BOUND ON ERRORS IN PARTICLE FILTERING WITH INCORRECT MODEL ASSUMPTIONS AND ITS IMPLICATION FOR CHANGE DETECTION

Namrata Vaswani

Dept. of Electrical and Computer Engineering & Center for Automation Research,
University of Maryland, College Park, MD 20742, USA

ABSTRACT

We study the errors in particle filtering with incorrect system model parameters. The total error in approximating the posterior distribution of the actual process (state) given noisy observations, can be split into modeling error and particle filtering error in tracking with the incorrect model. We show that the bound on both errors is a monotonically increasing function of the error in the system model per time step. The bound on the particle filtering error blows up very quickly since it has increasing derivatives of all orders. We apply this result to bounding the errors in approximating our statistic for slow change detection in nonlinear systems.

1. INTRODUCTION

Given noisy observations, $\{Y_t\}$, of some function, $\{h_t\}$, of the state, $\{X_t\}$, the non-linear filtering problem is (i) to evaluate the posterior probability distribution, π_t, of X_t given observations $Y_{0:t}$ and (ii) for any given function, ϕ, of the state, evaluate its expectation under π_t (posterior expectation). For most non-linear systems, exact filters do not exist and hence one resorts to approximate filtering methods like particle filtering [1] (which is a sequential Monte-Carlo method). Finite number of particles, N, in particle filtering (PF) introduce PF error which has been shown to converge to zero as $N \to \infty$ [1]. Also, often the correct system model (state transition kernel) is not known. If the system model error lasts for a finite time and assumptions for asymptotic stability [2] are satisfied, the asymptotic ($t \to \infty$) contribution of that error to the total error in the posterior can be shown to go to zero. We use bounds on these errors from [2] and show here that the bounds are an increasing function of the magnitude of the system model error per time step (quantified by a "distance metric" between the correct and incorrect state transition kernel).

We then apply this result to the slow change detection problem in non-linear systems with unknown change parameters [3]. The change detection statistic in [3] is a posterior expectation of a function of the state. Model error occurs due to the change. The "distance metric" now quantifies the rate of change (change magnitude per time step). In this case our result implies that the error in approximating the statistic for the changed system using a PF optimal for the original system, is smaller for slower changes.

Some other examples of PF under model error are: PF when the model parameters are learnt from insufficient training data; PF for tracking changing systems where change system parameters are learnt on the fly [4, 5]; PF of systems operating in multiple discrete modes (each with a different system model) [6].

Related Work: There has been a lot of recent research on stability of the optimal nonlinear filter. Asymptotic stability results w.r.t. initial condition were first proven in [7]. The Hilbert projective metric has been used to prove stability w.r.t. the initial condition and also w.r.t the model when the transition kernel is mixing [8]. We use in this paper results from [2] in which the authors have proved stability and also extended it to prove particle filter convergence, without assuming a mixing transition kernel.

In [9], PFs have been used for sudden change detection in non-linear systems with known change parameters. It defines a change detection statistic using observation likelihood ratio. When change parameters are unknown, one can modify this to thresholding observation likelihood or use tracking error [3] to detect sudden changes. For detecting slow changes, we have proposed a statistic called ELL [3] which is the posterior expectation of the negative log of the prior state pdf. We show here that the approximation of this statistic is more accurate for slower changes (which is intuitive).

System Model: We assume that we have an $\Re^{n_x}$ valued state process, $X = \{X_t\}$ and an $\Re^{n_y}$ valued observation process, $Y = \{Y_t\}$[1]. The system (or state transition) process $\{X_t\}$ is assumed to be a Markov process with state transition kernel $Q_t(x_t, dx_{t+1})$ and the observation process is defined by $Y_t = h_t(X_t) + w_t$ where w_t is an i.i.d. noise process and h_t is, in general, a nonlinear function. The prior initial state distribution, denoted by $\pi_0(dx)$, the conditional distribution of observation given state, $G_t(x_t, dy_t)$, with pdf given by $g_t(x, Y_t) \triangleq \psi_{t,Y_t}(x)$, and the state transition kernel, $Q_t(x_t, dx_{t+1})$, are known and assumed absolutely continuous [2].

