Particle Filtering for Large Dimensional Problems with Multimodal Likelihoods

Namrata Vaswani
Dept of Electrical & Computer Engineering
Iowa State University
http://www.ece.iastate.edu/~namrata
HMM Model & Tracking

• Hidden state sequence: \( \{X_t\} \), observations: \( \{Y_t\} \)
  – state sequence, \( \{X_t\} \), is a Markov chain
  – \( Y_t \) independent of past & future given \( X_t \)
  – \( p(X_t|X_{t-1}) \): state transition prior (known)
  – \( p(Y_t|X_t) \): observation likelihood (known)

• **Tracking**: recursively get the optimal estimate of \( X_t \) at each \( t \) using observations, \( Y_{1:t} \)
  – compute/approximate the posterior, \( \pi_t(X_t) := p(X_t|Y_{1:t}) \)
  – use \( \pi_t \) to compute any “optimal” state estimate, e.g. MMSE, MAP,…
Problem Setup

• Observation Likelihood is often multimodal or heavy-tailed
  – e.g. some sensors fail or are nonlinear
  – e.g. clutter, occlusions, low contrast images
  – If the state transition prior is narrow enough, posterior will be unimodal: can adapt KF, EKF
    • If not (fast changing sequence): req. a Particle Filter

• Large dimensional state space (LDSS)
  – e.g. tracking the temperature field in a large area
  – e.g. deformable contour tracking
  – PF expensive: requires impractically large \( N \)
Narrow prior: Unimodal posterior

Temperature measured with 2 types of sensors, each with nonzero failure probability

Broad prior: Multimodal posterior
Large Dim. & Multimodal Examples

- **Sensor Networks**
  - Spatially varying physical quantities, e.g. temperature
  - Boundary of a chemical spill or target emissions

- **Image Sequences**
  - Boundary contour of moving & deforming objects
  - Deforming shapes of “landmark” points
  - Rigid motion & Illumination variation (over space & time)

- **Time-varying system transfer functions**
  - Time varying AR model for speech (e.g. STV-PARCOR)

- **Observation likelihood is frequently multimodal in most of the above problems**
Multimodal likelihood examples - 1

• Nonlinear sensor [Gordon et al’93]
  – sensor measuring the square of temperature corrupted by Gaussian noise
    \[ Y_t = X_t^2 + w_t, \quad w_t \sim N(0, \sigma^2) \]
  • whenever \( Y_t > 0 \), \( p(Y_t|X_t) \) is bimodal as a function of \( X_t \) with modes at \( X_t = Y_t^{1/2}, -Y_t^{1/2} \)

• Observ\( n = \) many-to-one function of state + noise
  – \( Y_t = h(X_{t,1})g(X_{t,2}) + w_t \): \( h, g \) monotonic functions
  – e.g. illumination & motion tracking [Kale-Vaswani’07]
Multimodal likelihood examples - 2

- Sensors with nonzero failure probability
  - temperature measured with 2 sensors, each with some probability of failure
  - bimodal likelihood if any of them fails
Multimodal likelihood examples - 3

- Deformable contour tracking [Isard-Blake’96][Vaswani et al’06]

through low contrast images (tumor region in brain MRI)

through overlapping background clutter
Particle Filter [Gordon et al’93]

• Sequential Monte Carlo technique to approx the Bayes’ recursion for computing the posterior

\[ \pi_t(X_{1:t}) = p(X_{1:t}|Y_{1:t}) \]

– approximation approaches true posterior as the # of M.C. samples ("particles") → \( \infty \) in most cases

• Does this sequentially at each time, \( t \), using Sequential Importance Sampling along with a Resampling step (to eliminate particles with very small importance weights)

– Our work: design of efficient importance densities
Particle Filter: Seq. Imp Sampling

- Sequential Imp Sampling for an HMM model
  - Replace $Y$ by $Y_{1:t}$, replace $X$ by $X_{1:t}$
  - Choose Imp Sampling density s.t. it factorizes as
    \[ q_{t,Y_{1:t}}(X_{1:t}) = q_{t-1,Y_{1:t-1}}(X_{1:t-1}) s_{X_{t-1},Y_t}(X_t) \]
    • allows for recursive computation of weights

