PF-EIS & PF-MT: NEW PARTICLE FILTERING ALGORITHMS FOR MULTIMODAL OBSERVATION LIKELIHOODS AND LARGE DIMENSIONAL STATE SPACES

Namrata Vaswani

Dept. of ECE, Iowa State University, Ames, IA 50011, namrata@iastate.edu

ABSTRACT

Consider tracking a state space model with multimodal observation likelihoods using a particle filter (PF). Under certain assumptions that imply narrowness of the state transition prior, many efficient importance sampling techniques have been proposed in literature. For large dimensional state spaces (LDSS), these assumptions may not always hold. But, it is usually true that at a given time, state change in all except a few dimensions is small. We use this fact to design a simple modification (PF-EIS) of an existing importance sampling technique. Also, importance sampling on an LDSS is expensive (requires large number of particles, N) even with the best technique. But if the "residual space" variance is small enough, we can replace importance sampling in residual space by Mode Tracking (PF-MT). This drastically reduces the importance sampling dimension for LDSS, hence greatly reducing the required N.

Index Terms: particle filter, mode tracking, importance sampling, Monte Carlo methods, sensor networks.

1. INTRODUCTION

Tracking is the problem of causally estimating a hidden state sequence, $\{X_t\}$, from a sequence of observations, $\{Y_t\}$ that satisfy the Hidden Markov Model assumption, i.e. $X_t \rightarrow Y_t$ is a Markov chain for each t, with observation likelihood (OL), $p(Y_t|X_t)$; and $X_{t-1} \rightarrow X_t$ is also Markov with state transition pdf (STP), $p(X_t|X_{t-1})$. The posterior $p(X_t|Y_{1:t}) \triangleq \pi_t(X_t)$. For nonlinear and/or nonGaussian state space models, π_t can be efficiently approximated using a particle filter (PF) [1, 2, 3]. One of two main issues in PF design is the choice of importance sampling density that reduces the variance of importance weights (improves effective particle size)[2].

The most commonly used importance sampling density is the STP, $p(X_t|X_{t-1}^i)$ [1](assumes nothing). But since this does not use knowledge of Y_t, the weight variance can be large. For situations where the OL is multimodal, but the STP is unimodal and narrow enough to ensure that $p^*(X_t) \triangleq p(X_t|X_{t-1}^i, Y_t)$ is unimodal, [4] proposes to approximate p^* by a Gaussian at its mode and importance sample from it. Other solutions that also assume p^* is unimodal are [2, 5, 6]. In many situations, p^* may be multimodal but conditioned on a small part of the state space, denoted $X_{t,s}$, it is unimodal (Assumption 1). When this holds, we propose to modify Doucet [4]'s method as follows. Let $X_t = [X_{t,s}, X_{t,r}]$. Sample $X_{t,s}$ from its STP but compute a Gaussian approximation to $p^*(X_t|X_{t,s}^i) = p^*(X_{t,r}|X_{t,s}^i)$ about its mode and importance sample $X_{t,r}$ from it. We refer to this idea as PF-EIS (Algorithm 1).

For large dimensional state spaces (LDSS), which have dimension more than 10 or 12, the number of particles required for reasonable accuracy is very large [1, 7] and this makes PF an impractical algorithm. One class of techniques for LDSS is [3, Ch 13],[8] which resample more than once within a time interval. Alternatively, if conditioned on a small part of the state ($X_{1:t,s}$), the rest ($X_{t,r}$) has a linear Gaussian state space model, Rao Blackwellization (RB-PF) [9, 7] can be used. Now, this assumption may not always hold. But, in most large dimensional systems, at any given time, "most of the state change" occurs in a small number of dimensions ("effective basis") while the change in the rest of the state space ("residual space") is "small" [10, 11]. If the variance of residual state change is "small enough" so that Theorem 1 is applicable, Assumption 1 will hold. In addition, if it is even "smaller" to ensure that Theorem 2 holds with a small enough ϵ, the importance sampling of $X_{t,r}$ can be replaced by Mode Tracking (MT). We call this idea PF-MT (Algorithm 2).

MT reduces the importance sampling dimension from $\dim(X_t)$ to $\dim(X_{t,s})$ (huge reduction for large dimensional problems). Of course, the error in the estimate of $X_{t,r}$ will also increase. But for 200-250 dim problems such as contour tracking [12, 13], this error is more than offset by the reduction in error due to improved effective particle size. Note that PF-MT is a generalization of the contour tracking idea of [13] which was first generalized in [10, 11] and used in [12]. It can also be understood as an approximate RB-PF [9, 7].

Some example applications are as follows. (i) When there are two different types of sensors tracking temperature at one location, each with some probability of failure, OL will be bimodal if one of them fails (Example 1). When tracking temperature at a large number of nodes in a sensor network, the state space dimension can be very large and also the number of possible OL modes can be very large. For the situation discussed in Example 2, Theorem 1 applies with $\Delta^* = \infty$. In other situations, Theorem 1 may still apply with a finite Δ^* if temperature change at a subsampled set of sensor nodes is used as the effective basis. (ii) In visual tracking problems such as deforming contour tracking [12, 13, 8] or tracking illumination change of moving objects [14], OL is multimodal (due to multiple objects, occlusions or clutter) and state space dimension is large.

Note PF-EIS or PF-MT will still work if Theorem 1 applies most of the time. Also, if system model changes with time, effective basis dimension can be changed over time. Also, note that PF-EIS is also applicable to smaller dimensional problems and PF-MT is also useful in situations where p^* is actually unimodal.

