Many thanks for the comments, folks. I've made changes to address all but one, with which I have a bit of a philosophical disagreement. Reviewer 3's comments were particularly useful and prompted me to rethink carefully the purpose of the simulations section. Responses are inline with the individual points below.

Reviewer Comments

Reviewer: 1

Recommendation: AQ - Publish In Minor, Required Changes

Comments

This is an interesting, well written paper that suggests new sampling strategies for smoothing in variable rate models. The proposed method rejuvenates the variable rate filtering particles that usually have poor smoothing performance due to the loss of particle diversity particularly when state discontinuities occur.

The paper relays on the key idea, using the previously generated filtering particles as a proposal to the target conditional distributions, which is inherited from the previous work of the author [Godsill et al. 2004] . The proposed algorithms apply a backward pass by sampling the target conditional distribution $P(\theta_n^-|.)$ for $n = N1$. Two methodologies are proposed for sampling the target conditional distribution of the variable rate structured "conditionally linear Gaussian models" and "conditionally deterministic models". The former samples from a discrete set of re-weighted filtering particles according to well known Rao-Blackwellization approach and the other uses an MCMC sampling scheme where the proposals are constructed by a modified filtering particle set.

Both algorithms are sophisticated techniques for rejuvenating the particles and can be interpreted as a way of reconstructing the particle history by using the filtering particles sampled at each filtering step. Reference on the convergence property of the algorithms and a discussion to understand the practical aspects of the the algorithm (Ex. whether is it possible to increase the online filtering performance by using retrospective smoothing steps, the ability of the algorithm to find the untracked modes of the posterior - this can also be suported with simulation results) would be useful for the reader in order to understand the limits of the algorithms.

I've added a brief note on convergence, which is inherited from the basic results for the forward-filtering-backward-sampling algorithm. I've also added an explicit example in the simulations of the smoother representing multi-modality where some particle contain a jump and others don't. Fixed interval smoothing won't help online filtering. Fixed-lag smoothing would, and this is covered elsewhere, especially in [Whitely et al. 2011].

In the simulations section, smoothing performance of the algorithms are compared to filtering and the filtering- smoothing results. In both tests improvement is achieved in means of RMSE. I see that in the simulations filter and filter- smoother is able to represent the solution to the data. However it would be much more interesting and convincing to see an illustration of a scenario where the filter is not able to represent the solution due to discarded particles in the resampling steps (Ex. poor estimate for the changepoint location or parameters, misses a jump or unable to track the maneuver due to multi-modality) but the smoother is able track the solution.

Because the smoother uses filter particles as a proposal, if the filter fails catastrophically (e.g. track loss) then the smoother will also fail. I have added an example where the multi-modality of the particle representation is lost due to resampling in the filter-smoother, but not the filter.

Minor comments:

Page: 37 line 36 : " θ " should be " θ' " Fig: 4 step 11 " $\theta*$ " should be " $\theta'*$ "

Well spotted. Now correct.

Reviewer: 2

Recommendation: A - Publish Unaltered

Comments

(There are no comments. Please check to see if comments were included as a file attachment with this e-mail or as an attachment in your Author Center.)

No file attached, so I assume there are genuinely no comments.

Reviewer: 3

Recommendation: RQ - Review Again After Major Changes

Comments

The authors propose a smoothing PF for variable rate models. The proposed method is essentially an adaptation of existing methods for fixed-rate models to the variable rate model. I feel there is enough substance in the adaptation process to merit publication provided some revisions are performed. My main concerns are with the performance analysis:

- It is never clarified but I'm assuming that the smoothing results are for fixed-interval smoothing, i.e., estimates at time $t_n, n = 1, ..., N$ use measurements from times $t_1, ..., t_N$. This should be clearly stated.

I feel this is clear from the introduction and the maths, but it's now explicitly stated as well.

- Only 10 realisations are used! It's difficult to draw conclusions with such a small sample size.

Ok. Now 100.

- In Figs 7 and 10 the effective sample size would be a better measure of performance than the number of unique particles.

- The principal improvement offered by the more sophisticated particle smoothers compared to the filter-smoother is increased particle diversity. This should improve estimation performance but the improvements in consistency may be even greater. This is alluded to in the paper and partially covered by Figs 6 and 9 which show plots of sample trajectories. However, a more rigorous consistency analysis should be performed. For example, a simple approach would be to construct a Gaussian approximation to the posterior, establish a confidence ellipsoid and then count the number of times the true value falls within this ellipsoid.