- Seq Imp Sampling: At each $t$, for each $i$,
  - Importance Sample: $X_t^i \sim s_{X_{t-1}^i,Y_t}(X_t)$
  - Weight: $w_t^i \propto w_{t-1}^i p(Y_t|X_t^i) p(X_t^i|X_{t-1}^i) / s_{X_{t-1}^i,Y_t}(X_t^i)$
  - Posterior, $\pi_t(X_{1:t}) \approx \pi_t^N(X_{1:t}) = \sum_i w_t^i \delta(X_{1:t} - X_{1:t}^i)$
Outline

- Goal & Key Ideas
- PF - Efficient Importance Sampling (PF-EIS)
- Testing for posterior unimodality
- PF-EIS with Mode Tracking (PF-EIS-MT)
- Some More Applications & Open Issues
Our Goal

• Design efficient importance sampling techniques for PF, when
  – the likelihood is multimodal and the state transition prior is broad in at least some dimensions
  and/or
  – the state space dimension is large (compared to the available particle budget, N)
Key Idea 1: “LDSS property”

- In most cases, at any given time, most of the state change occurs in a small number of dimensions.
- The state change in the rest of the dimensions is small (state transition prior narrow).
  - Different from dim. reduction or from marginalizing over stationary distribution [Chorin et al.’04] [Givon et al.’08]
  - Related to the “compressibility” assumption used in lossy compression & in compressed sensing.
Key Idea 2: “Unimodality”

- Split the state space s.t. the posterior conditioned on a **small “multimodal”** part of the current state is unimodal
  - Possible to do this if the state change in the rest of dimensions is small enough (LDSS property)
  - We derive sufficient conditions to test for unimodality

- When this holds, we can
  - sample the “multimodal” states from the prior
  - use existing efficient sampling techniques for unimodal posteriors for the rest of the states (“unimodal states”)
Key Idea 3: “IS-MT”

• If for a part of the “unimodal” state space, state change still smaller: its conditional posterior will be quite narrow (besides being unimodal)

• If a sampling density is unimodal & narrow enough:
  – Any sample from it will be close to its mode w.h.p.
  – A valid approximation: use the mode as the sample

• Mode tracking (MT) approx. of importance sampling (IS) introduces some extra error but greatly reduces IS dimension
  – Lower approx. error when available N is small
PF with Efficient Importance Sampling (PF-EIS)
The Problem (recap)

• Observation likelihood is frequently multimodal

• State transition prior is broad and/or multimodal in at least some dimensions
Existing Work

• **PF-Original**: Importance Sample from prior [Gordon et al’93]
  – always applicable but is inefficient

• **Optimal IS density**: \( p^*(X_t) := p(X_t | X_{t-1}, Y_t) \) [D’98][older works]
  – cannot be computed in closed form most cases

• **When the optimal IS density, \( p^* \), is unimodal**
  – Adapt KF, EKF, PMT [Brockett et al’94][TZ’92][Jackson et al’04]
    • Possible if the posterior is unimodal too
  – PF-D: IS from Gaussian approx to \( p^* \) [Doucet’98]
  – Unscented PF [VDDW,NIPS’01]: UKF to approx to \( p^* \)

• **MHT, IMM, Gaussian Sum PF** [Kotecha-Djuric’03], …
  – practical only if total # of (possible) modes is small
PF-Efficient IS (PF-EIS)
[Vaswani, Trans. SP, Oct’08]

• Large dim problems w/ freq. multimodal likelihoods
  – No. of possible likelihood modes large: MHT, GSPF impractical
  – Prior broad in at least some dims: p* multimodal

• But, because of the LDSS property (small state change in most dimensions), it is possible to
  – split the state, X_t, into X_{t,s} (“multimodal” state) & X_{t,r} (“unimodal” state) s.t. p* conditioned on X_{t,s} is unimodal

• Modify existing algorithms as follows
  – sample X_{t,s} from its state transition prior
  – sample X_{t,r} from a Gaussian approx. to p* given X_{t,s}
PF-EIS algorithm

- Split $X_t = [X_{t,s}, X_{t,r}]$

- At each $t$, for each particle $i$
  - Imp Sample $X_{t,s}^i \sim p(X_{t,s}^i | X_{t-1}^i)$
  - Compute mode of $p^*$ conditioned on $X_{t,s}^i$ as
    - $m_t^i = \text{arg min}_x \left[ \log p(Y_t | x) + \log p(x | X_{t-1}^i, X_{t,s}^i) \right]$
  - Imp Sample $X_{t,r}^i \sim N(m_t^i, \Sigma_t^i)$
  - Weight
    - $w_t^i \propto w_{t-1}^i p(Y_t | X_t^i) p(X_{t,r}^i | X_{t-1}^i, X_{t,s}^i) / N(X_{t,r}^i ; m_t^i, \Sigma_t^i)$

