MIT Open Access Articles

A nonparametric belief solution to the Bayes tree

A Nonparametric Belief Solution to the Bayes Tree

Dehann Fourie, John Leonard and Michael Kaess

Abstract— We relax parametric inference to a non-parametric representation towards more general solutions on factor graphs. We use the Bayes tree factorization to maximally exploit structure in the joint posterior thereby minimizing computation. We use kernel density estimation to represent a wider class of constraint beliefs, which naturally encapsulates multi-hypothesis and non-Gaussian inference. A variety of new uncertainty models can now be directly applied in the factor graph, and have the solver recover a potentially multi-modal posterior. For example, data association for loop closure proposals can be incorporated at inference time without further modifications to the factor graph. Our implementation of the presented algorithm is written entirely in the Julia language, exploiting high performance parallel computing. We show a larger scale use case with the well known Victoria park mapping and localization data set inferring over uncertain loop closures.
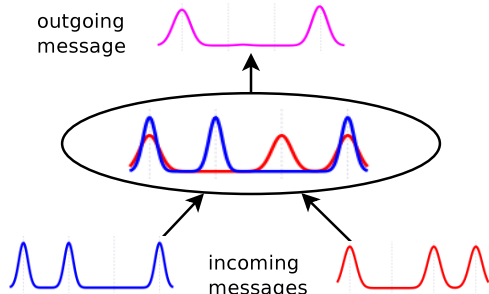
Fig. 1. Illustration of a Bayesian clique operation as part of a larger multi-modal belief propagation on a Bayes tree. Two incoming messages are *combined* with local potentials to produce one outgoing message during the upward pass procedure towards the root. Multi-modality is allowed to exist amongst cliques, rather than selecting a single mode as a maximum-product type algorithm would.

I. INTRODUCTION

Robots navigating in an unfamiliar world generally use a *front-end* process to combine various sensor measurements into a common data representation over which inference is performed to recover the spatial relationships, also known as state estimation. The inference process, referred to as the *back-end*, often faces erroneous constraints in the problem definition due to data association errors or sensor limitations which become more likely over time. For example, an imaging system might incorrectly identify two similar features as the same object. Our work is aimed at improving robustness and information efficiency in the presence of uncertain data associations. The key difference in our approach is that inference is performed over multi-modal posteriors using a non-Gaussian approximation to portions or even the entire system, and result in non-Gaussian posterior estimates.

Naive centralized processing over all sensory data would quickly generate an intractably enormous inference task. Our effort draws inspiration from the work of Tanner and Wong [1], as well as the Bayes tree [2] (closely related to the Junction tree), and extends their approaches to perform inference over multi-modal non-Gaussian constraints represented in the factor graph. Our goal is to use approximate belief propagation to enable multi-modal navigation.

We break away from previous unimodal parametric, i.e. non-linear least squares optimization, by explicitly tracking

D. Fourie is with the MIT/WHOI Joint Program with J.J. Leonard at the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. {dehann, jleonard}@mit.edu

M. Kaess is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA. kaess@cmu.edu

non-Gaussian belief in the joint distribution. Fig. 1 conceptually shows multi-modal belief propagation through a clique as part of a larger Bayes tree. The proposed algorithm captures ambiguity as uncertainty in the belief space, allowing consensus to be found through Bayesian computation across all available data. The final solution may very well still have several modes, which is in contrast to most existing SLAM solutions that force the solution to adopt a single mode result. Section III-D will explore these aspects in more detail.

In this paper we briefly present relevant background on existing robust robot navigation techniques. We then develop a methodology for finding consensus among all available constraints. A one dimensional robot state estimation example is used to illustrate concepts. We process the well known Victoria park dataset with ambiguous associations amongst (loop closure) tree detections.

II. PREVIOUS WORK

Most literature in simultaneous localization and mapping (SLAM), bundle adjustment (BA) and structure from motion (SfM) have focused on using parametric optimization routines to perform back-end inference. These use Gaussian parametric representations with a Quasi-Newton iterative optimization process to infer maximum likelihood estimates (MLE) for variable assignments.

Several robustness efforts have focused on removing "bad" constraints from the factor graph, either through more intelligent front-end, or preprocessing of the factor graph before actual variable assignment is done. Latif et al. [3] show the value of finding consensus at the front-end stage, delaying loop closure constraints until several new constraints agree and adding them to the factor graph as a batch of new constraints.

Recently in the SLAM community Graham [4] used an expectation maximization (EM) approach to smoothly transition poorly matched measurements to assumed outliers. The EM algorithm iterates between covariance weight selection and optimal variable assignments and suppresses "misbehaved" measurements by emphasizing the majority of constraints with consensus.

The method of Graham is local, and an analog variant with discrete switch variables is proposed by Sunderhauf et al. [5]. Switch variables introduce binary slack variables into the optimization that can enable or disable each measurement. Measurements which are inconsistent with the rest of the graph are discarded through multiplication by zero. Switch variables are comparable to a null-hypothesis approach [6], and has the disadvantage of ignoring information and relying heavily on tuning parameters.

Rather than retrofitting a single parametric solution with null-hypothesis, as suggested by all the afore mentioned methods, Huang et al. [7] and others propose parametric multi-hypothesis methods. Here each possible hypothesis is explicitly solved with a slightly different and distinct underlying factor graph. If all possible permutations are tracked, the correct solution will be contained by one out of the many available solutions. It is important to note that using a method like RANSAC on this collection of solutions is not valid, since each solution represents a different problem and not a Monte Carlo style variation on the same problem.

