# The "Finkelstein" Particle Filter: A tool for the high-dimensional filtering problem

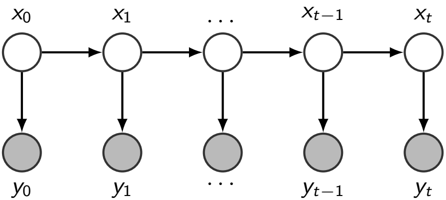
Jameson Quinn

1/25/2017

# Structure of talk (if there's not a fire alarm)

- Basic filtering problem
- Curse of dimensions for particle filters
- Existing state of the art (Rebeschini and van Handel, 2015): "Frankenstein"
- Proposed algorithm ("Finkelstein") and justification
- Numerical results
- Future work

# Basic filtering problem




What is $\pi_t \equiv (x_t | y_0, ..., y_t)$?

# Particle filter algorithm

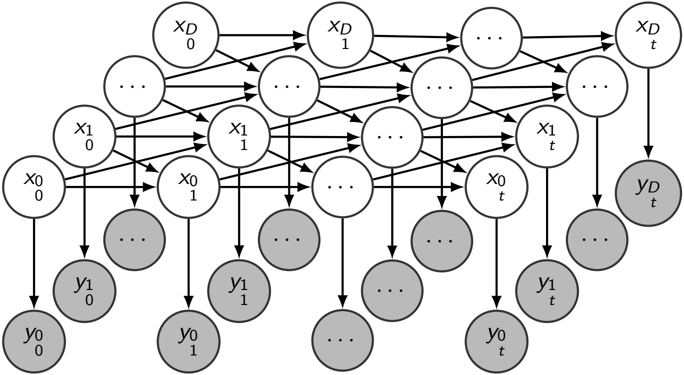
Algorithm:

- Assume we have $M$ particles $x_{t-1}^{1..M}$, taken to be "representative" samples from $(x_{t-1}|y_0, ..., y_{t-1})$.
- For each $x_{t-1}^i$, "progress" it to get $x_t^{i*} \sim (x_t|x_{t-1})$.
- Find weights $w^{i*} \equiv f(y_t|x_t^{i*})$
- Sample (with replacement) $M$ new $x_t^i$ from $\{x_t^{i*}\}$ using probabilities $\{w^{i*}\}$.

Characteristics:

- Clearly not iid, but $\frac{1}{M}\Sigma_{i=1}^M g(x_t^i)$ estimates $E_{\mathcal{F}}(g(x_t))$.
- Error falls like $\frac{1}{\sqrt{M}}$ and does *not* grow with $t$ (under weak assumptions).

# High-dimensional filtration




What is $\pi_t \equiv (x_t^{\cdots} | y_0^{\cdots}, ..., y_t^{\cdots})$?

# Curse of dimensionality

Each particle $x_t^i$ encompasses $x\overset{i}{\underset{t}{0}}, ..., x\overset{i}{\underset{t}{D}}$?

- Particle filter still "works" in that estimation error falls like $\frac{1}{\sqrt{M}}$ and doesn't grow with $t$.
- ... but error is exponential in $D$.

Separate particles for each $x\overset{i}{\underset{t}{l}}$?

- Works if $(x_t^l | y_{\underset{0}{\cdot\cdot}}, ..., y_{\underset{t}{\cdot\cdot}}) \perp\!\!\!\perp (x_t^k | y_{\underset{0}{\cdot\cdot}}, ..., y_{\underset{t}{\cdot\cdot}})$

# Existing state of the art (Rebeschini and van Handel, 2015)

- Assume we have $M$ particles $x^{1..M}_{\substack{\cdots\\t-1}}$, taken to be "representative" samples from $(x_{\substack{\cdots\\t-1}} | y_{\substack{\cdots\\0}}, \ldots, y_{\substack{\cdots\\t-1}})$.
- For each $x^{i}_{\substack{\cdots\\t-1}}$, "progress" it to get $x^{i*}_{\substack{\cdots\\t}} \sim (x_{\substack{\cdots\\t}} | x_{\substack{\cdots\\t-1}})$.
- *Split each* $x^{i*}_{\substack{\cdots\\t}}$ *into neigborhoods* $\{x^{i*}_{\substack{N_l\\t}} : 0 < l < L < D\}$
- Find weights $w^{i*}_{N_l} = f_{x_{\substack{N_l\\t}} | y_{\substack{N_l\\t}}}(x^{i*}_{\substack{N_l\\t}})$
- Sample (*independently*, with replacement) each neighborhood $N_l$ of $x^{i}_{\substack{N_l\\t}}$ from $\{x^{i*}_{\substack{N_l\\t}}\}$ using probabilities $\{w^{i*}_{N_l}\}$.

