated using the approximate manifold product method integrated with the non-Gaussian multimodal-iSAM factor graph solver [2], [1], [41]. Successive range, bearing measurements are introduced every couple hundred milliseconds (shown as red factors in Figure 4), with

B. Experiment: Acoustic Localization, $SE(2)$

A physical experiment is used to demonstrate and evaluate inference on the $SE(2) = SO(2) \times E(2)$ mixed-manifold. The manifold is composed of 2D Euclidean XY plane, combined with a bounded circular domain for rotation on $[-\pi, +\pi)$. The experiment intends to measure the speed of sound in air using inexpensive hardware and performing

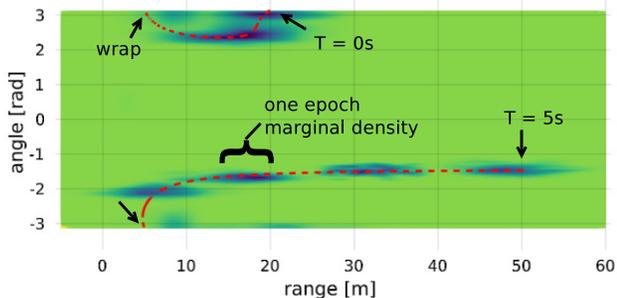


Fig. 5: Belief over cylindrical manifold of range and angle to object being tracked by a simulated radar. The dashed red line indicates the simulated trajectory, which wraps around at around $T = 2s$. Blue shading indicates different target marginal posterior belief estimates stacked for visualization. The target is successfully tracked along the manifold wrap-around, and note the marginal estimates are coupled across dimensions.

the localization inference using the approximate manifold products methodology outlined above.

The experiment consists of two acoustic beacons (speakers) placed at two corners of an equilateral triangle with the microphone array making the third. The microphone array’s position $[X, Y, \Theta | \mathbf{Z}]$ (a marginal posterior) can be estimated by measuring the actual angles using ultra-short-baseline beam forming [42] – a computational process to determine the amount of signal energy received per angular direction, based on constructive/destructive interference that is induced by the geometry of the received signals (phase shifts). The two up-chirps—shown in figure 6—are used to distinguish transmitted signals from each beacons (i.e. matched filtering [43]). Time separated chirps can also be used.

The captured energy per look-angle from the beam forming and matched filtering correlator is approximated with an on-manifold kernel density estimate function – an example is shown in figure 7. Assuming the four-element microphone array is pointed exactly perpendicular to the baseline between the two beacons, the acoustic wavefronts will arrive from either $+30^\circ$ or -30° , respectively.

A speed of sound parameter is usually known, however, this experiment will do a line search over the sound speed based on a comparison of localized vs. known geometry.

The estimated in air sound speed is around 350 m/s (at sea level at around 18°C), which is within 3% of the known value around 341 m/s . The measurement error is most likely

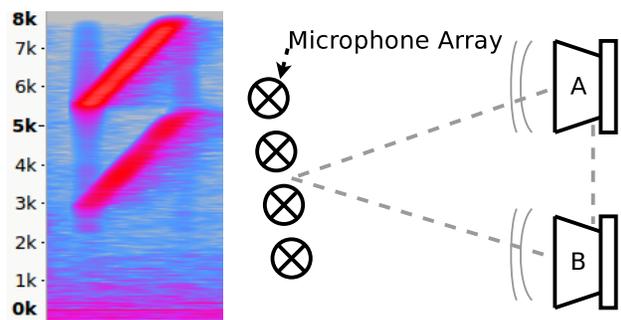


Fig. 6: Illustration of experiment where two acoustic sources transmitting unique chirps and recording microphone array. *Left*: example spectrogram of recorded up chirps from beacon A/B respectively, indicated in kHz . Chirps could be time separated or simultaneous.

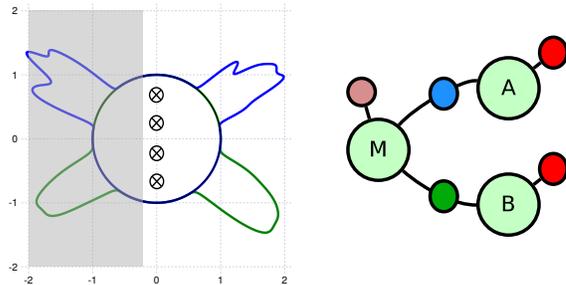


Fig. 7: *Left*: two beam patterns blue and green from beacon A and B. The shaded region is the symmetric (multi-modal) possibility behind microphone array. *Right*: The accompanying factor graph used for non-Gaussian acoustic localization with blue and green bearing factors with the non-Gaussian acoustic measurements. Red factors are priors on beacons, and a partial Y dimension constraint on the array location. X and orientation Θ are freely estimated.

due to experimental setup, and could be pursued as an experimental method to study algorithmic errors including bias or parasitics. Regardless of the 3% measurement tolerance, the trend in localization error when sound speed is altered in computation is consistent with our expectations, as illustrated in Figure 8.