Organization: In Sec. 2, we explain PF-EIS, give sufficient conditions for Assumption 1 to hold for an LDSS model and show how to verify these. PF-MT is explained in Sec. 3. Comparisons with existing PF methods and discussion are given in Sec. 4.

2. PF-EIS: PF-EFFICIENT IMPORTANCE SAMPLING

The "optimal" importance sampling density, i.e. one that minimizes the conditional variance of weights is [4] $p(X_t|X_{t-1}^i, Y_t) \triangleq p^*(X_t)$. In most cases, this cannot be computed analytically. [4] suggests approximating p^* by a Gaussian about its mode, *when p^* is unimodal*. But when OL is multimodal, p^* will be unimodal only if the STP is narrow enough in at least some dimensions. When p^* is multimodal, we propose the following modification. Split the state vector X_t as $X_t = [X_{t,s}, X_{t,r}]$ so that variance of $X_{t,r}$ is small enough s.t.

Assumption 1 *Conditioned on $X_{t,s}$, p^* is unimodal, i.e.*

$$p^{**,i}(X_{t,r}) \triangleq p^*(X_t|X_{t,s}^i) = p(X_{t,r}|X_{t-1}^i, X_{t,s}^i, Y_t) \textit{ is unimodal}$$

Algorithm 1 PF-EIS. Going from π_{t-1}^N to $\pi_t^N(X_t) = \sum_{i=1}^N w_t^{(i)} \delta(X_t - X_t^i)$, $X_t^i = [X_{t,s}^i, X_{t,r}^i]$

1. *Importance Sample $X_{t,s}$:* $\forall i$, sample $X_{t,s}^i \sim p(X_{t,s}^i | X_{t-1}^i)$.
2. *Importance Sample $X_{t,r}$:* $\forall i$, sample $X_{t,r}^i \sim \mathcal{N}(X_{t,r}^i; m_t^i, \Sigma_{IS}^i)$. Here $m_t^i(X_{t-1}^i, X_{t,s}^i, Y_t) = \arg\min_{X_{t,r}} L^i(X_{t,r})$ and $\Sigma_{IS}^i \triangleq (\nabla^2 L^i(m_t^i))^{-1}$ where $L^i(X_{t,r}) \triangleq -\log[p^{**,i}(X_{t,r})] = -\log[p(X_{t,r}|X_{t-1}^i, X_{t,s}^i, Y_t)]$.
3. *Weight & Resample:* Compute $w_t^i = \frac{\tilde{w}_t^i}{\sum_{j=1}^N \tilde{w}_t^j}$ where $\tilde{w}_t^i = w_{t-1}^i \frac{p(Y_t|X_t^i) p(X_{t,r}^i | X_{t-1}^i, X_{t,s}^i)}{\mathcal{N}(X_{t,r}^i;\, m_t^i,\, \Sigma_{IS}^i)}$ & resample. $t \leftarrow t+1$ & go to step 1.

Algorithm 2 PF-MT. Going from π_{t-1}^N to $\pi_t^N(X_t) = \sum_{i=1}^N w_t^{(i)} \delta(X_t - X_t^i)$, $X_t^i = [X_{t,s}^i, X_{t,r}^i]$

1. *Importance Sample $X_{t,s}$:* $\forall i$, sample $X_{t,s}^i \sim p(X_{t,s}^i | X_{t-1}^i)$.
2. *Mode Track $X_{t,r}$:* $\forall i$, set $X_{t,r}^i = m_t^i$.
3. *Weight & Resample:* Compute $w_t^i = \frac{\tilde{w}_t^i}{\sum_{j=1}^N \tilde{w}_t^j}$ where $\tilde{w}_t^i = w_{t-1}^i p(Y_t|X_t^i) p(X_{t,r}^i | X_{t-1}^i, X_{t,s}^i)$ & resample. $t \leftarrow t+1$, go to step 1.

When this holds for each particle and for each time, we can use the Gaussian approximation idea of [4] to approximate $p^{**,i}$ and sample from it. In practice, even if it holds for most particles at most times, our proposed algorithm will work. Thus we propose to importance sample (IS) as follows. Select $X_{t,s}$ as the minimum number of dimensions of X_t required to ensure that Assumption 1 holds. Sample $X_{t,s}^i$ from its STP (to sample the possibly multiple modes of p^*). Sample $X_{t,r}^i$ from a Gaussian approximation[4] to $p^{**,i}$ about its mode, i.e. sample $X_{t,r}^i$ from $\mathcal{N}(m_t^i, \Sigma_{IS}^i)$ where

$$m_t^i = m_t^i(X_{t-1}^i, X_{t,s}^i, Y_t) \triangleq \min_{X_{t,r}} L^i(X_{t,r}), \ \ and$$

$$\Sigma_{IS}^i \triangleq [(\nabla^2 L^i)(m_t^i)]^{-1}, \ L^i(X_{t,r}) \triangleq -\log[p^{**,i}(X_{t,r})] + \text{const}$$

$\nabla^2 L^i$ denotes the Hessian of L^i. We refer to the above algorithm as PF with Efficient IS or PF-EIS. It is summarized in Algorithm 1. *For $X_{t,r} = X_t$, Algorithm 1 reduces to Doucet's algorithm [4] and if $X_{t,s} = X_t$, Algorithm 1 reduces to the original PF [1].*

2.1. Unimodality of $p^{**,i}(X_{t,r})$ for LDSS Models

For the LDSS examples of the introduction, the state dynamics can be written in the form of equations (1)-(4) of [10]. It is a generic form of the second order motion model for nonEuclidean state spaces. The quantity C_t (e.g. contour or temperature) has "velocity" (time derivative), v_t, split as $v_t = B_s v_{t,s} + B_r v_{t,r}$ where B_s denotes the effective basis directions and B_r denotes a basis for the residual space. $v_{t,s}, v_{t,r}$ are the corresponding coefficients. For e.g., B_s can be the dominant eigenvectors of the covariance of v_t or it can be an interpolation basis. *Also, effective basis dimension,* $\dim(v_{t,s}) = K$.