Olson et al. [6], [8] proposed the max-mixtures approach which selects the local maximal weighted Gaussian from a mixture of Gaussians before continuing with a parametric optimization routine. Their approach is akin to max-product inference, which greatly simplifies the inference problem by discarding multiple hypotheses when neighboring nodes share information. By analogy, the maximum-a-posteriori approach does slightly better by selecting the maximally weighted mode from each clique, i.e. connected grouping of variables.

FastSLAM [9] is conceptually similar by emphasizing a subset of permuted parametric solutions. FastSLAM is a midway between belief methods and multi-hypothesis parametric solutions. For example, under correct data association, a single particle can be used to recover the full SLAM solution. However, FastSLAM does not exploit structure within the joint probability distribution, unlike the Bayes tree [2]. Critically, the Bayes tree precisely encodes the type of structure needed for multihypothesis tracking, but this has not previously been studied in detail.

Hybrid inference on the Bayes tree, by Segal et al. [10], [11], used discrete states to enable or disable measurements like switch variables, but encapsulated their assignment in belief. By explicitly splitting posterior belief and optimal variable assignment computations, a best fit solution is found by searching local to the initialized state in the Bayes tree. The work suggests an underlying synergy between ambiguity through belief and consensus amongst multiple hypotheses.

Furthermore, other near parametric solutions use summarizing statistics, such as the expectation propagation algorithm [12], by propagating an estimate of the first two or three statistical moments of the posterior. This option may not be well suited for drawing accurate metric solutions from multimodal distributions. Two modes imply two different solutions are possible, but the mean of such a distribution would collapses the two possibilities into a single parametric value.

In turn, loopy belief propagation (BP) transmits belief directly on the factor graph [13]. Loopy belief propagation requires repeated iteration across the entire network, in a seemingly random pattern, with little guarantee that an acceptable solution will emerge. Guarantees that the correct modes and covariances can be recovered from graphs with cycles are difficult to find.

The work of Kuehnel [14] investigated Bayesian sampling and inference techniques which focus on posterior estimation for structure from motion, and points to the other major alternative of non-parametric representations. Kernel density estimators (KDEs) [15] use either Dirac delta kernels (particles) or piecewise smooth kernels such as Gaussian, Laplace or Epanetchnikov to represent variable belief. Our work builds on approximating the true posterior through kernel density estimation.

III. DEVELOPMENT AND THEORY

By propagating belief on the Bayes tree, we are able to encapsulate the multi-hypothesis solution in one computational methodology. Although exact Bayesian inference on any tree remains exponentially complicated, inference over the full joint may be avoided by exploiting sparsity in its structure.

A. Factor Graph to Bayes Tree

We use bipartite factor graphs [16] as a graphical model to describe variables of interest through constraint relations originating from sensor measurement likelihood functions. Fig. 2 shows an example factor graph, with seven pose locations x_i, a new landmark sighting l_1 and introduction of prior knowledge through unary constraints. In general, a factor graph encodes the joint probability function relating all state variables Θ and sensor measurements Z:

$$[\Theta \mid Z] = p(\Theta \mid Z) \propto \prod_i^N \varphi_i(\theta_i, z_i) \prod_j \psi_j(\theta_j) \quad (1)$$

If we can compute the posterior, we can find point variable assignments at the most probable location: $\theta_* = \mathrm{argmax}_\theta\, p(\Theta|Z)$, also known as maximum posteriori (MAP) estimates.

A general factor graph contains cycles making inference more difficult. We can discover cliques in cyclic factor graphs through variable elimination, and construct an acyclic tree model representing the same re-factored system. We follow the Bayes tree algorithm [2] which successfully handles cycles in a near optimal manner. The Bayes tree refactorization of our example factor graph is also shown in fig. 2.
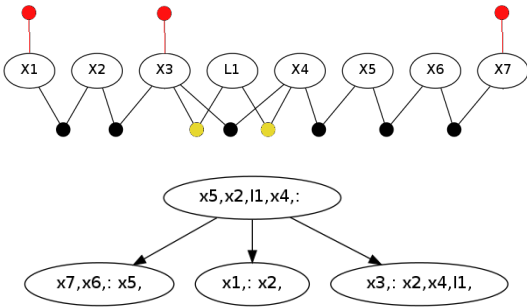
Fig. 2. Top, an example factor graph with seven poses X_i and one feature of interest L_1. Solid disks represent independent measurements from different sensor types. The associated Bayes tree for elimination order $x_3, x_4, x_1, x_6, l_1, x_7, x_2, x_5$ is shown below.

The belief over all variables can now be computed over all r cliques using the tree factorization instead:

$$[\Theta|Z] \propto [\theta_{fr,0} \,|\, z_0] \prod_{k=1}^{r} [\theta_{fr,k} \,|\, \theta_{sp,k}, z_k]. \tag{2}$$

Each clique k consists of frontal variables $\theta_{fr,k}$, separator variables $\theta_{sp,k}$ and likelihood of sensor measurements z_k. Here $k = 0$ denotes the root clique.

The re-factorization is absolutely fundamental to find a computationally tractable solution of the full joint density. The heuristics used to construct the Bayes tree are special in that they minimize modifications to the tree when the underlying factor graph is modified. Modifications to the factor graph are common in simultaneous localization and mapping (SLAM)—as the system accrues more information over time. Aside, this aspect has not been heavily considered in other main stream Bayesian inference approaches.