Particle Filtering Algorithm: The problem of nonlinear filtering is to compute at each time t, the conditional probability distribution, of the state X_t given the observation sequence $Y_{1:t} = (Y_1, Y_2, ...Y_t)$, $\pi_t(dx) = Pr(X_t \in dx | Y_{1:t})$. The transition from π_{t-1} to π_t is defined using the Bayes recursion as follows:

$$\pi_{t-1} \;\text{—>}\; \pi_{t|t-1} = Q_t \pi_{t-1} \;\text{—>}\; \pi_t = \psi_t . \pi_{t|t-1} \triangleq \frac{\psi_t \pi_{t|t-1}}{(\pi_{t|t-1}, \psi_t)}$$

For nonlinear or nonGaussian system or observation model, except under very special cases, the filter is infinite dimensional and hence one approximates it by sequential Monte Carlo methods like particle filtering. The **particle filter (PF)** [1] or Bayesian Bootstrap Filtering [10] is a recursive algorithm which produces at each time t, a cloud of N particles $\{x_t^{(i)}\}$ whose empirical measure, π_t^N (a random measure), closely "follows" π_t. It starts

[1] We use the subscript 't' (e.g. X_t, Y_t) instead of 'n' for (discrete) time instants, to avoid confusion with N used for number of particles in Particle Filtering

[2] Note that for ease of notation, we denote the pdf by the same symbol as the probability distribution

with sampling N times from π_0 to approximate it by $\pi_0^N(dx) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x_0^{(i)}}(dx)$. The Bayes recursion then runs as follows:

$$\pi_{t-1}^N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x_{t-1}^{(i)}}(dx) \dashrightarrow \pi_{t|t-1}^N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{\bar{x}_t^{(i)}}(dx)$$

$$\dashrightarrow \bar{\pi}_t^N \triangleq \sum_{i=1}^N w_t^{(i)} \delta_{\bar{x}_t^{(i)}}(dx) \dashrightarrow \pi_t^N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x_t^{(i)}}(dx)$$

$$\begin{aligned} \text{where} \quad \bar{x}_t^{(i)} &\sim Q_t(x_{t-1}^{(i)}, dx), \\ x_t^{(i)} &\sim \text{Multinomial}(\{\bar{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^N) \\ w_t^{(i)} &\triangleq \frac{\psi_t(\bar{x}_t^{(i)})}{(\pi_{t|t-1}^N, \psi_t(\bar{x}_t^{(i)}))} \end{aligned} \quad (1)$$

Notation: We discuss here some more notation: $(.,.)$ is used to denote the inner product and $||.||$ denotes the total variation norm of any signed measure. "$a.s.$" denotes almost surely w.r.t. the probability distribution of the observation sequence $Y_{0:t}$. Also, $Q(\mu) \triangleq \int_x Q(x, dx')\mu(dx)$ where μ is a probability distribution, and $\bar{Q}(\mu) \triangleq \frac{Q(\mu)}{Q(\mu)(E)} = \frac{Q(\mu)}{\int_{x' \in E}\int_{x \in E} Q(x,dx')\mu(dx)}$.

Organization of the Paper: In Section 2, we state the results from past work and then present our own result on approximation error bounds. Section 3 explains the change detection problem and the application of our result to it. Conclusions are given in Section 4 and Section 5 (Appendix) contains the proof of our result.

2. BOUNDING APPROXIMATION ERRORS

Let (π_t, ϕ) is the posterior expectation of ϕ under the exact model and exact nonlinear filter. If there is a model error, i.e state transition kernel is Q_t^b while the non-linear filter uses Q_t^a, the estimated posterior is denoted by $\pi_t^{b,a}$. This is approximated using a PF with N particles, and then the posterior is denoted as $\pi_t^{b,a,N}$. In this section, we first provide some definitions and the convergence theorems from [2] and then present our result on bounding on the total error, $e_t^N = |(\pi_t, \phi) - (\pi_t^{b,a,N}, \phi)|$.