If in the LDSS model of [10], C_t belongs to a vector space, we have $g(C_{t-1}, v_t) = v_t$ and $\dim(C_t) = M$. Then it simplifies to:

$$\begin{aligned}
C_t &= C_{t-1} + B_s v_{t,s} + B_r v_{t,r}, \\
v_{t,s} &= f_s(v_{t-1,s}) + \nu_{t,s}, \ \nu_{t,s} \sim \mathcal{N}(0, \Sigma_s), \ \Sigma_s = diag\{\Delta_p\}_{p=1}^K \\
v_{t,r} &= f_r(v_{t-1,r}) + \nu_{t,r}, \nu_{t,r} \sim \mathcal{N}(0, \Sigma_r), \ \Sigma_r = diag\{\Delta_p\}_{p=K+1}^M \\
& p(Y_t|X_t) = p(Y_t|C_t) \triangleq \alpha \exp[-E_{Y_t}(C_t)] \qquad (1)
\end{aligned}$$

Here $X_{t,s} = v_{t,s}$ and $X_{t,r} = [v_{t,r}, C_t]$. For the purpose of sampling, $X_{t,r} = v_{t,r}$ since C_t is a deterministic function of $C_{t-1}, v_{t,s}$, and $v_{t,r}$. Also, for the above model, $p(X_{t,s}|X_{t-1}) = p(X_{t,s}|X_{t-1,s})$ and $m_t^i = m_t^i(X_{t-1,r}^i, X_{t,s}^i, Y_t)$. We obtain sufficient conditions for Assumption 1 for this model and extend them to the model of [10].

For the above model, we have $p^{**,i}(X_{t,r}) = p^{**,i}(v_{t,r}) = p(v_{t,r}|v_{t-1,r}^i, C_{t-1}^i, v_{t,s}^i, Y_t) = p(v_{t,r}|v_{t-1,r}^i, \tilde{C}_t^i, Y_t)$.
Let $f_r(v_{t-1,r}^i) \triangleq f_r^i$ and $C_{t-1}^i + B_s v_{t,s}^i \triangleq \tilde{C}_t^i$. Then,

$$p^{**,i}(v_{t,r}) \propto \exp[-E(\tilde{C}_t^i + B_r v_{t,r})] \, \mathcal{N}(v_{t,r}^i; f_r^i, \Sigma_r) \qquad (2)$$

Thus $L^i(v_{t,r}) = -\log[p^{**,i}(v_{t,r})] + \text{const}$ is

$$L^i(v_{t,r}) = E(\tilde{C}_t^i + B_r v_{t,r}) + \sum_{p=1}^{M-K} \frac{([v_{t,r} - f_r^i)]_p)^2}{2\Delta_{p+K}} \qquad (3)$$

where $[.]_p$ denotes the p^{th} coordinate of a vector. Now, $p^{**,i}$ will be unimodal iff L^i has a unique minimizer. The second term in (3) is strongly convex with a unique minimizer at $v_{t,r} = f_r^i$. But $E(C_t)$ (and hence E as a function of $v_{t,r}$) can have multiple minimizers since OL can be multimodal. If we can ensure that Σ_r is small enough so that L^i has a single minimizer that lies in the neighborhood of $f_r^i = f_r(v_{t-1,r}^i)$, we will be done. This idea leads to:

Theorem 1 (Unimodality) *Denote $f_r(v_{t-1,r}^i) \triangleq f_r^i$ and $C_{t-1}^i + B_s v_{t,s}^i \triangleq \tilde{C}_t^i$. For the model of (1), $p^{**,i}(v_{t,r})$ will be unimodal if*

1. *E is twice differentiable almost everywhere.*
2. *$\tilde{C}_t^i + B_r f_r^i$ is close enough to a minimizer of E so that $E(C)$ is strongly convex in its neighborhood.*
3. *$\Delta_{p+K}, p = 1, 2, \ldots M-K$ satisfy*

$$\inf_{v_{t,r} \in \mathcal{G}} \max_{p=1,\ldots M-K} (\gamma_p(v_{t,r}) - \Delta_{p+K}) > 0,$$
$$\mathcal{G} \triangleq \cap_{p=1}^{M-K} (\mathcal{A}_{K,p} \cup \mathcal{Z}_{K,p}) \qquad (4)$$

$$\gamma_p(v_{t,r}) \triangleq \begin{cases} \frac{|[\nabla D]_p|}{|[\nabla E]_p|}, & v_{t,r} \in \mathcal{A}_{K,p} \\ 0, & v_{t,r} \in \mathcal{Z}_{K,p} \end{cases} \qquad (5)$$