B. Approximate Belief Propagation

Reminiscent of previous particle methods in SLAM, we combine nonparametric representations with the Bayes tree re-factorization to encapsulate multi-hypothesis inference. We can use the symbolic structure of the Bayes tree to represent the information flow of dominant modes in the original graph. The conditional independence represented by the Bayes tree factorization allows belief propagation (BP) to consider only local interactions at each clique. Each clique in the tree represents a posterior over the local frontal and child clique variables. This local encoding has much lower dimension than the full posterior distribution.

We can compute the partial posterior at each clique by passing the required marginals first up, and then back down, the tree. These posteriors and marginals encapsulate the consensus from available measurements by maintaining dominant modes and simultaneously eliminating low likelihood modes through stochastic approximation. Information, in the form of belief messages, is passed up from the leaves to the root clique. The problem is essentially solved at the root, and all marginals can then be recovered by passing messages back down the tree.

The crux computation in belief propagation is efficiently approximating the posterior density at each clique from multiple incoming child messages and other local potentials. Once we have the local posterior, the independent variables are marginalized out before sending a new message to the parent clique. This process is also referred to as the Chapman-Komolgorov transit integral:

$$p(y) = [Y] = \int_{\mathbb{R}} p(x|y) p(y) \, dx \tag{3}$$

This problem has been intensely studied by the statistics community at large [17], [18], [19], [20] and is the subject of ongoing research.

C. Kernel Densities and Their Product

We use kernel density estimation [15] to approximate non-Gaussian and multi-modal belief in each of the continuous state variables:

$$\left[\hat{X}\right] = \sum_{i=1}^{N} w^{(i)} \mathcal{N}\left(x; x^{(i)}, \Lambda^{(i)}\right). \tag{4}$$

where $\left[\hat{X}\right]$ is the estimated probability density over X; $x^{(i)}$ is the center of, and $w^{(i)}$ the weighting factor, for each of the N kernels. Note that our intent is to use evenly weighted particles throughout the Bayes tree belief propagation update procedure. Lastly, Λ is a bandwidth parameter that determines the smoothness of the approximated belief.

We use leave-one-out likelihood cross validation (LOOCV) of Silverman [15] to compute an appropriate bandwidth for the KDE. We have found that in the multi-modal case many of the heuristic methods drastically oversmooth the bandwidth estimates. LOOCV produces reliable bandwidth estimates with a minimum of ~ 75 samples per mode in each dimension.

1) Multiscale Product Approximation: Products and convolutions are the core operations in Bayesian computation. Sudderth and Ihler [21] developed an efficient method of directly computing the product between two KDEs through multiscale Gibbs sampling. Their method employs a KD-tree representation over the kernels used in the density approximation. The product is computed using Gibbs sampling, but iterations are staged across increasingly finer and finer scales to promote sampling from all modes. We refer the reader to their work at [21], [22].

We use the multiscale product computation to estimate conditional beliefs within each clique as we progress in the posterior approximation process. The original work of Sudderth and Ihler explicitly avoids belief propagation on trees; they argue that the increased dimensionality within cliques of a tree produce an exponential increase in complexity. While this remains true, the Bayes tree performs well in finding near optimal variable clustering, which minimizes the clique sizes.

D. Algorithmic Goal

We are looking for the partial posterior to all clique variables given local potentials and incoming message information. We assume that an initial, and consistent, state for all beliefs is available. The local belief propagation step at each clique during the upward (forward) direction is depicted in fig. 1. We seek to find the partial joint distribution (partial posterior) over variables θ of clique C. We note that frontal variables only contain information from this and cliques lower down in the tree, denoted by Y_z:

$$p_C = [\theta \mid Y_z] \qquad (5)$$

where z implies all measurements from the current and lower cliques of this branch in the tree. The root clique represents the last step in the upward pass. The marginal posterior beliefs for all variables are recovered during the downward belief propagation pass.

Since the exact partial posterior at each clique $[\theta \mid z]$ has exponential complexity, a practical algorithm must make an approximation at some point during its computation. We denote an approximated belief as:

$$\hat{p}_C{}^{(i)} = \left[\hat{\theta} \mid Y_z \right]^{(i)} \qquad (6)$$

at the ith iteration of some approximating process.

Finally, we want the accuracy of the approximation process to improve based on some tunable parameter $\mathcal{Q}$. Therefore the error (measured by some distance or divergence) of approximated belief from the exact belief should decrease as computational load is increased via parameter $\mathcal{Q}$,

$$D\left(p \,\|\, \hat{p}_{\mathcal{Q}}\right) < D\left(p \,\|\, \hat{p}_{\mathcal{Q}'}\right) \qquad \mathcal{Q}' \prec \mathcal{Q}. \qquad (7)$$

And since probability distributions integrate to one, we are in the realm of convergent but infinite vector spaces.

