LS-N-IPS: AN IMPROVEMENT OF PARTICLE FILTERS BY MEANS OF LOCAL SEARCH

Péter Torma * **Csaba Szepesvári** **

* *Eötvös Loránd University, Budapest*
** *Mindmaker Ltd.*
H-1121 Budapest,
Konkoly Thege Miklós út 29-33.

Abstract: A modification of N-IPS, a well known particle filter method is proposed and is shown to be more efficient than the baseline algorithm in the small sample size limit and when the observations are "reliable". The algorithm called LS-N-IPS adds local search to the baseline algorithm: in each time step the predictions are refined in a local search procedure that utilizes the most recent observation. The uniform stability of LS-N-IPS is studied and results of experiments are reported both for a simulated and a real-world (visual) tracking problem.

Keywords: Particle filter, N-IPS, Visual Tracking

1. INTRODUCTION

With the increase of computational power non-linear filtering and in particular methods based on particle filtering are gaining importance. Variants of the N-Interacting Particle System (N-IPS) filter (Del Moral, 1998) have been successfully used in many applications, e.g. visual tracking, radar image localisation and volatility prediction (see e.g.(Isard and Blake, 1998) and (Pitt and Shephard, 1999)).

Although the computational power of today's computers made particle filtering practical in some applications, the majority of on-line filtering tasks still remains outside of the realm of problems solvable with this class of algorithms. Since the computational requirements of N-IPS grow linearly with the number of particles, a natural idea is to try to reduce the number of particles whilst keeping the performance the same so as to broaden the class of problems that can be tackled by particle filtering.

In this article we propose a modification of the N-IPS algorithm, which we shall call LS-N-IPS (local search+N-IPS). LS-N-IPS is intended to improve the efficiency of the N-IPS algorithm under the assumption that the observations are sufficiently "reliable". Roughly, this assumption means that the conditional variance of the observation noise given the past states is smaller than that of the dynamics.

It is well known that under this assumption N-IPS does not perform very well since most of the particles will bear a small likelihood owning to the sensitivity (peakiness) of the observation density. Therefore, in such a case typically a large number of particles are needed to achieve even a moderate precision.

We believe that the "reliability assumption" holds in many practical cases and is worth of being studied. In particular, in this article we provide an example with this property from the domain of vision based object tracking. We show both on this tracking problem and in a simulated scenario (where qualitative measurements are easier to make) that LS-N-IPS is significantly more effective than N-IPS. We also prove a uniform stability theorem stating the uniform boundedness of the error of approximation of the posterior estimated by the LS-N-IPS algorithm in the L^1 norm.

The article is organized as follows: In Section 2 we introduce the LS-N-IPS algorithm. In Section 3 we prove the above mentioned stability theorem. Results of some experiments in a simulated scenario are presented in Section 4, whilst the results on a real-world object tracking problem are presented in Section 5. A short discussion of related work is presented in Section 6, and conclusions are drawn in Section 7.

2. THE LS-N-IPS ALGORITHM

Let us consider the filtering problem defined by the system

$$X_{t+1} = f(X_t) + W_t, \tag{1}$$
$$Y_t = g(X_t) + V_t, \tag{2}$$

where $t = 0, 1, 2, \ldots$, $X_t, W_t \in \mathcal{X}$, $Y_t, V_t \in \mathcal{Y}$, and W_t, V_t are zero mean i.i.d. random variables. Let the posterior given the observations $Y_{0:t} = (Y_0, \ldots, Y_t)$ be π_t:

$$\pi_t(A) = P(X_t \in A | Y_{0:t}),$$

where $A \subset \mathcal{X}$ is any measurable set.

The proposed "Local Search"-modified N-IPS algorithm (LS-N-IPS) is as follows (N is the number of particles):

(1) Initialization:
  - Let $X_0^{(i)} \sim \pi_0$, $i = 1, 2, \ldots, N$ and set $t = 0$.

(2) Repeat forever:
  - Compute the proposed next states by $Z_{t+1}^{(i)} = S_\lambda(f(X_t^{(i)}) + W_t^{(i)}, Y_{t+1})$, $i = 1, 2, \ldots, N$.
  - Compute $w_{t+1}^{(i)} \propto g(Y_{t+1} | Z_{t+1}^{(i)})$, $i = 1, 2, \ldots, N$.[1]
  - Sample $k_{t+1}^{(i)} \propto (w_{t+1}^{(1)}, \ldots, w_{t+1}^{(N)})$, $i = 1, 2, \ldots, N$.
  - Let $X_{t+1}^{(i)} = Z_{t+1}^{(k_{t+1}^{(i)})}$, $i = 1, 2, \ldots, N$.

The only difference between the LS-N-IPS and the N-IPS algorithms is in the update of the proposed states. Here LS-N-IPS uses a non-trivial local search operator, S_λ, to "refine" the predictions $f(X_t^{(i)}) + W_t^{(i)}$, whilst in N-IPS one has $S_\lambda(x, y) = x$.

Our requirement for S_λ in LS-N-IPS is that it should satisfy $g(y|S_\lambda(x, y)) \geq g(y|x)$. The parameter $\lambda > 0$ defines the "search length": S_λ is usually implemented as a (local) search trying to maximize $g(y|\cdot)$ around x, in a neighborhood with size λ, e.g. $S_\lambda(y|x) = \mathrm{argmax}\{g(y|\tilde{x}) \,|\, \|\tilde{x} - x\| \leq \lambda\}$.