$$\begin{aligned}
\nabla E &\triangleq B_r^T \nabla_C E(\tilde{C}_t^i + B_r v_{t,r}) \\
\nabla D &\triangleq v_{t,r} - f_r^i \qquad (6)
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_{K,p} &\triangleq \{v_{t,r} \in \mathcal{R}_{K,LC}^c : [\nabla D]_p . [\nabla E]_p < 0\}, \\
\mathcal{Z}_{K,p} &\triangleq \{v_{t,r} \in \mathcal{R}_{K,LC}^c : [\nabla E]_p = 0 \ \& \ [\nabla D]_p = 0\}, \\
\mathcal{R}_{K,LC} &\triangleq \{v_{t,r} \in \mathbb{R}^{M-K} : \tilde{C}_t^i + B_r v_{t,r} \in \mathcal{R}_{LC}\}, \qquad (7)
\end{aligned}$$

where $\mathcal{R}_{LC} \subseteq \mathcal{S} = \mathbb{R}^M$ is the largest contiguous region in the neighborhood of $\tilde{C}_t^i + B_r f_r^i$ which contains a minimizer of E and where $E(C)$ is convex. Also, $|.|$ denotes absolute value and $[.]_p$ denotes p^{th} coordinate of a vector.

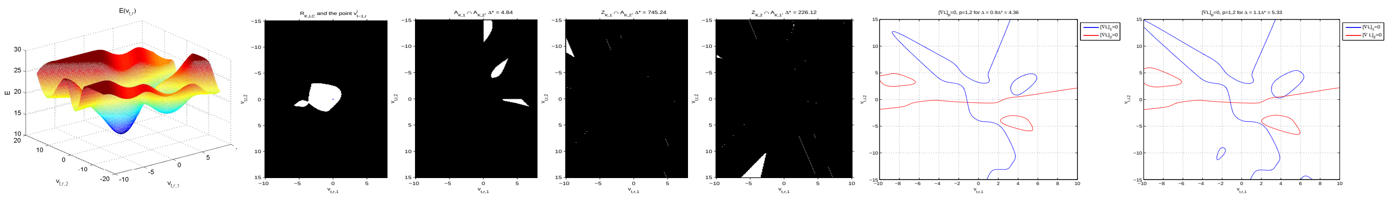


Fig. 1. Computing Δ_K^* for Example 1 ($M = 3, K = 1$). We used $\alpha_1 = 0.1$, $\alpha_2 = 0.4$, $a = 10$, $\sigma_{obs}^2 = 1$, $\Delta_1 = 5.4$, $B_s = [0.64\ -0.56\ 0.53]'$, $B_r = [0.73\ 0.66\ -0.18;\ -0.25\ 0.5\ 0.83]'$ (we use MATLAB notation). Also, $C_{t-1}^i = [0\ 0\ 0]'$, $v_{t-1,r}^i = [0\ 0]'$, $v_{t-1,s}^i = 0$, $Y_t = [6.43\ 1.68\ -3.59\ -2.5\ 1.59\ 1.49]'$ and $v_{t,s}^i = 2.9$ (simulated from $\mathcal{N}(0, \Delta_1)$). Col. 1: mesh plot of E as a function of $v_{t,r}$. Col. 2: $\mathcal{R}_{K,LC}$, note that the point $f_r^i = v_{t-1,r}^i$ lies inside it. Col. 3,4,5: the 3 regions constituting $\mathcal{G}$, $\mathcal{A}_{K,1} \cap \mathcal{A}_{K,2}$, $\mathcal{Z}_{K,1} \cap \mathcal{A}_{K,2}$, $\mathcal{Z}_{K,1} \cap \mathcal{A}_{K,2}$ along with the computed value of Δ^* in the 3 regions (4.84, 745.24, 226.12). The final value Δ_K^* is the minimum of these 3 values, i.e. we have $\Delta_K^* = 4.84$. Col. 6: contours of $[\nabla L]_1 = 0$ and $[\nabla L]_2 = 0$ for L computed with $\Delta_2 = \Delta_3 = 0.9\Delta_K^*$. The contours have only one point of intersection (only one point where $\nabla L = 0$). Col. 7: contours of of $[\nabla L]_j = 0, j = 1, 2$ for $\Delta_2 = \Delta_3 = 1.1\Delta_K^*$. There are 3 intersection points (3 points where $\nabla L = 0$).

An easy to verify sufficient condition to ensure (4) holds is

$$\max_{p=1,\ldots M-K} \Delta_{p+K} < \inf_{v_{t,r} \in \mathcal{G}} \max_{p=1,\ldots M-K} \gamma_p(v_{t,r}) \triangleq \Delta_K^* \quad (8)$$

Proof: We need a set of conditions that ensure that L^i has a unique minimizer.The second term in (3) is strongly convex everywhere. Consider the region $\mathcal{R}_{LC}$. Condition 2 ensures that $\mathcal{R}_{LC}$ exists. By its definition, $E(C)$ as a function of C is convex in $\mathcal{R}_{LC}$. By condition 1, this implies that the Hessian $\nabla_C^2 E \geq 0$ in $\mathcal{R}_{K,LC}$, which in turn implies that $\nabla_{v_{t,r}}^2 E = B_r^T[\nabla_C^2 E]B_r \geq 0$ in $\mathcal{R}_{K,LC}$. Thus in $\mathcal{R}_{K,LC}$, $L^i(v_{t,r})$ is strongly convex, i.e. it has a unique minimizer. We need to show that outside $\mathcal{R}_{K,LC}$, i.e. in $\mathcal{R}_{K,LC}^c$, L^i has no minimizers. *A sufficient condition for this is that* $\nabla_{v_{t,r}} L \neq 0$, $\forall v_{t,r} \in \mathcal{R}_{K,LC}^c$.