E. A Possible Solution: Assuming Fixed Points

One approach, inspired by works such as [1], [20], might involve an iterative procedure for approximating conditional beliefs over clique variables. Consider that the joint probability of the clique occurs in the space (S, d) such that

$$S : \mathbb{R} \times \mathbb{R} \times \ldots \mathbb{R} \times \mathcal{X} \qquad (8)$$

up to clique variable dimension, along with augmented variables alphabet $\mathcal{X}$, and distance $d\left(f, g\right)$ between distributions. While probabilistic distances are difficult to pin down, we acknowledge the Wasserstein metric as one of the better standard definitions towards satisfying the triangle inequality. In certain cases we can revert to using non-symmetric Kullback-Leibler divergence as substitute for checking progression of some algorithm, $D_{KL}\left(f\|g\right)$, i.e. we have:

$$d : S \times S \to \mathbb{R}. \qquad (9)$$

Taking the current belief estimate of the joint partial posterior as a function $g_i\left(\theta\right) = \hat{p}_C{}^{(i)}\left(\theta \mid y_{z:}\right)$, we aim to

develop a transformation operator T to modify the belief approximation towards the true belief:

$$g_{i+1}\left(\theta\right) = \left(T g_i\right)\left(\theta\right) \qquad (10)$$
$$\text{s.t.} \qquad D\left(p_C \| g_{i+1}\right) < D\left(p_C \| g_i\right). \qquad (11)$$

Furthermore, if we know that a fixed point solution exists,

$$\left(T g_*\right)\left(\theta\right) = g_*\left(\theta\right), \qquad (12)$$

which approaches the true belief $g_* \to p_C$ as $\mathcal{Q} \to \mathcal{Q}_{sup}$, we can focus our algorithmic development on a generic operator $T_{\mathcal{Q}}$. For such an approach to be reliable, we would also have to show the operator T to be a contraction mapping:

$$T : S \to S, \qquad (13)$$
$$D\left(\left(T p_C\right) \| \left(T g_i\right)\right) < D\left(p_C \| g_i\right) \qquad \forall g, p_C \in S, \qquad (14)$$

along with a necessary condition that $\left(T^\infty p_C\right) = p_C$. A more practical version is bounded approximation error i.e. $D\left(p_C \| \left(T_{\mathcal{Q}}{}^k p_C\right)\right) \le \varepsilon, \quad \forall k \in \mathbb{N}$. If the above holds, we further expect some relation between the operator fixed point and detailed balance used in MCMC.

An empirical convergence test involves monitoring the belief of each variable in the clique. The chain is assumed to have converged to its stationary distribution when the KL-divergence between consecutive approximated marginal beliefs drops below the operator quality $\mathcal{Q}$ bound: $D_{KL}\left(g_i'\left(\theta_j\right) \| g_{i+1}'\left(\theta_j\right)\right) \le \varepsilon_{\mathcal{Q}}$. We refer the interested reader to the analysis of Tanner and Wong [1].

F. Imputation with Gibbs Sampling Algorithm

We propose a Markov chain approach to draw samples from the partial posterior of each clique. The samples are used to produce new approximations of variable beliefs at each iteration of the chain. The intuition is that samples will iteratively pass through implicit constraint equations within each clique.

To construct the Markov chain with Gibbs sampling, we require samples from the conditional distributions of each variable in a clique. Consider a clique with frontal variables θ_1, θ_2 and separator variable θ_3. At each iteration, we wish to draw N samples from the approximated conditional distributions:

$$\{\theta_{1,1}, \theta_{1,2}, \ldots, \theta_{1,N}\}^{(i+1)} \sim \left[\hat{\Theta}_1 \mid \Theta_2^{(i)}, \Theta_3^{(i)} \right]$$
$$\{\theta_{2,1}, \theta_{2,2}, \ldots, \theta_{2,N}\}^{(i+1)} \sim \left[\hat{\Theta}_2 \mid \Theta_1^{(i+1)}, \Theta_3^{(i)} \right]$$
$$\{\theta_{3,1}, \theta_{3,2}, \ldots, \theta_{3,N}\}^{(i+1)} \sim \left[\hat{\Theta}_3 \mid \Theta_1^{(i+1)}, \Theta_2^{(i+1)} \right] \qquad (15)$$

In contrast, existing methods generally try to sample a single event from the product of conditional potentials. We construct each conditional distribution by taking the required product between local likelihood, prior constraint and incoming belief message functions involving that variable. In this example, the conditional is

$$\left[\hat{\Theta}_1 \mid \Theta_2^{(i)}, \Theta_3^{(i)} \right] \propto \left[\hat{\Theta}_1 \right] \left[\hat{\Theta}_1 \mid \Theta_2^{(i)} \right] \left[\hat{\Theta}_1 \mid \Theta_3^{(i)} \right]. \qquad (16)$$

The multiscale Gibbs sampling approach [21], presented earlier, approximates the belief over Θ_1 by taking samples from the product of individual elements of the incoming belief densities.

We focus on the belief $\left[\hat{\Theta}_1 \,|\, \Theta_2^{(i)} \right]$, which we compute directly from the sensor measurement likelihood model $\varphi_k \left(\theta_1, \theta_2; z_k \right)$ in the factor graph. We find N values of θ_1 that correspond to each of the N current θ_2 particles:

$$
\begin{aligned}
\text{solve implicit} \quad & \theta_1^{(i)}, \qquad \forall i \in [1, N] \\
s.t. \quad & z_{j,k} \sim \eta_j \left(z, \alpha \right) \\
& 0 = \delta_j \left(\theta_1^{(i)}, \theta_2^{(i)}; z_{j,k} \right)
\end{aligned} \tag{17}
$$

where $\eta_j \left(\cdot \right)$ represents the noise distribution of that constraint function, imposed through the residual δ_j which depends on sensor parameters α. Note that the use of N values of θ_2 constitutes a further approximation and loss of information, since we are essentially approximating the convolution operation at the point. Larger values of N will improve estimation accuracy.