Definition 1 *The* **unnormalized filter kernel** *[2] for a system with state transition kernel Q_t and probability of observation given state $\psi_{t,Y_t}(x)$, is given by $R_{t,Y_t}(x, dx') = \psi_{t,Y_t}(x')Q_t(x, dx')$ [2]. So $R_{t,Y_t^a}^a = \psi_{t,Y_t^a}Q_t^a$ is the unnormalized optimal filter kernel for filtering observations coming from hypothesis a (denoted by H_a). In this paper we study the behavior of the non-optimal filter whose unnormalized kernel is given by $R_{t,Y_t^b}^a = \psi_{t,Y_t^b}Q_t^a$ i.e. the transition kernel under H_a used to track observations coming from hypothesis b (H_b).*

Definition 2 *A nonnegative kernel Q defined on state space E is* **mixing** *[2] if there exists a constant, $0 < \epsilon \leq 1$ and a nonnegative measure λ s.t.*

$$\epsilon\lambda(A) \leq Q(x, A) \leq \frac{1}{\epsilon}\lambda(A) \; \forall x \in E, \quad \forall \text{ Borel subsets } A \subset E$$

For a mixing kernel, the Birkhoff's contraction coefficient (explained in [2]), $\tau \leq \tilde{\tau}(\epsilon) \triangleq \frac{1-\epsilon^2}{1+\epsilon^2} < 1$.

Theorem 1 *(Model error bound, Theorem 4.6 of [2]): If for all k, the kernel R_{k,Y_k} is a.s. mixing ($\epsilon_k > 0, a.s.$ & $\tau_k \leq \tilde{\tau}_k(\epsilon_k) < 1, a.s.$), then the weak norm between the correct optimal filter density μ_t and the incorrect one μ'_t is upper bounded as follows:*

$$\sup_{\phi:||\phi||_\infty \leq 1} |(\mu_t - \mu'_t, \phi)| \leq \delta_t + \frac{\delta_{t-1}}{\epsilon_t^2} + \sum_{k=1}^{t-2} \tilde{\tau}_{t:k+3} \frac{\delta_k}{\epsilon_{k+1}^2 \epsilon_{k+2}^2}$$
$$\triangleq \theta_t(\delta_k, \epsilon_k, 0 \leq k \leq t), a.s.$$
$$\textit{where } \; \delta_k \triangleq \sup_{\phi:||\phi||_\infty \leq 1} |(\mu'_k - \bar{R}_{k,Y_k}(\mu'_{k-1}), \phi)|$$

Theorem 2 *(PF error bound, Theorem 5.7 of [2]): If for all k, the kernel R_{k,Y_k} is a.s. mixing ($\epsilon_k > 0, a.s.$ & $\tau_k \leq \tilde{\tau}_k(\epsilon_k) < 1, a.s.$), and $\sup_{x \in E_{x,y}} \psi_k(x) < \infty, a.s.$, then the weak norm between the correct optimal filter density μ_t and the approximation μ_t^N (evaluated using the PF) is upper bounded as follows [3]:*

$$\begin{aligned} \sup_{\phi:||\phi||_\infty \leq 1} \Xi_{pf}[|(\mu_t - \mu_t^N, \phi)|] &\leq \frac{2}{\sqrt{N}}(\rho_t + \frac{\rho_{t-1}}{\epsilon_t^2} + \sum_{k=1}^{t-2} \tilde{\tau}_{t:k+3} \frac{\rho_k}{\epsilon_{k+1}^2 \epsilon_{k+2}^2} \\ &\triangleq \frac{\beta_t(\rho_k, \epsilon_k, 0 \leq k \leq t)}{\sqrt{N}}, a.s. \\ \textit{where } \; \rho_k &\triangleq \frac{\sup_{x \in E} \psi_{k,Y_k}(x)}{\inf_{\mu \in \mathcal{P}(E)}(Q_k\mu, \psi_{k,Y_k})} < \infty, a.s. \quad (2 \end{aligned}$$

Using the above theorems, the total error in estimating posterior expectation of ϕ (assuming ϕ is a bounded function with $M = ||\phi||_\infty$) can be bounded as

$$e_t^{M,N} \leq |(\pi_t, \phi) - (\pi_t^{b,a}, \phi)| + |(\pi_t^{b,a}, \phi) - (\pi_t^{b,a,N}, \phi)| \leq M\theta_t + \frac{M\beta_t}{\sqrt{N}}$$

In a previously submitted work [11], we proved using the results stated above and some additional assumptions that if the system model error lasts for a finite time (say $[t_b : t_f]$) [4] and ϕ is bounded, the filtering error, e_t^N, tends to zero as $t \to \infty$ and for a given t, as $N \to \infty$. Also if ϕ is bounded from below but unbounded from above (or vice versa), one can consider its bounded approximation $\phi^M(x) \triangleq \min\{\phi(x), M\}$ and show convergence as $M \to \infty$. Thus we have $\lim_{M\to\infty}(\lim_{t\to\infty}(\lim_{N\to\infty} e_t^{M,N})) = 0$ in this case. Although this result was proven for the particular case of the change detection statistic which is the posterior expectation of an unbounded function of the state, it actually holds for any function of the state. We show here that the model error bound θ_t and the PF error bound coefficient β_t (and hence also the total error [5], e_t^N) are upper bounded by increasing functions of the distance between the incorrect and correct transition kernels.