Let T_2 denotes the second term of (3). Now, $\nabla L = \nabla E + \nabla T_2$ can be zero only in regions of $\mathcal{R}_{K,LC}^c$ where, for all p, $[\nabla E]_p$ and $[\nabla T_2]_p = \frac{[\nabla D]_p}{\Delta_{p+K}}$ either have different signs or are both zero. Thus in $\mathcal{R}_{K,LC}^c$, ∇L can be zero only in $\mathcal{G} = \cap_{p=1}^{M-K}(\mathcal{A}_{K,p} \cup \mathcal{Z}_{K,p})$.

We need a condition on Δ_{p+K} that ensures that

$$[\nabla L]_p \neq 0, \text{ for some } p, \ \forall v_{t,r} \in \mathcal{G} \quad (9)$$

Now T_2 only has a minimizer inside $\mathcal{R}_{K,LC}$ and none outside it. Thus if we can find a condition on Δ_{p+K} that ensures that $[\nabla L]_p > 0$ where $[\nabla T_2]_p > 0$ and $[\nabla L]_p < 0$ where $[\nabla T_2]_p < 0$ for at least one p for all points in $\mathcal{G}$, we will be done. Thus, at any point $v_{t,r} \in \mathcal{G}$, we need $[\nabla L]_p.[\nabla T_2]_p > 0$ for at least one p.

Inside $\mathcal{G}$, $v_{t,r}$ can either belong to $\cap_{p=1}^{M-K}\mathcal{A}_{K,p}$ or it can belong to $\mathcal{Z}_{K,p_1} \cap \ldots \mathcal{Z}_{K,p_j} \cap [\cap_{p \neq p_1,p_2,\ldots p_j} \mathcal{A}_{K,p}]$ for some $j > 0$. First consider a $v_{t,r} \in \cap_{p=1}^{M-K}\mathcal{A}_{K,p}$. For a given p, it is easy to see that $[\nabla L]_p.[\nabla T_2]_p > 0$ if $\Delta_{p+K} < \frac{|[\nabla D]_p|}{|[\nabla E]_p|}$. Thus for (9) to hold, we need that for each $v_{t,r} \in \cap_{p=1}^{M-K}\mathcal{A}_{K,p}$, there is at least one p for which $\Delta_{p+K} < \frac{|[\nabla D]_p|}{|[\nabla E]_p|}$. This is equivalent to requiring that

$$\max_{p=1,\ldots(M-K)} \left[\frac{|[\nabla D]_p|}{|[\nabla E]_p|} - \Delta_{K+p}\right] > 0, \ \forall v_{t,r} \in \cap_{p=1}^{M-K}\mathcal{A}_{K,p} \quad (10)$$

Now consider $v_{t,r} \in \mathcal{Z}_{K,p_1} \cap \ldots \mathcal{Z}_{K,p_j} \cap [\cap_{p \neq p_1,p_2,\ldots p_j} \mathcal{A}_{K,p}] \triangleq \mathcal{B}$. Here $[\nabla L]_{p_k} = 0, \forall k = 1, 2, ..j$. So in this case, for (9) to hold, we need

$$\max_{p \neq p_1,p_2,..p_j} \left[\frac{|[\nabla D]_p|}{|[\nabla E]_p|} - \Delta_{p+K}\right] > 0 \ \forall v_{t,r} \in \mathcal{B} \quad (11)$$

Now (10) and (11) can be combined and compactly written as: $\max_p[\gamma_p(v_{t,r}) - \Delta_{p+K}] > 0$ for all $v_{t,r} \in \mathcal{G}$ or that $\inf_{v_{t,r} \in \mathcal{G}} \max_p[\gamma_p(v_{t,r}) - \Delta_{p+K}] > 0$ where γ_p is defined (5). This is the same as (4).

Condition (8) is sufficient for (4) because $\max_p[g_1(p) - g_2(p)] > \max_p g_1(p) - \max_p g_2(p)$.

Remark 1 *If $E(C)$ is Lipschitz, we will always get $\Delta_K^* > 0$ and hence we can always find a $\Sigma_r > 0$ for which $p^{**,i}$ is unimodal.*

Corollary 1 *For the LDSS model of [10], Theorem 1 applies with the following modifications: (a) Replace $B_r f_r^i$ by $g(B_r f_r^i)$ everywhere. (b) Redefine $\nabla E \triangleq B_r^T \nabla_v g(B_r v_{t,r}) \nabla_C E(\tilde{C}_t^i + g(B_r v_{t,r}))$ with $(\nabla_v g)_{i,j} \triangleq \frac{\partial g_j}{\partial v_i}$. (c) Directly define $\mathcal{R}_{K,LC} \subseteq \mathbb{R}^{M-K}$ as the largest contiguous region in the neighborhood of f_r^i where $E(\tilde{C}_t^i + g(B_r v_{t,r}))$ is convex as a function of $v_{t,r}$.*

Note, the above result is more general than that of [10].