- Resample when needed
Testing for Posterior Unimodality
The Problem (Simplified Version)
[Vaswani, Trans. SP, Oct’08]

- Posterior $\propto$ likelihood . prior
  - $p(x|y) = \alpha p(y|x) p(x)$

- The Problem:
  - Likelihood is multimodal
    - $E(x) := -\log p(y|x)$ has multiple local minima
  - Typically prior is strongly log-concave, e.g. Gaussian
    - $D(x) := -\log p(x)$ is strongly convex with a unique min. at $x_0$
  - How narrow should the prior be for the posterior to be unimodal
    - i.e. for $L(x) := -\log p(x|y) = E(x) + D(x)$ to have a unique minimizer?
Sufficient Conditions [Vaswani, Trans. SP, Oct’08]

- A posterior is unimodal if
  - the prior strongly log-concave, e.g. Gaussian
  - its unique mode, $x_0$, is close enough to a likelihood mode s.t. likelihood is locally log-concave at $x_0$
  - spread of the prior narrow enough s.t. $\exists$ an $\epsilon_0 > 0$ s.t.

$$\left[ \inf_{x \in \cap_p (A_p \cup Z_p)} \max_p \gamma_p(x) \right] > 1$$

$$\gamma_p(x) := \begin{cases} 
\frac{|[\nabla D(x)]_p|}{\epsilon_0 + |[\nabla E(x)]_p|} & x \in A_p \\
\frac{|[\nabla E(x)]_p|}{\epsilon_0 - |[\nabla E(x)]_p|} & x \in Z_p 
\end{cases}$$

$$Z_p := R_{LC'} \cap \{x : [\nabla E]_p \cdot [\nabla D]_p \geq 0, |[\nabla E]_p| < \epsilon_0\}$$

$$A_p := R_{LC'} \cap \{x : [\nabla E]_p \cdot [\nabla D]_p < 0\}$$
Implications: Posterior unimodal if

• Either the likelihood is unimodal or
• The prior is Gaussian-like and
  – its mode, $x_0$, is close enough to a likelihood mode
  – its maximum variance is small compared to distance b/w nearest & second-nearest likelihood mode to $x_0$

• Max variance upper bound increases with decreasing strength of second-nearest mode
Scalar case: Plots of $L(x)$, $\nabla L(x)$

$\nabla L$ has 1 zero crossing (stationary pt) for $\sigma_{\text{prior}}^2 = 0.9\Delta^*$

$\nabla L$ has 3 zero crossings (stationary pts) for $\sigma_{\text{prior}}^2 = 1.1\Delta^*$
Vector case (2D):
Contours of $[\nabla L(x)]_1=0$ (blue), $[\nabla L(x)]_2=0$ (red)

$\Delta_1 = \Delta_2 = 0.9 \Delta^*$
one intersection point
(one stationary point)

$\Delta_1 = \Delta_2 = 1.1 \Delta^*$
three intersection points
(three stationary points)
Application: Tracking random fields

- **State transition model**: State, $X_t = [C_t, v_t]$
  - temperature vector at time $t$, $C_t = C_{t-1} + Bv_t$
  - temperature change ($Bv_t$) is spatially correlated
  - temperature change coefficients along eigen-directions, $(v_t)$: Gaussian random walk

- **Observation**, $Y_t = \text{sensor measurements}$
  - different sensor measurements independent given $X_t$
  - with probability $\alpha_j$, sensor $j$ can fail
  - likelihood multimodal w.r.t. temperature at node $j$, if sensor at node $j$ fails (or sees outlier noise)
Choosing multimodal state, $X_{t,s}$

Practical heuristics motivated by the unimodality result

- Get the eigen-directions of the covariance of temperature change

- If one node has older sensors (higher failure probability) than other nodes:
  - choose temperature change along eigen-directions (a) which are most correlated to temperature at this node and (b) which have the largest eigenvalues, as $X_{t,s}$
Simulation Results: Sensor failure

• Tracking temperature at M=3 sensor nodes, each with 2 sensors

• Node 1 had much higher failure probability than rest

• PF-EIS: $X_{t,s} = v_{t,1}$

• PF-EIS (black) outperforms PF-D, PF-Original & GSPF
Simulation Results: Nonlinear sensor