2.1. Error Bounds as a function of Distance between $\mathbf{Q_t^b}, \mathbf{Q_t^a}$

We first define the distance metric, D_Q, between the state transition kernels Q_t^b, Q_t^a in terms of the distance between their corresponding unnormalized filter kernels. We also define another distance $\tilde{D}$ which measures the total model error in the posterior.

[3] Ξ_{pf} denotes expectation over different realizations of the particle filter, each of which produces a different random measure π_t^N

[4] even though the model error is for a finite time, it modifies the pdf of the state X_t (and hence also its posterior π_t) permanently.

[5] For ease of notation, we use e_t^N instead of $e_t^{M,N}$

Definition 3 *We define the* **distance metric between state transition kernels** Q_t^b **and** Q_t^a, $D_{Q,Y_t}(Q_t^b, Q_t^a)$ *(or* $D_{Q,t}$*), for a given observation* Y_t *as the following distance between* R_{t,Y_t}^b, R_{t,Y_t}^a:

$$\begin{aligned} D_{Q,Y_t}(Q_t^b, Q_t^a) &\triangleq D_R(R_{t,Y_t}^b, R_{t,Y_t}^a) \\ &\triangleq \sup_x \int_E |R_{t,Y_t}^b(x,x') - R_{t,Y_t}^a(x,x')|dx' \\ &= \sup_x \int_E \psi_{t,Y_t}(x')|Q_t^b(x,x') - Q_t^a(x,x')|dx' \end{aligned}$$

$D_{Q,t}$ [6] *quantifies the* **system model error per time step at time t**

Definition 4 *The* **total model error in the posterior** *is defined as total variation norm of the difference between the posteriors evaluated using correct and incorrect model scaled by* $\lambda_{k,Y_k^b}^b(E)$ *where* $\lambda_{k,Y_k^b}^b$ *is the invariant measure corresponding to* $R_{k,Y_k^b}^b$ [7] *:*

$$\tilde{D}_{t,Y_{0:t}} \triangleq \lambda_{k,Y_k^b}^b(E)||\pi_t^{b,a} - \pi_t^{b,b}|| \quad (3)$$

Now we state the main result of this paper (proof in Appendix):

Theorem 3 *Assuming (i)* $A_k > D_{Q,k}$ *and (ii)* $C > (\tilde{D}_{k-1})/\epsilon_k^b - D_{Q,k}$*, the following holds:*

$$\delta_k \le \frac{2D_{Q,k}}{A_k} \le \frac{2D_{Q,k}}{C - \frac{\tilde{D}_{k-1}}{\epsilon_k^b}} \triangleq f_\delta(D_{Q,k}, \tilde{D}_{k-1})$$

$$\rho_k \le \frac{\sup_x \psi_{k,Y_k}(x)}{{\epsilon_k^{b,a}}^2 (A_k - D_{Q,k})} \le \frac{\sup_x \psi_{k,Y_k}(x)}{{\epsilon_k^{b,a}}^2 (C - \frac{\tilde{D}_{k-1}}{\epsilon_k^b} - D_{Q,k})} \triangleq f_\rho(D_{Q,k}, \tilde{D}_{k-1}),\ a.s.$$

$$\text{where}\quad A_k \triangleq R_{k,Y_k^b}^b(\pi_{k-1}^{b,a})(E), \quad C \triangleq R_{k,Y_k^b}^b(\pi_{k-1}^{b,b})(E) \quad (4)$$

i.e. δ_k *and* ρ_k *are upper bounded by increasing functions of* $D_{Q,k}$ *and* $\tilde{D}_{k-1}$*. Also* f_δ *is a linear function of* $D_{Q,k}$ *while* f_ρ *is nonlinear (strictly convex), with derivatives of all orders also increasing so that* $f_\delta + f_\rho$ *also has the same property.*