2.2. Numerical Verification of Unimodality

When trying to verify (3) using numerical (finite difference) computations of gradients and Hessians, 0 needs to be replaced by a small number $\epsilon_0 > 0$, i.e. we need conditions to ensure $|[\nabla L]_p| > \epsilon_0$ for some p for all $v_{t,r} \in \mathcal{R}_{K,LC}^c$. To ensure $|[\nabla L]_p| > \epsilon_0$ for some p for all $v_{t,r} \in \mathcal{R}_{K,LC}^c$, the following two modifications are needed: redefine $\mathcal{Z}_{K,p}$ and $\gamma_p(v_{t,r})$ as follows

$$\mathcal{Z}_{K,p} \triangleq \{v_{t,r} \in \mathcal{R}_{K,LC}^c : |[\nabla E]_p| < \epsilon_0, \ \&[\nabla E]_p.[\nabla D]_p \geq 0\}$$

$$\gamma_p(v_{t,r}) \triangleq \begin{cases} \frac{|[\nabla D]_p|}{\epsilon_0 + |[\nabla E]_p|}, & v_{t,r} \in A_{K,p} \\ \frac{|[\nabla D]_p|}{\epsilon_0 - |[\nabla E]_p|}, & v_{t,r} \in Z_{K,p} \end{cases}$$

Example 1 (Computing Δ_K^*) Consider tracking temperature (denoted C_t) at M locations. Temperature at each location is measured using two types of sensors that have failure probabilities α_1 and α_2. If the sensor fails it outputs a random number distributed according to a pdf $p_f(y)$. We assume here that $p_f(y) = Unif(y; -a, a)$. If the sensor is working, the measured temperature is the actual temperature plus Gaussian noise. The noise is independent of the noise at other sensors. Failure of all the $2M$ sensors are also independent. Thus we have the following observation likelihood (OL):

$$p(Y_t|C_t) = \prod_{p=1}^{M} p(Y_{t,p}^1, Y_{t,p}^2|C_{t,p}) = p(Y_{t,p}^1|C_{t,p})p(Y_{t,p}^2|C_{t,p})$$

$$p(Y_{t,p}^j|C_{t,p}) = (1 - \alpha^j)\mathcal{N}(Y_{t,p}^j; C_{t,p}, {\sigma_{obs}^j}^2) + \alpha p_f(Y_{t,p}^j) \quad (12)$$

The state dynamics follows (1), i.e. change in temperature over time (v_t) at the different sensor locations is assumed to be zero mean and spatially correlated. The eigenvectors of the covariance of v_t are $[B_s\ B_r]$ and the eigenvalues are $\{\Delta_p\}$. The coefficients along B_s, B_r, denoted $v_{t,s}, v_{t,r}$, are assumed to follow a random walk model with $f_s(v_s) = v_s$ and $f_r(v_r) = v_r$.

Consider $M = 3$ and $K = 1$ so that $v_{t,r} \in \mathbb{R}^2$. We need to find a condition on Δ_2, Δ_3 that ensures that assumption 1 holds. Here $\mathcal{G}$ is a subset of the 2D plane and consists of 3 types of regions: $\mathcal{A}_{K,1} \cap \mathcal{A}_{K,2}$, $\mathcal{Z}_{K,1} \cap \mathcal{A}_{K,2}$ and $\mathcal{A}_{K,1} \cap \mathcal{Z}_{K,2}$. Since the second term of (3) is convex with minimizer f_r^i which belongs to $\mathcal{R}_{K,LC}$, there is no point in $\mathcal{R}_{K,LC}^c$ where both $[\nabla D]_1 = 0$ and $[\nabla D]_2 = 0$. Thus $\mathcal{Z}_{K,1} \cap \mathcal{Z}_{K,2}$ will always be an empty set. We show an example compuation of Δ_K^* in Fig. 1 for which we got $\Delta_K^* = 4.84$.

3. PF-MT: PF WITH MODE TRACKER

LDSS problems very often have a small dimensional effective basis, $X_{t,s}$, in which most of the state change occurs and a large dimensional residual space, $X_{t,r}$, in which the variance of the state change is small, i.e. trace(Σ_r) is small. Thus trace(Σ_{IS}^i) $\leq$ trace(Σ_r) will also be small. When this is true, a valid approximation is to replace importance sampling of $X_{t,r}^i$ from $\mathcal{N}(m_t^i, \Sigma_{IS}^i)$ (step 2 in Algorithm 1) by deterministically setting $X_{t,r}^i = m_t^i$. We call this the Mode Tracking (MT) approximation since m_t^i is the mode of $p^{**,i}$. Another valid approximation, when Σ_r is small, is to set $\Sigma_{IS}^i = \Sigma_r$. This and the fact that $X_{t,r}^i = m_t^i$ makes the denominator of $\tilde{w}_t^i$ constant (and hence it can be removed). The above modifications, called PF-MT, are summarized in Algorithm 2.

Consider the model of (1). We show below that when trace(Σ_r) is small, with high probability, there is little error in replacing a random sample from $\mathcal{N}(m_t^i, \Sigma_{IS}^i)$, by m_t^i, which is the mode of $p^{**,i}$.

Theorem 2 (IS-MT) *For (1), assume that conditions of Theorem 1 are satisfied. Let* $v_{t,r}^i \sim \mathcal{N}(m_t^i, \Sigma_{IS}^i)$. *Then,* $v_{t,r}^i$ *converges to* m_t^i *in probability as trace*(Σ_r) $\to 0$, *for almost all values of* $v_{t-1,r}^i, C_{t-1}^i, v_{t,s}^i, Y_t$.