- Tracking temperature at M=3 nodes, each with 1 sensor per node

- Node 1 has a squared sensor (measures square of temperature in Gaussian noise)
  - likelihood multimodal when $Y_t>0$

- PF-EIS (black) outperforms all others
PF-EIS with Mode Tracker (PF-EIS-MT)
The Problem (recap)

- As importance sampling dimension increases, N required for accurate tracking also increases
  - effective particle size reduces

- Regular PF impractical for very large dimensional problems
Existing Work

• If a large part of state space conditionally linear Gaussian or can be vector quantized
  – use Rao Blackwellized PF [Chen-Liu’00][SGN,TSP’05]

• If a large part of state space is asympt. stationary
  – marginalize over it using MC [Chorin et al’04][Givon et al’08]

• If cannot do either: PF-EIS with Mode Tracker

• Other work: Resampling modifications
  – Look ahead resampling: Auxiliary PF [Pitt-Shepherd’99]
  – Repeated resampling within a single t [Oudjane et al’03]
PF-EIS with Mode Tracking

[Vaswani, Trans. SP, Oct’08]

• If for a part of the unimodal state (“residual state”), the conditional posterior is narrow enough,
  – it can be approx. by a Dirac delta function at its mode

• Mode Tracking (MT) approx of Imp Sampling (IS)
  – MT approx of IS: introduces some error
  – But it reduces IS dimension by a large amount (improves effective particle size)
  – Net effect: lower error when N is small (if residual states carefully chosen)
PF-EIS-MT algorithm design

• Select the multimodal state, $X_{t,s}$, using heuristics motivated by the unimodality result, the rest of the states are $X_{t,r}$

• Split $X_{t,r}$ further into $X_{t,r,s}$, $X_{t,r,r}$ s.t. the conditional posterior of $X_{t,r,r}$ (residual state) is narrow enough to justify IS-MT
PF-EIS-MT algorithm

At each $t$, split $X_t = [ X_{t,s}, X_{t,r,s}, X_{t,r,r} ]$ &

- for each particle, $i$,
  - sample $X_{t,s,i}$ from its state transition prior
  - compute the conditional posterior mode of $X_{t,r}$
  - sample $X_{t,r,s,i}$ from Gaussian approx about mode
  - compute mode of conditional posterior of $X_{t,r,r}$ and set $X_{t,r,r,i}$ equal to it
  - weight appropriately

- resample
PF-MT

• PF-MT: computationally simpler version of PF-EIS-MT
  – Combine $X_{t,s}$ & $X_{t,r,s}$ & sample from the state transition prior for both
  – Mode Track $X_{t,r,r}$: compute conditional posterior mode of $X_{t,r,r}$ and set $X_{t,r,r}$ equal to it
Simulation Results: Sensor failure

- Tracking on M=10 sensor nodes, each with two sensors per node. Node 1 has much higher failure prob than rest.

- PF-MT (blue) has least RMSE
  - using K=1 dim multimodal state
Simulation Results: Modeling error

- Tracking on $M=5$ sensor nodes, each with one sensor per node
- Actual $p_{\text{fail}}$ of node 1 much larger than modeled one
- PF-EIS-MT (blue) has least RMSE
  - Using $K=1$, $K_2=2$
More Applications
Applications

- Tracking changes in spatially varying physical quantities, using a network of sensors [Vaswani, Trans. SP, Oct 2008]

- Tracking spatially varying illumination change of moving objects [Kale et al, ICASSP'07]

- Deformable contour tracking
  - Affine PF-MT [Rathi et al, CVPR'05, PAMI'07]
  - Deform PF-MT [Vaswani et al, CDC’06, Trans IP (to appear)]

- Global motion & shape change of a large set of “landmarks” (feature points of interest) from image sequences [Vaswani et al, Asilomar’07]
  - group of interacting people from a distance, human body parts during an action, e.g. dancing
Illumination & Motion Tracking: PF-MT

[Kale et al, ICASSP’07]

• Large dim: spatially & temporally varying illumination (object near enough to light source)
  – State = Motion (3 dim) + Illumination (7 dim)