By iterating the chain expressed in eqs. (15) and (17), we obtain a series of samples from the approximated clique joint density. Conditional function estimates are themselves improved by enforcing all local constraint potentials and incoming message information as the chain progresses.

The combination of tree re-factorization, and elimination of low probability modes and approximating the convolution through implicit equation solving is our mechanism for reducing computational complexity.

Modeling multi-modal belief: The last component we need is how to model data association uncertainty. While several new avenues for modeling measurement uncertainty are possible with a non-parametric solution, we show a multi-modal loop closure where the likelihood function represents a measurement originating from one of several ambiguous landmarks:

$$
z_j \sim \left[Z \,|\, \Gamma, \Theta_j \right], \qquad \gamma \sim \left[\Gamma \right]_\rho = Cat \left(\rho \right), \tag{18}
$$

where γ represents the data association selection based on some weighting ρ. Our formulation differs from previous methods in that the discrete variable is marginalized out:

$$
\left[Z \,|\, \Theta_j \right] = \int \left[Z \,|\, \gamma, \Theta_j \right] \left[\gamma \right]_\rho \, d\gamma. \tag{19}
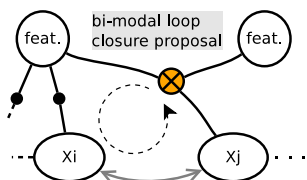$$

Fig. 3. Example of bi-modal loop closure proposal, modeled directly in the factor graph with poses X and ambiguous data association for measurement from a feature sighting.

IV. RESULTS

We implemented the proposed algorithm in a new programming language called Julia [23], for its efficient functional syntax, high performance, and parallel computing capabilities. We explain the process in more detail with a modified robot door sighting localization example from [24], chapter 8. Thereafter we apply the algorithm to the familiar, and larger scale, Victoria Park dataset.

A. Illustrative Example

Fig. 4 illustrates the localization problem where a robot is driving and sensing familiar landmarks. The robot instantiates poses at points of interest along its trajectory that will form the backbone of the inference task. Please see fig. 2 for the associated factor graph, as well as its re-factorization into a Bayes tree.
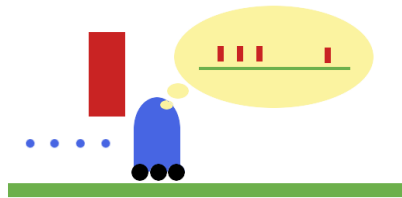

Fig. 4. Pictorial of a robot moving in a one dimensional world, modified from original example in [24].

Given a prior map containing four landmarks at $z_l \in \{-100, 0, 100, 300\}$, the robot believes it must be at one of these four locations when a sighting is made. We can represent these four different landmark positions hypotheses as four modes in the associated measurement factor:

$$
\psi_l \left(x \right) = \left[\hat{X} \,|\, z_l \right] = \frac{1}{4} \sum_k^{\{a:d\}} \mathcal{N} \left(x; \, \mu_k, \, \sigma^2 \right) \tag{20}
$$

1) First Three Poses: After the first landmark sighting at pose x_1, the robot moves forward 50 units to new pose x_2. After a further 50 units, a second landmark sighting is made at pose x_3. The odometry measurement is modeled with a Gaussian likelihood function

$$
\begin{aligned}
z_k \sim \left[Z_{odo} \,|\, X_i, X_{i+1} \right] &= \mathcal{N} \left(\mu = f_{odo} \left(\cdot \right), {\sigma_j}^2 \right) \\
f_{odo} \left(x_i, x_{i+1} \right) &= x_{i+1} \ominus x_i
\end{aligned} \tag{21}
$$

where $\ominus$ is used to denote the difference on the manifold, which is important for generalization to higher dimensions. The constraint functions $\varphi_j \left(x_{i+1}, x_i \right)$, indicated by solid filled factor vertexes in fig. 5, are denoted as

$$
\varphi_j \left(x_i, x_{i+1}; \sigma_j \right) \propto \left[Z_{odo} \,|\, X_i, X_{i+1} \right].
$$

Finally, we construct the proposal sampling function as a residual for use with eq. (17)

$$
0 = \delta_j \left(x_{i+1}^{(k)}, x_i^{(k)}; z_k \right) = z_k - f_{odo} \left(x_i, x_{i+1} \right). \tag{22}
$$

The associated factor graph and Bayes tree at this point is shown in fig. 5. Note multimodal landmark sightings, $\psi_l \left(x \right)$, are shown as red unary factors.
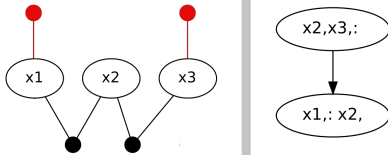
Fig. 5. A multi-hypothesis factor graph and Bayes tree representing 100 units driving distance and two independent sightings of four indistinguishable but known landmarks.