Corollary 3: *The above theorem implies that both* $\theta_t(\delta_k, \epsilon_k^b, t_b \le k \le t)$ *and* $\beta_t(\rho_k, \epsilon_k^{b,a}, 0 \le k \le t)$ *are upper bounded by increasing functions of the vector of distances* $[D_{Q,k}, k = t_b, ...t]$ *and consequently* e_t^N *is also upper bounded by an increasing function of* $[D_{Q,k}, k = t_b, ...t]$ *with derivatives of all orders increasing. Also* e_t^N *increases with* t *as long as the change persists.*

Proof of Corollary: The corollary follows immediately from the definitions of θ_t, β_t (equations (2) & (2)), the above theorem and the following three facts: (i)ϵ_k^b is independent of $D_{Q,k}$, (ii) $\epsilon_k^{b,a}$ is a decreasing function of the total model error (shown in [11]) [8] and (iii) $\tilde{D}_{k-1}$ (total model error) is an increasing linear function of the past model errors per time step $[D_{Q,j}, j = t_c, ...k-1]$ (this is intuitively obvious, and so we skip the formal proof) and it also increases with k as long as the change persists.

[6] We use $D_{Q,t}$ to denote $D_{Q,Y_t^b}(Q_t^b, Q_t^a)$ for ease of notation.

[7] Scale by $\lambda_{k,Y_k^b}^b(E)$ only for ease of notation in stating the theorem

[8] with increasing total model error, the overlap between Y_k^b and Q_k^a decreases and so the kernel $R_k^{b,a}$ becomes less mixing ($\epsilon_k^{b,a}$ decreases)

3. THE CHANGE DETECTION PROBLEM

Consider a nonlinear partially observed system with state transition kernel Q_k^0 and probability of observation given state $\psi_{k,Y_k}(x)$. Now suppose there is a change in the system model at some unknown time t_c and the state transition kernel becomes Q_t^c (unknown) while the particle filter still uses Q_t^0 to track the observations. The aim is to detect the change with minimum delay.

If the change is sudden (large change magnitude per time step, quantified by $D_{Q,Y_k}(Q_k^c, Q_k^0)$), and the changed system parameters (Q_k^c) are known, log of observation likelihood ratio has been used to detect it [9]. For unknown change parameters, one can adapt this to use negative log of observation likelihood (denoted by OL), $OL_k^{c,0} \triangleq -\log Pr(Y_k^c|Y_{0:k-1}, H_0) = -\log(R_{k,Y_k^c}^0(\pi_{k-1}^{c,0})(E))$. OL exceeding a threshold was used by us [11] to detect a sudden change. But if the change is slow (change magnitude per time step is small, so D_Q small), $OL^{c,0}$ does not increase enough to detect the change. This is because,

$$OL_k^{c,0} \le -\log(A_k - D_{Q,k}) \le -\log(C - \frac{\tilde{D}_{k-1}}{\epsilon_k^c} - D_{Q,k}) \quad (5)$$

i.e. *it is upper bounded by an increasing function of* D_Q (follows by applying inequalities (8) & (9) (in Appendix) with $a = 0, b = c$).

To detect such slow changes which get missed by OL, we proposed using "Expected (negative) Log Likelihood" or ELL [3, 11]. ELL is the posterior expectation of the negative log of likelihood of the state (under prior distribution of the state, p_t^0), i.e.

$$ELL(Y_{0:t}) = E_{\pi_t}[-\log p_t^0(x)] = (\pi_t, -\log p_t^0). \quad (6)$$

It has been shown that when the change becomes detectable and errors in ELL approximation are small enough, ELL will detect the change correctly most of the time (small miss probability) [11]. Now, approximating ELL for changed observations using a PF optimal for unchanged observations fits in the framework of particle filtering with incorrect model assumptions. Here the model error per time step corresponds to rate of change (magnitude of change per time step). The function ϕ in this case is $\phi(x) = -\log p_t^0(x)$. Applying corollary 3, the error, $e_t^{c,0,N}$, in evaluating $ELL(Y_{0:t}^c)$ is upper bounded by a nonlinear increasing function (with increasing derivatives of all orders) of the vector of distances $[D_{Q,Y_k^c}(Q_k^c, Q_k^0), t_c \le k \le t]$ (distances now quantify the rate of change). Thus $ELL(Y_{0:t}^c)$ approximation is accurate for slow changes, for some time (until total change magnitude is small) and blows up quickly with increasing rate of change or increasing total change magnitude. Also, for the original system, $D_{Q,k}^0 = 0$ (no model error) and hence the only source of error, $e_t^{0,0,N}$, is finite particle size in PF, i.e. $\theta_t = \delta_t = 0\ \forall t$ and $\rho_k \le \frac{M \sup_x \psi_{k,Y_k}(x)}{C\sqrt{N}}$.