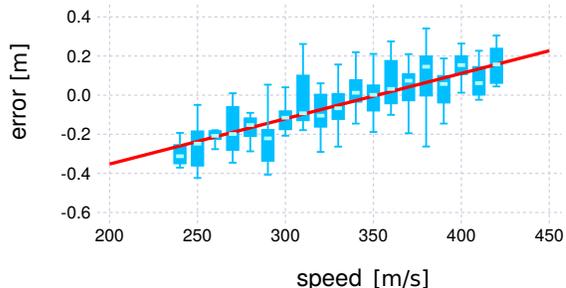


Fig. 8: Result of estimating the speed of sound with a US\$5 acoustic 4-element microphone array (Sony PS3eye), by monitoring the array localization error while varying the speed of sound parameter. The setup is illustrated in figures 6 and 7. Vertical axis represents the localization error of the array and horizontal axis is the chosen sound speed for each calculation, with the final estimate at the line-fit zero crossing near 350 m/s . See text for details.

V. CONCLUSION

This paper presents a manifold extension of the multi-scale Gibbs sampling method for approximating products between kernel density estimates – i.e. extending NBP [27]. The on-manifold extension is designed to accommodate most, if not, all Riemannian manifolds, allowing posterior Bayesian inference on many common and useful manifolds using the same methodology. Definition/implementation of an on-manifold \oplus and \ominus operator is the only additional requirement. Although comparisons to existing on-manifold parametric or MCMC methods are important, we feel that task is somewhat out-of-scope in this presentation since the method is a general (rather than specific) inference tool which should simplify many challenging inference tasks currently faced in robotics today: a user may directly use

this implementation on a new problem for which other *full-functional* methods likely first require a development cycle to build the necessary samplers or belief representations. Furthermore, the proposed method was validated on both real data and simulation canonical examples, and is currently used with the mm-iSAM method [1]. In future work, we wish to explore robust visual-inertial odometry by performing non-Gaussian posterior (belief) inference by propagating mixed manifolds of 20 dimensions or more.

ACKNOWLEDGEMENTS

This work was partially supported by the Office of Naval Research under grants N00014-18-1-2832 and N00014-16-1-2628, and the MIT CSAIL Systems That Learn initiative.