These two comments concern the same question: How should particle approximations of a distribution be compared? There is, of course, much more to a distributional estimate than how close its mean is to the true value, so we do need something more than just the RMSE, as you suggest. Ideally, we would like to compare our set of particles to the true posterior distribution, using statistics such as the effective sample size (ESS), or normalised estimation error squared (NEES). However this true posterior is not available. As a substitute, either a Gaussian approximation or a particle approximation with loads of particles are sometimes used as "truth". This is far from ideal, and I find the latter particularly unsettling, because we end up testing our particle algorithms only against their own best performance.

My preferred alternative for measuring consistency is a mean "empirical NEES" statistic which takes a value between 0 and 1 and takes into account the estimation error, covariance estimate and diversity of the particle approximation. I've added details and plots of this metric to the paper. I think that together with the RMSE values, this provides sufficient evidence of the improvement provided by the smoother.

- Comparing the proposed method and the filter-smoother for equal sample sizes doesn't really make sense given that the former has $O(N^2)$ expense while the latter has $O(N)$ expense. The filter-smoother should be implemented with a much larger sample size such that both algorithms have approximately equal expense. This approach, which is taken in (Fearnhead et al, 2010), seems much fairer.

I completely disagree. I think this would be misleading and is unnecessary for demonstrating the purpose of the new smoothers.

For a start, the filter-smoother and the smoother provide estimates with different qualities, and comparing them like this would be to compare apples with oranges. The principal reason for using a particle smoother is to rejuvenate a particle population with low path-space diversity. This is demonstrated by the increased number of unique particles and the improved consistency. (The reduced RMSE is a consequent bonus.) In contrast, [Fearnhead et al. 2010] are comparing a variety of smoothing algorithms all of which do the same thing, produce a diverse set of marginal smoothing particles. Their methods are not appropriate here.

Furthermore, if we were trying to solve a particular practical problem, such as producing the best state estimator possible for an offline tracker, then it would be a sensible test. However, the example simulations here are entirely fictional, cooked up simply to demonstrate that the new smoothing algorithm functions as *expected. The problem is that whatever this test demonstrated about relative performance, the result could be reversed by changing the set-up of the simulation.*

- Results should be shown for a range of sample sizes so we can get an idea of the required sample size for (close to) optimal performance.

Yes. I've added graphs showing the effect of varying the number of smoothing trajectories for the tracking case. The finance results are similar and so omitted for space.

Some minor comments: - A Rao-Blackwellised VRPF was also developed in (Morelande and Gordon, 2009). Their formulation was more general than that used here since it permits the model to be partially linear-Gaussian.

Yes. This should be and now is referenced. Actually, their formulation is not more general, just different. The Rao-Blackwellisation in our paper is of the state variables when the dynamics are linear-Gaussian, while [Morelande and Gordon, 2009] Rao-Blackwellise the linear parts of the motion parameters, but in a conditionally deterministic model. We could extend our work by using their Rao-Blackwellisation scheme on the piecewise-deterministic smoother.

- Recently, O(N) schemes for particle smoothing which improve upon the filter-smoother have been proposed by (Fearnhead et al, 2010) and (Douc et al, 2011). Perhaps these could be used for VR filtering. They should at least be mentioned.

I've added a reference to [Douc et al., 2011]. This is similar to our Metropolis-Hastings-based method which is already referenced (but unhelpfully so as it is still awaiting publication). ([Bunch and Godsill, 2012]). The Fearnhead method is for marginal (as opposed to joint) smoothing distributions only, so it doesn't really help. I've mentioned this in section IV.

- My understanding is that in (33) $\tau_{K_{n-1}+1}$ should be drawn from a truncated distribution which forces it to be greater than t_{n-1}. According to the notation of (4), shouldn't the denominator then be $S(\tau_{K_{n-1}}, ...)$ rather than $S(\tau_{K_{n-1}+1}, ...)$?

Yes. Well spotted! Corrected.

- On page 13, just before (51), there is a $\tilde{\theta}$ which should be a θ.

Fixed.

- In Fig. 9(a), are these results for the filter-smoother?

Yes. Clarified.

- I'm not sure all the appendices are necessary since the results can be found in other, easily obtained sources. At least

I agree. I've cut the first two, giving the derivation and initialisation of the two-filter smoother. The discretisation of the finance model remains, as I can't find a complete and equivalent version in a published work.