3.1. Improving ELL Approximation

Now for change detection to work best (detect change with minimum delay), error in approximating ELL should be small for both $ELL(Y_{0:t}^0)$ and $ELL(Y_{0:t}^c)$. Now $e_t^{0,0,N}$ is small (N chosen large enough), but $e_t^{c,0,N}$ depends on $D_{Q,k}$ being small $\forall k \le t$ (change slow, total change magnitude small). The nonlinearity of the error bounds, suggests that a small value of both $D_{Q,k}^0$ and $D_{Q,k}^c$ will introduce smaller total error $e_t^{0,0,N} + e_t^{c,0,N}$ than $D_{Q,k}^0 = 0$ and large $D_{Q,k}^c$. Thus instead of using Q_k^0 as the transition kernel in particle filtering ($Q_k^{pf} = Q_k^0$) [11], using a Q_k^{pf} that is closer to Q_k^c (even if its distance from Q_k^0 is not zero) will be a better idea. If Q_k^c is known, one could attempt to use a mixture of Q_k^0 and Q_k^c as Q_k^{pf}. For unknown Q_k^c, one could use Q_k^0 with a larger system noise variance as Q_k^{pf}. Both these ideas have been used in past works on tracking/filtering using a particle filter [10, 5]; we have in this paper provided a rigorous justification for using them.

4. CONCLUSION

We have upper bounded the approximation error in PF with incorrect model, by a nonlinear increasing function (with increasing derivatives of all orders) of the vector of model errors per time step, $[D_{Q,k}, k \leq t]$. We have applied this result to bound errors in approximating the ELL [3, 11] by a nonlinear increasing function of the rate of change and suggested ways to improve ELL approximation based on our result. As part of future work, we intend to study the change detection performance improvement and the bounds on approximation errors with (i) learning change parameter on the fly and (ii) with using ELL of a sequence of past states.

5. APPENDIX

Proof of Theorem 3: For ease of notation, denote $\sup_x \psi_{k,Y_k}(x) \triangleq S$. We first prove the following three inequalities below and then apply them to bound δ_k, ρ_k. Note that $R_{k,Y_k} = R^b_{k,Y_k^b}$ when applying Theorem 1 (model error bound) but $R_{k,Y_k} = R^a_{k,Y_k^b}$ when using Theorem 2 (PF error bound for incorrect model).

$$\begin{aligned}
&||R^a_{Y_k^b}(\pi^{b,a}_{k-1}) - R^b_{Y_k^b}(\pi^{b,a}_{k-1})|| \\
&\leq \int_x \int_{x'} |R^a_{Y_k^b}(x,x') - R^b_{Y_k^b}(x,x')| \pi^{b,a}_{k-1}(x) dx' dx \\
&\leq \sup_x \int_{x'} |R^a_{Y_k^b}(x,x') - R^b_{Y_k^b}(x,x')| dx' \\
&\triangleq D_R(R^a_{Y_k^b}, R^b_{Y_k^b}) = D_{Q,k} \qquad (7)
\end{aligned}$$

Also,

$$\begin{aligned}
|A_k - R^a_{k,Y_k^b}(\pi^{b,a}_{k-1})(E)| &= |R^b_{k,Y_k^b}(\pi^{b,a}_{k-1})(E) - R^a_{k,Y_k^b}(\pi^{b,a}_{k-1})(E)| \\
&\leq \int_{x'} |\int_x (R^a_{Y_k^b}(x,x') - R^b_{Y_k^b}(x,x'))\pi^{b,a}_{k-1}(x) dx| dx' \\
&= ||R^a_{Y_k^b}(\pi^{b,a}_{k-1}) - R^b_{Y_k^b}(\pi^{b,a}_{k-1})|| \overset{(a)}{\leq} D_{Q,k} \qquad (8)
\end{aligned}$$

Inequality (a) follows from of (7).