• IS on motion (3 dim) & MT on illumination
  – Illumination changes very slowly
  – Image likelihood is usually unimodal conditioned on motion (i.e. as a function of illumination)
    • even if it is multimodal (e.g. in case of occlusions), the modes are usually far apart compared to the illumination change variance
Face tracking results (N=100)[Kale et al’07]
Face tracking: RMSE from ground truth

[Kale et al, ICASSP’07]

Comparing PF-MT with 10 dim regular PFs (original, auxiliary) & with PF- K dim (not track illumination at all). N = 100
Deformable Contour Tracking

- State: contour, contour point velocities
- Observation: image intensity and/or edge map

- Likelihood: neg. expo. of segmentation energies
  - Region based: observation = image intensity
    - Likelihood = probability of image being generated by the contour
    - Multimodal in case of low contrast images
  - Edge based: observation = edge locations (edge map)
    - Likelihood = probability of a subset of these edges being generated by the contour; of others being generated by clutter or occlusions or being missed due to low contrast
    - Multimodal due to clutter or occlusions or low contrast
Two proposed PF-MT algorithms

• **Affine PF-MT** [Rathi et al, CVPR’05, PAMI’07]
  – Imp sample on 6-dim space of affine deformations, Mode Track on residual deformation
  – Assumes either that likelihood modes separated only by affine deformation or that non-affine deformation per frame is small (slowly deforming sequence)

• **Deform PF-MT** [Vaswani et al, CDC’06, Trans IP (to appear)]
  – Imp sample on translation & on deformation at K sub-sampled locations around the contour
  – Useful when likelihood modes separated by non-affine deformation (e.g. due to overlapping clutter or low contrast) & large non-affine deformation per frame
Tracking using Affine PF-MT
[Rathi et al, CVPR’05, PAMI’07]

- Tracking humans from a distance (small deformation per frame)
- Deformation due to perspective camera effects (changing viewpoints), e.g. UAV tracking a plane

Condensation (PF 6-dim) fails
Tracking using Deform PF-MT

Low contrast images (tumor region in brain MRI)

Overlapping background clutter
Landmark Shape Tracking

- Track change in shape & global motion of a set of M “landmarks” from a seq. of images

- Advantage separating shape & motion dynamics
  - Learn shape dynamics with any camera, learn motion dynamics of the camera used for tracking

- **State** = [motion, shape, shape velocity]
- **Observation** = [M strongest edge locations]
  - or KLT-points or motion blobs or local PCA matching
Landmark Shape Tracking

PF-EIS: Red, PF-Gordon: blue, PF-Doucet: green

- PF-Gordon samples from STP: low effective particle size
- N=50 particles
- PF-D: assumes unimodal p*
- Often samples around wrong mode when p* multimodal
Summary

• Efficient Importance Sampling techniques that do not require unimodality of optimal IS density

• Derived sufficient conditions to test for posterior unimodality
  – developed for the conditional posterior, \( p^{\ast\ast}(X_{t,r}) := p(X_{t,r} | X_{t,s}^i, X_{t-1}^i, Y_t) \)
  – used these to guide the choice of multimodal state, \( X_{t,s} \), for PF-EIS

• If the state transition prior of a part of \( X_{t,r} \) is narrow enough, its conditional posterior will be unimodal & also very narrow
  – approx by a Dirac delta function at its mode: IS-MT
  – improves effective particle size: net reduction in error

• Demonstrated applications in
  – tracking spatially varying physical quantities using unreliable sensors
  – deformable contour tracking, landmark shape tracking, illumination
Collaborators

• Deformable contour tracking
  – Anthony Yezzi, Georgia Tech
  – Yogesh Rathi, Georgia Tech
  – Allen Tannenbaum, Georgia Tech

• Illumination tracking
  – Amit Kale, Siemens Corporate Tech, Bangalore

• Landmark shape tracking
  – Samarjit Das, Iowa State Univ. (my student)
Open Issues

• Parallel implementations, speed-up posterior mode comp.

• Current conditions for posterior unimodality expensive to verify, depend on previous particles & current observation
  – develop heuristics based on the result to efficiently select multimodal states on-the-fly, or
  – modify the result s.t. unimodality can be checked offline (select multimodal states offline)

• Residual space directions usually change over time
  – How do we either dimension-reduce (for PF-EIS) or select the MT directions (for PF-EIS-MT) on-the-fly
    • can we use Compressed Sensing or Kalman filtered CS [Vaswani’08] on the state change vector to do this?

• Analyze the IS-MT approx, prove stability of PF-MT