# Characteristics

- Rebeschini and van Handel show that the error using their method is bounded by $||\pi_{\substack{I\\t}} - \hat{\pi}^{M}_{\substack{I\\t}}||_{MC} \leq \alpha(\frac{e^{\beta|N_I|}}{\sqrt{M}} + e^{-\gamma \inf |I-b \in N_I^C|})$, where the constants $\alpha, \beta, \gamma$ do not depend on $t$.
- Thus there is a tradeoff: using smaller neighborhoods and/or more particles will control the term $\frac{e^{\beta|N_I|}}{\sqrt{M}}$, while using larger neighborhoods will control the term $e^{-\gamma \inf |m-b \in N_I^C|}$.

# “Finkelstein” solution 1/3

- Assume we have $M$ particles $x^{1..M}_{\substack{\cdots\\ t-1}}$, taken to be “representative” samples from $(x_{\substack{\cdots\\ t-1}} | y_{\substack{\cdots\\ 0}}, \ldots, y_{\substack{\cdots\\ t-1}})$.
- For each $x^{\,i}_{\substack{\cdots\\ t-1}}$, “progress” it to get a “full particle” $x^{i,0}_{\substack{\cdots\\ t}} \sim (x_{\substack{\cdots\\ t}} | x_{\substack{\cdots\\ t-1}})$ whose spatial “subparticles” are known as $x^{\,i}_{\substack{l\\ t}}$.
- Find weights for each subparticle locus, denoted $w^{\,i}_{l} \equiv f(y_{\substack{l\\ t}} | x^{\,i}_{\substack{l\\ t}})$; and probabilities conditional on $x^i_{t-1}$, denoted $f_i(x^{\,k}_{\substack{l\\ t}}) \equiv f(x^{\,k}_{\substack{l\\ t}} | x^{\,i}_{\substack{\cdots\\ t-1}})$.
- For each full particle $x^{i,0}_{\substack{\cdots\\ t}}$, run an MCMC chain targeting the filtration distribution up to convergence at $x^{i,C}_{\substack{\cdots\\ t}}$

# 2/3: Metropolis-Hastings MCMC (Starting at $x^{P,s}_{\underset{t}{\dots}}$)

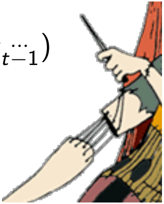
- Choose a spatial locus $l \in 1 \dots D$ and draw, with probability weighted by $w^{i}_{l}$, a proposed replacement subparticle for that locus $x^{k}_{\underset{t}{l}}$.
- Accept this replacement as $x^{P,s+1}_{\underset{t}{l}}$ with probability: $1 \wedge$

$$\frac{\sum_{i \in \{k, h^{P,s}(l)\}} f_i(x^{k}_{\underset{t}{l}}) \prod_{l \in n(l) \setminus l} f_i(x^{P,s}_{\underset{t}{l}})}{\sum_{i \in \{k, h^{P,s}(l)\}} \prod_{l \in n(l)} f_i(x^{P,s}_{\underset{t}{l}})} \quad (1)$$

- Sum of probabilities contributions from neighborhood *history*: $h^{i,j}(l) \equiv \{m : \exists k \in n(l) : x^{i,j}_{\underset{t}{k}} = x^{m,0}_{\underset{t}{k}}\}$
- Each contribution a product of locus probabilities over *neighborhood*: $n(l \in 1 \dots D) \equiv \{k\} : (x_{\underset{t}{l}} | x_{\underset{t-1}{k \in n(l)}}) \perp\!\!\!\perp x_{\underset{t-1}{n(l)^C}}$.

## 3/3: Why does this work?

Imagine $M \to \infty$ so $\frac{1}{M}\Sigma_i \delta(x_{\overset{\cdots}{t-1}}) \to \pi_{\overset{\cdots}{t-1}}$.



$$f(x_{\overset{\cdot\cdot}{t}} | y_{\overset{\cdot\cdot}{0}}, \ldots, y_{\overset{\cdot\cdot}{t}}) = f(x_{\overset{\cdot\cdot}{t}} | \pi_{\overset{\cdots}{t-1}}, y_{\overset{\cdot\cdot}{t}}) \propto f(y_{\overset{\cdot\cdot}{t}} | x_{\overset{\cdot\cdot}{t}}) f(x_{\overset{\cdot\cdot}{t}} | \pi_{\overset{\cdots}{t-1}})$$

$$= [\Pi_l f(y_{\underset{t}{l}} | x_{\underset{t}{l}})] \int_{\pi_{t-1}} \Pi_l f(x_{\underset{t}{l}} | x_{\overset{\iota}{\overset{\cdots}{t-1}}}) dx_{\overset{\iota}{\overset{\cdots}{t-1}}}$$

Converting the integral to a sum, the appropriate M-H acceptance for switching from $x^{0}_{\underset{t}{l}}$ to $x^{1}_{\underset{t}{l}}$ would be:

$$\frac{\cancel{w^{1}_{l}}\ \cancel{\Pi_{-l} w^{-l}} \qquad \Sigma_i[f_i(x^{1}_{\underset{t}{l}}) \Pi_{-l} f_i(x_{\underset{t}{-l}})]}{\cancel{w^{0}_{l}}\ \cancel{\Pi_{-l} w^{-l}} \qquad \Sigma_i[f_i(x^{0}_{\underset{t}{l}}) \Pi_{-l} f_i(x_{\underset{t}{-l}})]} \qquad \frac{\cancel{w^{0}_{l}}}{\cancel{w^{1}_{l}}} \qquad (2)$$

Exp. (1) above ignores "far away" and/or "unrelated" terms for computational simplicity, but unbiased estimator of exp. (2).