Proof: From (1) and (3), we have

$$(\Sigma_{IS}^i)^{-1} = (\nabla_{v_{t,r}}^2 L^i)(m_t^i) = (\nabla_{v_{t,r}}^2 E)(\tilde{C}_t^i + B_r m_t^i) + \Sigma_r^{-1} \tag{13}$$

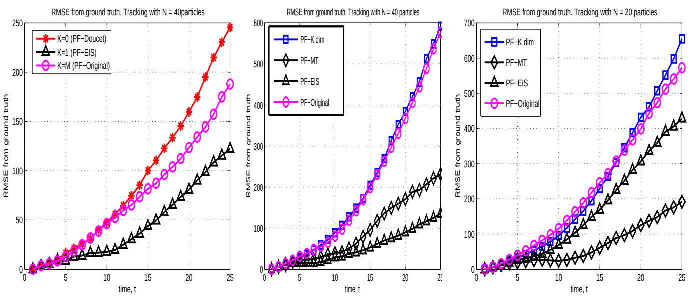
By Theorem 1, m_t^i is the unique minimizer of L^i. Also, as explained in the proof of Theorem 1, m_t^i lies inside $\mathcal{R}_{K,LC}$ and inside $\mathcal{R}_{K,LC}$, E is convex, i.e. $(\nabla_{v_{t,r}}^2 E)(\tilde{C}_t^i + B_r m_t^i) \geq 0$. Thus $(\Sigma_{IS}^i)^{-1} \geq \Sigma_r^{-1}$ or equivalently $\Sigma_{IS}^i \leq \Sigma_r$. This implies that

$$\text{trace}(\Sigma_{IS}^i) \leq \text{trace}(\Sigma_r) \tag{14}$$

Note that m_t^i is a function of $v_{t-1,r}^i, C_{t-1}^i, v_{t,s}^i, Y_t$. For any given values of $v_{t-1,r}^i, C_{t-1}^i, v_{t,s}^i, Y_t$, we have that

$$\begin{aligned} Pr(||v_{t,r}^i - m_t^i|| > \epsilon) &\leq \sum_{p=1}^{M-K} Pr([v_{t,r}^i - m_t^i]_p^2 > \frac{\epsilon^2}{M-K}) \\ &\leq \sum_{p=1}^{M-K} \frac{(\Sigma_{IS}^i)_{p,p}}{\epsilon^2/(M-K)} = \frac{\text{trace}(\Sigma_{IS}^i)}{\epsilon^2/(M-K)} \\ &\leq \frac{\text{trace}(\Sigma_r)(M-K)}{\epsilon^2} \triangleq \frac{\Delta_{K,tot}(M-K)}{\epsilon^2} \end{aligned}$$

The first inequality follows since $\{v_{t,r}^i : ||v_{t,r}^i - m_t^i||^2 > \epsilon^2\} \subseteq \{v_{t,r}^i :$ there exists at least one p for which $[v_{t,r}^i - m_t^i]_p^2 > \epsilon^2/(M-K)\}$ and applying the union bound on the probability of this event. The second inequality follows by applying Markov's inequality to $[v_{t,r}^i - m_t^i]_p)^2$. The third inequality holds because of (14).



(a) Examp 2,N=45 (b) Examp 3,N=40 (c) Examp 3,N=20

Fig. 2. RMSE (root mean square error) of the tracked temperatures from true values. RMSE is computed as square root of mean of squared norm of the error over 10 random simulations of the state space model. (a): Comparing RMSE of PF-EIS with that of original PF[1] and of PF-Doucet[4] for Example 2. (b) and (c): Comparing RMSE of PF-MT with that of PF-K dim, PF-EIS and original PF (M dim) for Example 3. (b): N=40 particles, (c): N=20 particles.

Thus, for every $\epsilon > 0$, we can find a $\delta_\epsilon = \epsilon^3/(M-K)$ s.t. if $\Delta_{K,tot} < \delta_\epsilon$, then $Pr(||v_{t,r}^i - m_t^i|| > \epsilon) < \epsilon$. This holds for any given values of $v_{t-1,r}^i, C_{t-1}^i, v_{t,s}^i, Y_t$. Thus the theorem follows.

The MT approximation introduces some error in the estimate of $X_{t,r}$ (error decreases with decreasing spread of $p^{**,i}$). But it reduces the PF dimension from $\dim(X_t)$ to $\dim(X_{t,s})$ (huge reduction for large dimensional problems), thus greatly improving the effective particle size. For carefully chosen dimension of $X_{t,s}$, this results in much smaller total error when the available number of particles, N, is small. Note also, that for best performance, one may choose a smaller dimensional $X_{t,r}$ (larger dimensional $X_{t,s}$) for PF-MT than that for PF-EIS, i.e. split $X_{t,r}$ for PF-EIS into $X_{t,r,s}$ and $X_{t,r,r}$ and use the MT approximation only on $X_{t,r,r}$.

4. SIMULATION RESULTS AND DISCUSSION

Example 2 Consider Example 1 with $M = 5$ sensor nodes and $K = K_{sim} = 1$. But now assume that the sensors at locations $K+1$ to M have zero failure probability (new sensors) and that $[B_s\ B_r] = I$. Thus OL is multimodal only as a function of $C_{t,1:K}$. Because of the choice of $[B_s\ B_r]$, $C_{t,1:K}$ depends only on $v_{t,s}$ and hence OL is multimodal only as a function of $v_{t,s}$ (and not $v_{t,r}$). In fact E will be a convex function of $v_{t,r}$ and hence $\mathcal{R}_{K,LC}^c$ will be empty. Thus Theorem 1 holds for $K = K_{sim}$ with $\Delta_1^* = \infty$ and so PF-EIS can be applied for any values of Σ_r. Simulation parameters were $\sigma_{obs}^2 = 5$, $p_f(y) = Unif(-100, 100)$, $\Sigma_s = 10$,$\Sigma_r = 2I_4$.