REFERENCES

- [1] D. Fourie, “Multi-Modal and Inertial Sensor Solutions to Navigation-type Factor Graphs.” Ph.D. dissertation, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, 2017.
- [2] D. Fourie, J. Leonard, and M. Kaess, “A nonparametric belief solution to the Bayes tree,” in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, IROS*, Daejeon, Korea, Oct 2016.
- [3] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*. Springer, 2011, vol. 2.
- [4] G. Kurz, I. Gilitschenski, S. Julier, and U. D. Hanebeck, “Recursive Bingham filter for directional estimation involving 180 degree symmetry,” *Journal of Advances in Information Fusion*, vol. 9, no. 2, pp. 90–105, 2014.
- [5] J. Straub, O. Freifeld, G. Rosman, J. J. Leonard, and J. W. F. III, “The Manhattan frame model: Manhattan world inference in the space of surface normals,” February 2017.
- [6] I. Gilitschenski, G. Kurz, and U. D. Hanebeck, “Bearings-only Sensor Scheduling Using Circular Statistics,” in *Proceedings of the 16th International Conference on Information Fusion*. IEEE, 2013, pp. 515–521.
- [7] I. Gilitschenski, G. Kurz, S. J. Julier, and U. D. Hanebeck, “A New Probability Distribution for Simultaneous Representation of Uncertain Position and Orientation,” in *17th International Conference on Information Fusion (FUSION)*. IEEE, 2014, pp. 1–7.
- [8] R. Van Der Merwe, A. Doucet, N. De Freitas, and E. A. Wan, “The Unscented Particle Filter,” in *Advances in neural information processing systems*, 2001, pp. 584–590.
- [9] H. S. N. R. Rypkema, E. M. Fischell, “One-way travel-time inverted ultra-short baseline localization for low-cost autonomous underwater vehicles,” in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2017, pp. pre–print.
- [10] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, Cambridge, MA, 2005.
- [11] G. Kurz, F. Pfaff, and U. D. Hanebeck, “Application of discrete recursive bayesian estimation on intervals and the unit circle to filtering on $se(2)$,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 3, pp. 1197–1206, 2018.
- [12] F. Pfaff, G. Kurz, and U. D. Hanebeck, “Multimodal circular filtering using Fourier series,” in *2015 18th International Conference on Information Fusion (Fusion)*. IEEE, 2015, pp. 711–718.
- [13] —, “Filtering on the Unit Sphere using Spherical Harmonics,” in *2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2017, pp. 124–130.
- [14] J. D. McEwen and B. Leistedt, “Fourier-Laguerre transform, Convolution and Wavelets on the Ball,” *arXiv preprint arXiv:1307.1307*, 2013.
- [15] F. J. Simons, F. Dahlen, and M. A. Wiecek, “Spatiospectral Concentration on a Sphere,” *SIAM review*, vol. 48, no. 3, pp. 504–536, 2006.
- [16] L. Song, K. Fukumizu, and A. Gretton, “Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models,” *Signal Processing Magazine, IEEE*, vol. 30, no. 4, pp. 98–111, 2013.
- [17] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [18] R. Duits, E. J. Bekkers, and A. Mashtakov, “Fourier Transform on the Homogeneous Space of 3D Positions and Orientations for Exact Solutions to PDEs,” 2018.
- [19] M. H. Janssen, A. J. Janssen, E. J. Bekkers, J. O. Bescós, and R. Duits, “Design and Processing of Invertible Orientation Scores of 3d Images,” *Journal of Mathematical Imaging and Vision*, vol. 60, no. 9, pp. 1427–1458, 2018.
- [20] K. Li, G. Kurz, L. Bernreiter, and U. D. Hanebeck, “Nonlinear Progressive Filtering for SE(2) Estimation,” in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 712–719.
- [21] U. D. Hanebeck, “FLUX: Progressive State Estimation Based on Zakai-type Distributed Ordinary Differential Equations,” *arXiv preprint arXiv:1808.02825*, 2018.
- [22] U. D. Hanebeck, K. Briechele, and A. Rauh, “Progressive bayes: a new framework for nonlinear state estimation,” in *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2003*, vol. 5099. International Society for Optics and Photonics, 2003, pp. 256–268.
- [23] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.
- [24] E. Sudderth, A. Ihler, W. Freeman, and A. Willsky, “Nonparametric belief propagation,” in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2003.
- [25] A. T. Ihler, E. B. Sudderth, W. T. Freeman, and A. S. Willsky, “Efficient multiscale sampling from products of Gaussian mixtures,” *Advances in Neural Information Processing Systems*, vol. 16, pp. 1–8, 2004.
- [26] A. T. Ihler and D. A. McAllester, “Particle belief propagation,” in *International Conference on Artificial Intelligence and Statistics*, 2009, pp. 256–263.
- [27] E. Sudderth, A. Ihler, M. Isard, W. Freeman, and A. Willsky, “Nonparametric belief propagation,” *Communications of the ACM*, vol. 53, no. 10, pp. 95–103, 2010.
- [28] T. Liu, A. W. Moore, and A. Gray, “New Algorithms for Efficient High-Dimensional Nonparametric Classification,” *Journal of Machine Learning Research*, vol. 7, no. Jun, pp. 1135–1158, 2006.
- [29] F. Pfaff, G. Kurz, and U. Hanebeck, “Multivariate Angular Filtering Using Fourier Series,” *Journal of Advances in Information Fusion*, vol. 11, no. 2, pp. 206–226, 2016.
- [30] F. Kschischang, B. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inform. Theory*, vol. 47, no. 2, Feb. 2001.
- [31] U. Kjærulff, “Inference in Bayesian networks using nested junction trees,” in *Learning in Graphical Models*. Springer, 1998, pp. 51–74.
- [32] M. Kaess, V. Ila, R. Roberts, and F. Dellaert, “The Bayes tree: An algorithmic foundation for probabilistic robot mapping,” in *Intl. Workshop on the Algorithmic Foundations of Robotics, WAFR*, Singapore, Dec. 2010.
- [33] J. Pearl, “Bayesian networks,” *Department of Statistics, UCLA*, 2011.
- [34] D. E. Lake, “Nonparametric Entropy Estimation using Kernel Densities,” *Methods in enzymology*, vol. 467, pp. 531–546, 2009.
- [35] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. Van der Meulen, “Nonparametric Entropy Estimation: An Overview,” *International Journal of Mathematical and Statistical Sciences*, vol. 6, no. 1, pp. 17–39, 1997.
- [36] S. Shwartz, M. Zibulevsky, and Y. Y. Schechner, “Fast kernel entropy estimation and optimization,” *Signal Processing*, vol. 85, no. 5, pp. 1045–1058, 2005.
- [37] S. M. Omohundro, *Five Balltree Construction Algorithms*. International Computer Science Institute Berkeley, 1989.
- [38] J. K. Uhlmann, “Metric trees,” *Applied Mathematics Letters*, vol. 4, no. 5, pp. 61–62, 1991.
- [39] S. T. Smith, “Optimization techniques on Riemannian manifolds,” *Fields institute communications*, vol. 3, no. 3, pp. 113–135, 1994.
- [40] J.-P. Dedieu, P. Priouret, and G. Malajovich, “Newton’s method on Riemannian manifolds: covariant alpha theory,” *IMA Journal of Numerical Analysis*, vol. 23, no. 3, pp. 395–419, 2003.
- [41] Package / Ecosystem Contributors. (2019) Caesar.jl. <https://github.com/JuliaRobotics/Caesar.jl>.
- [42] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation and modulation theory*. Wiley Online Library, 2002, vol. 1.
- [43] H. L. Trees, *Detection, Estimation, and Modulation Theory: Part III. Radar-Sonar Signal Processing and Gaussian Signal in Noise*. John Wiley & Sons, Incorporated, 2001.