# Computational complexity:

- Progressing the particles and calculating weights: $O(MD)$
- Calculating conditional probabilities: $O(M^2D)$
- Running MCMC: $O(MDBC)$ where block size $B = |n(\dot{})| << D$ and $C$ is convergence time.
- Overall worst-case: $O(M^2DC) = O(\frac{D}{\epsilon^4}) << O(\frac{e^{\alpha D}}{\epsilon^2})$

# Assumptions involved; strong but probably unnecessary

We found the ratio $f(x^{i}_{\substack{l\\t}})/f(x^{j}_{\substack{l\\t}})$ using the expression $f(x_{\substack{\cdot\cdot\\t}}|x_{\substack{\cdots\\t-1}}) = \Pi_l f(x_{\substack{l\\t}}|x_{\substack{\cdots\\t-1}})$. This assumes:

- Can obtain density $f(x_{\substack{l\\t}}|x_{\substack{\cdots\\t-1}})$ (up to a constant).
- Conditional independence: $(x_{\substack{l\\t}}|x_{\substack{\cdots\\t-1}}) \perp\!\!\!\perp (x_{\substack{j\\t}}|x_{\substack{\cdots\\t-1}})$ for $l \neq j$.

Both are unrealistic in most interesting problems. But often we can still *estimate* that ratio feasibly.

# Error behavior: reasons for hope

Is the error $\propto \frac{1}{\sqrt{M}}$ and not dependent on $t$? Consider as density operators:

$$\pi_t = K_t P \pi_{t-1} = Q F_P K_t L P^* \pi_{t-1} \approx Q F_P K_t L S^N P^* \hat{\pi}^N_{t-1}$$

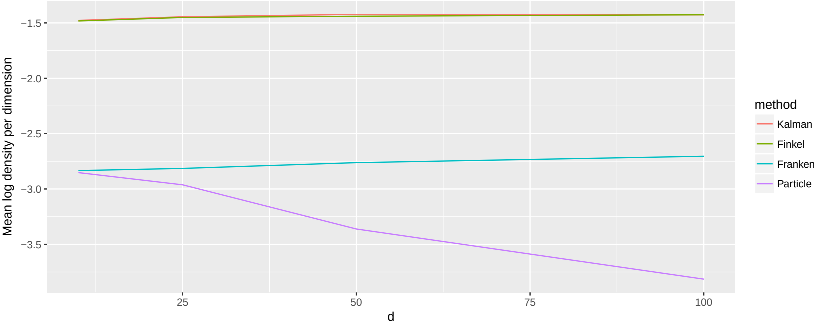
From right to left, in order of application, this last is:

1. Begin with $\hat{\pi}^N_{t-1}$ , the set of weighted particles at time $t-1$.
2. Progress with $P^*$, labeling each $x_t$ with its origin $x_{t-1}$.
3. Resample with $S^N$, for N new equally-weighted particles.
4. Localize with $L$, breaking the density over $\mathcal{W}_t \times \mathcal{W}_{t-1}$ into $D$ separate densities over each $\mathcal{L}_{\underset{t}{I}} \times \mathcal{W}_{t-1}$.
5. Reweight with $K_t$, based on $y_t$.
6. Rejoin with $F_P$, Finkelstein operator based on progression $P$.
7. Quash with $Q$, integrating over the history $\mathcal{W}_{t-1}$.

# Results

Fully Normal example, $d$ dimensions arranged linearly (2 neighbors), 5 random time steps from “tightish” prior, this computer, average of 16 runs with same model, non-optimized code (Julia)

Finkelstein: 50 particles, 35s; Frankenstein, 10K particles, 13s; Standard particle filter, 10K particles, 7.7s.



# I guess there wasn't a fire alarm

Further work:

- Loosen assumptions
  - Present states not independent conditional on past? Relatively easy.
  - Can's obtain density for present conditional on past? Less clear, but likely some of these ideas (MCMC chain targeting best-guess conditional distribution) still "help".
- Find applications/test data sets
  - Ideas?
- Finish proof of error behavior
  - Assuming perfect convergence (not unreasonable)
- Explore convergence issues
  - Effect of connectivity; conditions for phase change?

# Thank you!

Thanks to my advisor Luke Miratrix, and to Pierre Jacob for helping me refine these ideas and notation. All mistakes of course are my own.