RMSEs of the tracked temperatures from their true value for this system, obtained using using PF-EIS with $K = K_{sim} = 1$, $K = 0$ (PF-Doucet[4]), and $K = M$ (original-PF[1]) is shown in Fig. 2(a). As can be seen, RMSE is smallest for PF-EIS.

Example 3 Consider Example 1 with $M = 10$ sensor nodes. All sensors have nonzero failure probability; $K = 3$ and $[B_s\ B_r]$ was a randomly chosen $M \times M$ orthonormal matrix (not identity). Also, $\sigma_{obs}^2 = 5$, $p_f(y) = Unif(y; -10, 10)$, $\Sigma_s = 25I_3$, $\Sigma_r = 2.5I_7$.

To track this system, a regular PF (PF-original or PF-EIS) will have to sample on $M = 10$ dimensions. But PF-MT utilizes the fact that the variance in residual space, Σ_r, is much smaller than Σ_s. It approximates $v_{t,r}^i$ by its posterior mode at each t (instead of importance sampling for it). This way the importance sampling dimension is only $K = 3$, but because of the MT step, the performance is much better than just running a K-dim original PF (run the PF only on the first K dimensions and treat $v_{t,r} \equiv 0$ for all t). Also, for small number of particles, N, its effective particle size is much better than that for either PF-EIS or PF-Original (M dim) and hence error is much smaller. As can be seen from Fig. 2(c), both PF-K dim and either of PF-EIS or PF-Original (PF-M dim) perform much worse than PF-MT for $N = 20$ particles. If N is allowed to increase to 40, PF-EIS (exact algorithm but lower effective particle size than PF-original) has the best performance (Fig. 2(b)). If N is increased further, say $N = 100$, all PFs have similar performance.

Note that $M = 10$ is a large enough dimensional state space if reasonable accuracy is desired with as low as $N = 20$ particles. In other practical scenarios (which are difficult to run multiple Monte Carlo runs of) such as contour tracking [13, 12] or tracking temperature in a wide area with large number of sensors, the state dimension can be as large as 200 or 250 while one cannot use more than 50 or 100 particles (for computational reasons).

There are still some un-addressed issues for PF-MT. If all or most particles $[v_{t,s}^i, v_{t,r}^i]$ stick to a wrong region somehow (because of the strong prior term, this will happen only if there are a sequence of bad observations), future particles of $v_{t,s}^i$ may get back because of random sampling, but $v_{t,r}^i$ will take very long (again because of strong prior term and no random sampling). This will result in loss of track. This problem will be much lesser if the dynamics of $v_{t,r}$ is either temporally independent or at least temporally stationary. Under this assumption, one should be able to show convergence of PF-MT as ϵ (used in Theorem 2) goes to zero. Temporal independence is a valid model for problems where the state vector can be interpreted as a "spatial signal" (e.g. temperature in space or contour tracking) and the effective basis is velocity at a subsampled set of points. For such problems, the state change (temperature change or contour deformation) is usually approximately bandlimited (spatially) at a frequency much smaller than the sampling frequency of the sensors or the image and so the value of K (computed using Nyquist criterion for the approximate bandwidth) is much smaller than M [12].

5. REFERENCES

[1] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, "Novel approach to nonlinear/nongaussian bayesian state estimation," *IEE Proceedings-F (Radar and Signal Processing)*, pp. 140(2):107–113, 1993.

[2] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking," *IEEE Trans. Sig. Proc.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[3] A. Doucet, N. deFreitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, 2001.

[4] A. Doucet, "On sequential monte carlo sampling methods for bayesian filtering," in *Technical Report CUED/F-INFENG/TR. 310, Cambridge University Department of Engineering*, 1998.

[5] Rudolph van der Merwe, Nando de Freitas, Arnaud Doucet, and Eric Wan, "The unscented particle filter," in *Advances in Neural Information Processing Systems 13*, Nov 2001.

[6] V. Cevher and J. H. McClellan, "Proposal strategies for joint state-space tracking with particle filters," in *ICASSP*, 2005.

[7] T. Schn, F. Gustafsson, and P.J. Nordlund, "Marginalized particle filters for nonlinear state-space models," *IEEE Trans. Sig. Proc.*, 2005.

[8] J.P. MacCormick and A. Blake, "A probabilistic contour discriminant for object localisation," *IEEE Intl. Conf. on Computer Vision (ICCV)*, Mumbai, India, June 1998.

[9] R. Chen and J.S. Liu, "Mixture kalman filters," *Journal of the Royal Statistical Society*, vol. 62(3), pp. 493–508, 2000.

[10] N.Vaswani, A.Yezzi, Y.Rathi, and A.Tannenbaum, "Particle filters for infinite (or large) dimensional state spaces-part 1," in *ICASSP*, 2006.

[11] N.Vaswani, "Particle filters for infinite (or large) dimensional state spaces-part 2," in *ICASSP*, 2006.

[12] N. Vaswani, A. Yezzi, Y. Rathi, and A. Tannenbaum, "Time-varying finite dimensional basis for tracking contour deformations from image sequences," in *IEEE Conf. Decision and Control (CDC)*, 2006.

[13] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi, "Particle filtering for geometric active contours and application to tracking deforming objects," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[14] A. Kale, N. Vaswani, and C. Jaynes, "Particle filter with mode tracker (pf-mt) for visual tracking across illumination change," in *ICASSP*, 2007, submitted.