We know that the first three landmarks are each separated by 100 odometry units. The exact belief over the three poses, x_1, x_2 and x_3, therefore have two major hypotheses. Fig. 6 shows the exact and estimated full marginal beliefs over x_1, x_2 and x_3 in red and black, respectively. While the exact solution contains $4^3 = 64$ modes, only the two shown in fig. 6 are significant. This is a key aspect to reducing computational complexity in multi-hypothesis tracking.
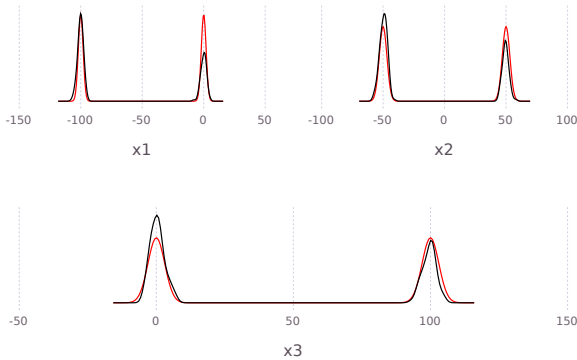

Fig. 6. Estimated marginal beliefs, in black, of all variables of factor graph in fig. 5 following one complete up and down belief propagation pass of the the Bayes tree. Ground truth belief is shown in red. Notice how the posterior has two significant modes as the final output of the inference process—a key difference from parametric SLAM solutions.

We compute the estimated belief using the MCMC chain methodology described in section III-F. Starting with the upward pass of belief over the Bayes tree, as shown in fig. 5. The joint probability of the leaf clique, which contains one odometry and a landmark sighting factor, is computed with Gibbs sampling:

$$\{x_{1,1}, \ldots, x_{1,N}\}^{(i+1)} \sim \left[\hat{X}_1 \mid X_2^{(i)}, z_{odo}\right] \left[X_1 \mid z_l\right]$$
$$\{x_{2,1}, \ldots, x_{2,N}\}^{(i+1)} \sim \left[\hat{X}_2 \mid X_1^{(i+1)}, z_{odo}\right]. \quad (23)$$

The transit conditionals were obtained through uniform prior assumption on X_2, and observation z_{odo}:

$$\left[\hat{X}_1 \mid X_2^{(i)}, z_{odo}\right] \propto \left[z_{odo} \mid X_1, X_2\right]\left[X_2\right], \quad (24)$$

since we only have a uniform prior on each of the sensor measurement terms. We approximate (24) using eqs. (22) and (17) to construct the kernel density estimate with LOOCV

bandwidth

$$\left[\hat{X}_1 \mid x_2^{(1:N)}, z_2\right] \propto \frac{1}{N} \sum_{i=1}^{N} \mathcal{N}\left(X_1; x_1^{(i)}, \Lambda_{loocv}\right). \quad (25)$$

The second potential function on X_1 is the multimodal prior defined by eq. (20). The belief product is estimated with the multi-scale product algorithm from section III-C.1.

Frontal variables of the clique are permanently stored as the MCMC chain progresses, while separator values are only cached. Once the MCMC chain of the leaf clique has converged, we can construct the outgoing belief message from the latest cached belief over separator variable, $\hat{m}(x_2) = [X_2 \mid z_{odo}, z_l]$. This message is sent to the root as per usual belief propagation.

Once the algorithm completes at the root, the process is repeated and new belief messages are passed back down the tree to recover all marginal beliefs. All estimated marginal beliefs are shown as black traces in fig. 6.

Note that the correct mode in each case may have a lower probability, and is an artifact of the approximations used. A sum-product belief propagation (as we use here) operates correctly with the lower probability mode and is therefore not a concern as would be in a max-product style approach. Max-product style algorithms would have selected the dominant mode at this point and failed.

2) A Third Landmark Sighting: Now the robot drives 200 more units through four more pose positions, to pose x_7, where a third and final familiar landmark sighting is made. While passing through poses x_3 and x_4, sightings of a new feature of interest, l_1, is made. New feature l_1 is included in the factor graph also. The complete system factor graph and tree was shown earlier in fig. 2.

The previous leaf clique $x_1 : x_2$ can be directly recycled, as per [2]. Two new leaf, and a root clique are formed, each with their own MCMC chains.
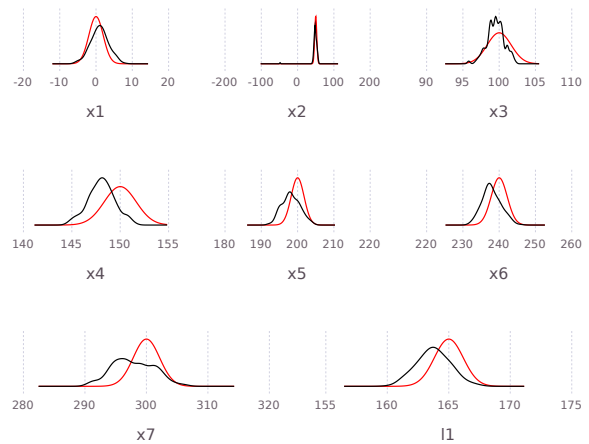

Fig. 7. Estimated marginal belief of all variables, as black traces, following one up and down pass of the Bayes tree in fig. 2. Baseline marginals, shown in red, computed via standard forward-backward solution over equivalent Hidden Markov Model solution. Note the exact marginal of X_1 can have up to 1024 modes, but only one significant mode at 0 corresponds to ground truth.

We note separate branches are known to be be independent and can be computed in parallel. The estimated and baseline marginals for all variables in the factor graph, following a single upward and downward belief propagation pass of the Bayes tree, is shown in fig. 7. Notice how consensus has been reached, showing a single mode in the posterior belief.

B. Victoria Park

The Victoria Park dataset consists of a car with planar laser scanner and automated tree trunk detection algorithm. We processed the first $23\ mins$ of driving and raw detection data without any known data association amongst tree sightings. We wish to recover the vehicle trajectory and local map of trees. The factor graph contains around 3800 variable dimensions and more than 8500 constraint dimensions. Furthermore, the problem contains 482 bi-modal loop closure proposal landmarks, which account for around 1700 bimodal constraint functions in the factor graph. Needless to say, the theoretic number of modes in the system is very large (greater than 2^{482}) and our focus is to estimate the much smaller number of prominent modes.