Next, we lower bound $A_k = C - (C - A_k)$:

$$\begin{aligned}
C - A_k = |C - A_k| &\leq ||R^b_{k,Y_k^b}(\pi^b_{k-1} - \pi^{b,a}_{k-1})|| \\
&\overset{(b)}{\leq} \frac{\lambda^b_{k,Y_k^b}(E)||\pi^b_{k-1} - \pi^{b,a}_{k-1}||}{\epsilon^b_k} \triangleq \frac{\tilde{D}_{k-1}}{\epsilon^b_k} \\
\text{Thus, } A_k &\geq C - \frac{\tilde{D}_{k-1}}{\epsilon^b_k} \qquad (9)
\end{aligned}$$

(b) follows from Lemma 3.5 of [2] and mixing property of R_k.

Now we use the above inequalities to bound δ_k:

$$\begin{aligned}
\delta_k &= \sup_{\phi:||\phi||_\infty \leq 1} |(\pi^{b,a}_k - \bar{R}^b_{Y_k^b}(\pi^{b,a}_{k-1}), \phi)| \\
&\leq ||\pi^{b,a}_k - \bar{R}^b_k \pi^{b,a}_{k-1}|| = ||\bar{R}^a_{Y_k^b}(\pi^{b,a}_{k-1}) - \bar{R}^b_{Y_k^b}(\pi^{b,a}_{k-1})|| \\
&\overset{(c)}{\leq} \frac{||R^a_{Y_k^b}(\pi^{b,a}_{k-1}) - R^b_{Y_k^b}(\pi^{b,a}_{k-1})|| + |A_k - R^a_{k,Y_k^b}(\pi^{b,a}_{k-1})(E)|}{A_k} \\
&\overset{(d)}{\leq} \frac{2D_{Q,k}}{A_k} \overset{(e)}{\leq} \frac{2D_{Q,k}}{C - \frac{\tilde{D}_{k-1}}{\epsilon^b_k}} \qquad (10)
\end{aligned}$$

Inequality (c) is an application of inequality (6) of [2], (d) follows by combining (7) and (8) and (e) follows from (9).

Now consider ρ_k:

$$\begin{aligned}
\rho_k &\overset{(f)}{\leq} \frac{S}{{\epsilon^{b,a}_k}^2 R^a_{k,Y_k^b}(\pi^{b,a}_{k-1})(E)} \\
&\overset{(g)}{\leq} \frac{S}{{\epsilon^{b,a}_k}^2 (A_k - D_{Q,k})} \overset{(h)}{\leq} \frac{S}{{\epsilon^{b,a}_k}^2 (C - \frac{\tilde{D}_{k-1}}{\epsilon^b_k} - D_{Q,k})}
\end{aligned}$$

Inequality (f) follows from Remark 5.10 of [2], (g) follows from (8) and assumption (i); (h) follows from (9) and assumption (ii).

6. REFERENCES

[1] P. DelMoral, "Non-linear filtering: Interacting particle solution," *Markov Processes and Related Fields*, pp. 555–580, 1996.

[2] LeGland F. and Oudjane N., "Stability and Uniform Approximation of Nonlinear Filters using the Hilbert Metric, and Application to Particle Filters," *Technical report, RR-4215, INRIA*, 2002.

[3] N. Vaswani, A. RoyChowdhury, and R. Chellappa, "Activity recognition using the dynamics of the configuration of interacting objects," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[4] H. Moon, R. Chellappa, and A. Rosenfeld, "3d object tracking using shape-encoded particle propagation," *IEEE International Conference on Computer Vision*, 2001.

[5] P. Fearnhead, "Sequential monte carlo methods in filter theory," in *PhD Thesis, Merton College, University of Oxford*, 1998.

[6] S. Zhou and R. Chellappa, "Probabilistic human recognition from video," in *European Conference on Computer Vision*, 2002, pp. 681–697.

[7] D. Ocone and E. Pardoux, "Asymptotic stability of the optimal filter with respect to its initial condition," *SIAM Journal of Control and Optimization*, pp. 226–243, 1996.

[8] F. LeGland and L. Mevel, "Exponential forgetting and geometric ergodicity in hidden markov models," *Mathematics of Control, Signals and Systems*, pp. 63–93, 2000.

[9] B. Azimi-Sadjadi and P.S. Krishnaprasad, "Change detection for nonlinear systems: A particle filtering approach," in *American Control Conference*, 2002.

[10] 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.

[11] N. Vaswani, "Change detection in partially observed nonlinear dynamic systems with unknown change parameters," in *Submitted to American Control Conference*, 2004.