The blue trace in fig. 8 represents the maximum a posteriori point estimate extracted from variable beliefs after multimodal posterior estimation has completed (MM-MAP). The magenta trace represents an equivalent maximum likelihood estimate (MLE, iSAM1) with all loop closures as classical unimodal constraints. Recovering all marginal beliefs for this size problem with the current implementation, *IncrementalInference.jl* available at [25], takes around 3 hours on a dual Xeon, $64\ Gb$ RAM computer, utilizing around 6 processing cores on average. Our implementation has room for several significant speed-ups, and we expect to better exploit multi-threaded operations for in-clique operations in future versions.

Fig. 8 shows how MAP estimates remain consistent in the presence of significant false proposals. Short of a few poses where an alternate mode happens to be "stronger"—further inference passes over the tree, or addition of new data will most likely emphasize the true mode. It is critical to remember that the multi-modal solution, as used in this example, does not throw out any information; we are not allowing null-hypothesis to happen. We hope to illustrate that each measurement is trusted, but their association to previous landmarks in the map are uncertain. Here, data association happens at inference time without modifying the structure of the graph in any way. This allows us to exploit the full structure of the problem throughout all *back-end* processes.

V. DISCUSSION

Previous histogram style Bayesian inference approaches, such as in [24], approximate solutions in a finite, discretized and fixed world grid. The proposed approach here aims to work freely on the infinite continuous set.

The statistics community have long developed several tree algorithms—analogues to the Bayes tree—for re-factoring any factor graph into a more computationally tractable form. Clique trees, cluster trees, rake-and-compress trees and most
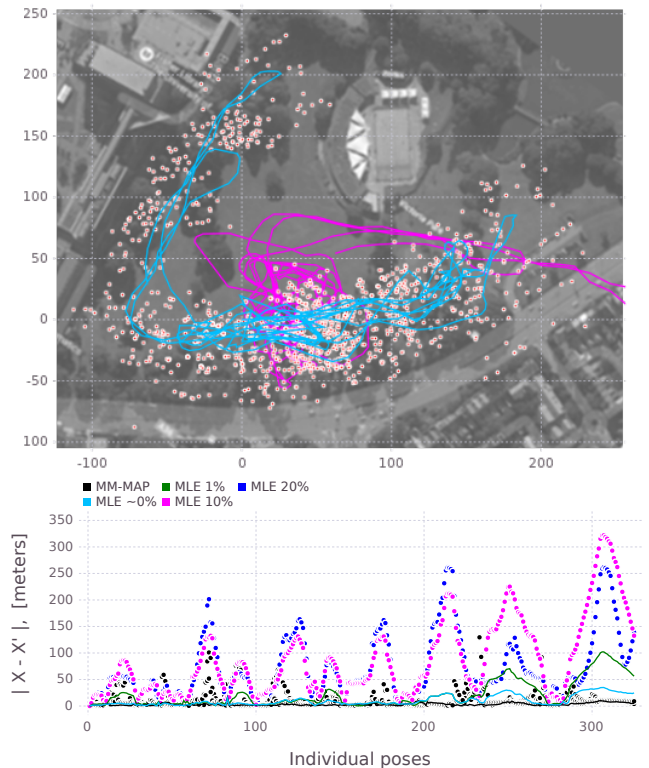

Fig. 8. Top: The blue trace shows Victoria park $argmax\ p\left(\Theta|\mathbf{Z}\right)$ point estimate, with 10% erroneous loop closure proposals. The magenta trace shows the same result for naive maximum likelihood estimate. Alternative modes, which were calculated, are difficult to visualize and not shown here. Bottom: shows the distance between similar poses, using MM-MAP 0% estimate as baseline, with varying levels of loop closure proposal corruption: 1%, 10% and 20% corruption. See legend for colors. The large values correspond to an erroneous unimodal MLE equivalent. Notice the black trace and dots representing equivalent MM-MAP estimates (for 1%, 10% and 20% corruption cases) which have much smaller errors. Keep in mind that in this example corrupted data implies less loop closures are in effect, reducing overall accuracy. Suspect minor error in baseline data association (visible in $\sim 0\%$ MLE), since 1%, 10% and 20% black traces are repeatable.

notably the junction tree, all re-factorize a "loopy" factor graph into an acyclic representation [26]. Most of the existing work consider trees much smaller than typically found in SLAM problems.

We have chosen to approximate the posterior densities of each clique in the Bayes tree with a set of samples. These samples can be used to construct belief propagation messages for inference across the entire tree. We limit the number of samples according to available computational resources and thereby concentrate computation around the prominent modes in the system.

The sample approach simultaneously introduces loss of information. Loss of information is a vital aspect in obtaining a computationally tractable algorithm and corresponds to discarding low likelihood modes, but may lose track of smaller modes which remain important. Fortunately, if new information is added the previously lost modes may be "revived" across an updated tree.

The key advantage of our approach is that when more information becomes available, the inference on the clique

is repeated and a new set of samples are generated which may now focus computation on an altered subset of dominant modes. Thereby data association is deferred into the back-end inference process and assignment is available through individual variable beliefs rather than factor graph structure.

Ambiguity in measurement data association can be modeled as multi-modal belief between multiple variables in the system. This would allow back-end solution to internally find Bayesian consensus amongst ambiguous data without modifying the structure of the factor graph. This is a vital difference to previous approaches.

The ability to process minimum incremental updates to the Bayes tree, which may now include multi-modality, is highly desirable and part of our future work efforts. Furthermore, the Bayes tree symbolically encodes the multi-hypothesis belief of an otherwise obscure variable interaction shown by the factor graph alone. We intend to further exploit the Bayes tree structure by combining classic unimodal parametric methods for portions of the tree which involve only uni-modal belief, and switching to nonparametric methods when multimodality is encountered. Finally, recent work on the connection between reproducing Hilbert space embeddings and conditional distributions [27], [28] offers a different and potentially more efficient mechanism for approximate Bayesian computations. Further methods, such as progressive Bayes [29] might prove very powerful indeed.

VI. CONCLUSION

A nonparametric solution is more costly to compute, but with increased computational power we can start looking to solutions beyond unimodal parametric Gaussian solvers. The advantage is that we can now start to solve more general problems. For example, factor graphs should not be limited to unimodal belief assumptions. If we are able to capture data association uncertainty in our models, as we have shown here, then we can focus development on solvers which adhere to the factor graph representation as a common language. The proposed method is unique in that data association ambiguity can be modeled in the factor graph and left up to the inference engine to solve. Most existing techniques use heuristics and local approximations to simplify the inference task, while the proposed method seeks to approximate the true Bayesian posterior.

REFERENCES

[1] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American statistical Association*, vol. 82, no. 398, pp. 528–540, 1987.

[2] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The International Journal of Robotics Research*, vol. 31, pp. 217–236, Feb. 2012.

[3] Y. Latif, C. Cadena, and J. Neira, "Robust loop closing over time for pose graph SLAM," *The International Journal of Robotics Research*, p. 0278364913498910, 2013.

[4] M. C. Graham, J. P. How, and D. E. Gustafson, "Robust incremental SLAM with consistency-checking," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2015.

[5] N. Sünderhauf and P. Protzel, "Switchable constraints for robust pose graph SLAM," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2012.

[6] E. Olson and P. Agarwal, "Inference on networks of mixtures for robust robot mapping," *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 826–840, 2013.

[7] G. Huang, M. Kaess, J. Leonard, and S. Roumeliotis, "Analytically-selected multi- hypothesis incremental MAP estimation," in *International Conference on Acoustics, Speech, and Signal Processing*, British Columbia, Canada, May 2013.

[8] E. Olson and P. Agrawal, "Inference on networks of mixtures for robust robot mapping," in *Robotics: Science and Systems (RSS)*, Jul. 2012.

[9] S. Thrun, M. Montemerlo, D. Koller, B. Wegbreit, J. Nieto, and E. Nebot, "FastSLAM: An efficient solution to the simultaneous localization and mapping problem with unknown data association," *Journal of Machine Learning Research*, vol. 4, no. 3, pp. 380–407, 2004.

[10] A. V. Segal and I. D. Reid, "Hybrid inference optimization for robust pose graph estimation," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 2675–2682.

[11] A. V. Segal, "Iterative local model selection for tracking and mapping," Ph.D. dissertation, Department of Engineering Science University of Oxford Michaelmas Term 2014 This thesis is submitted to the Department of Engineering Science, University of Oxford, 2015.

[12] T. P. Minka, "Expectation propagation for approximate bayesian inference," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2001, pp. 362–369.

[13] A. Ranganathan, M. Kaess, and F. Dellaert, "Loopy SAM," in *Intl. Joint Conf. on Artificial Intelligence*, Hyderabad, India, 2007, pp. 2191–2196.

[14] F. O. Kuehnel, "Robust bayesian estimation of nonlinear parameters on se (3) lie group," *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 735, pp. 176–186, 2004.

[15] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.

[16] 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.

[17] L. Tierney, "Markov chains for exploring posterior distributions," *the Annals of Statistics*, pp. 1701–1728, 1994.

[18] R. M. Neal, "Sampling from multimodal distributions using tempered transitions," *Statistics and computing*, vol. 6, no. 4, pp. 353–366, 1996.

[19] G. Celeux, M. Hurn, and C. P. Robert, "Computational and inferential difficulties with mixture posterior distributions," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 957–970, 2000.

[20] A. E. Gelfand and A. F. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American statistical association*, vol. 85, no. 410, pp. 398–409, 1990.

[21] 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.

[22] 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.

[23] J. Bezanson, S. Karpinski, V. B. Shah, and A. Edelman, "Julia: A fast dynamic language for technical computing," *arXiv preprint arXiv:1209.5145*, 2012.

[24] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, Cambridge, MA, 2005.

[25] D. Fourie, "Incrementalinference.jl," 2016. [Online]. Available: https://github.com/dehann/IncrementalInference.jl

[26] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, Cambridge, MA, 2009.

[27] 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.

[28] A. Beskos, F. J. Pinski, J. M. Sanz-Serna, and A. M. Stuart, "Hybrid monte carlo on hilbert spaces," *Stochastic Processes and their Applications*, vol. 121, no. 10, pp. 2201–2230, 2011.

[29] O. C. Schrempf, O. Feiermann, and U. D. Hanebeck, "Optimal mixture approximation of the product of mixtures," in *Information Fusion, 2005 8th International Conference on*, vol. 1. IEEE, 2005, pp. 8–pp.