As a simple example, aiming to show the improved performance of LS-N-IPS consider the system defined by $X_{t+1} = V_{t+1}, Y_t = X_t + W_t$, where $\mathrm{var}(W_t) << \mathrm{var}(V_t)$. If $S_\lambda(x, y) = y$ (assuming that $g(y|y) = \mathrm{argmax}_x g(y|x)$) then the local search renders all particles to Y, i.e., $X_t^{(i)} = Y$ and thus the estimate of the position of X_t is $\int x d\pi_t^N = Y_t$. Here (and in what follows) π_t^N denotes the estimated posterior corresponding to the particle system $\{X_t^{(i)}\}_{i=1}^N$ of size N.[2] On the other hand the estimate of the N-IPS model is given by

$$\overline{X}_t = \sum_{i=1}^{N} \frac{g(Y_t | V_t^{(i)})}{\sum_{j=1}^{N} g(Y_t | V_t^{(j)})} V_t^{(i)}.$$

Clearly, under a wide range of conditions $E(|Y_t - X_t|^2 \,|\, X_t) << E(|\overline{X}_t - X_t|^2 \,|\, X_t)$ and in this sense, the estimate of LS-N-IPS is better than that of the N-IPS model. If we assume that V_t and W_t are Gaussian with respective covariances Q and R then the posterior can be computed analitically (using the Kalman-filter euqations), yielding $\overline{X}_t^* = Q(Q + R)^{-1} Y_t$. Therefore X_t^* is close to Y_t provided that $\mathrm{var}(W_t) << \mathrm{var}(V_t)$.

3. UNIFORM CONVERGENCE OF LS-N-IPS

Provided that the system to be filtered is sufficiently regular, the N-IPS model is known to provide a uniformly "good" estimate of the posterior, for any given particle-set size (see Theorem 3.1 of (Del Moral, 1998)). Here we extend this result to LS-N-IPS. First let us remark that LS-N-IPS can be viewed as an N-IPS algorithm where in each step an approximate dynamics is used in the prediction step. The approximation error can be controlled by the search length λ. Hence we give the theorem for the general case when approximate models are used in N-IPS.

Theorem 3.1. *Let $\hat{g}, \hat{K}_t$ be approximate models, $\epsilon > 0$ and $\hat{\pi}_0$ be an approximation of the prior π_0. Assume that the Radon-Nikodym derivative of both K and $\hat{K}_t$ exist w.r.t. the Lebesgue-measure and let us denote them by $K(x, z)$ and $\hat{K}_t(x, z)$, respectively. Assume that for some $\hat{\varepsilon} > 0$ and $\hat{a} > 0$*

$$\frac{1}{\hat{a}} \leq \hat{g}(y|x), g(y|x) \leq \hat{a} \tag{3}$$

and

$$\hat{\varepsilon} \leq \hat{K}_t(x, z), K(x, z) \leq \frac{1}{\hat{\varepsilon}}. \tag{4}$$

[1] Here $g(y|x)$ denotes the observation density function of the system to be filtered (cf. Equation 2).

[2] $\pi_t^N(A) = (1/N) \sum_{i=1}^N \chi_A(X_t^{(i)})$, where χ_A is the characteristic function of the measurable set A.

Further, assume that for some $\beta > 0$ these approximate entities satisfy

$$\sup_y h(G_y, \hat{G}_y), h(F, \hat{F}_t), h(\pi_0, \hat{\pi}_0) \leq \beta,$$

where h is the Hilbert projective metrics (G.Birkhoff, 1967), $G_y : L^1(\mathcal{X}) \to L^1(\mathcal{X})$ is defined by $(G_y f)(x) = g(y|x)f(x)$ and F is the Markov operator corresponding to K. $\hat{G}_y, \hat{F}_t$ are defined in an analogueosly. Consider the N-IPS model based on the approximate models and the approximate prior. Consider the empirical posterior $\hat{\pi}_t^N$ as computed by this N-IPS algorithm. Then the uniform bound

$$\sup_{t\geq 0} E(|\int f d\hat{\pi}_t^N - \int f d\pi_t| \,|\, Y_{0:t}) \leq \frac{5\exp(2\hat{\gamma}')}{N^{\hat{\alpha}/2}} + \frac{4\beta}{\log(3)(1-\tanh(C(K)/4)} \quad (5)$$

holds for any $N \geq 1$ and for any measurable function f satisfying $\|f\|_\infty \leq 1$, where $\hat{\gamma}'$ and $\hat{\alpha}$ are defined by

$$\hat{\alpha} = \frac{\hat{\varepsilon}^2}{\hat{\varepsilon}^2 + \hat{\gamma}'} \quad \text{with} \quad \hat{\gamma}' = 1 + 2\log\hat{a}.$$

$C(K)$ is defined by

$$C(K) = \ln \sup_{x,y,x',y'} \frac{K(x,y)K(x',y')}{K(x',y)K(x,y')}$$

Proof. By the triangle inequality,

$$\begin{aligned}
&\sup_{t\geq 0} E(|\int f d\hat{\pi}_t^N - \int f d\pi_t| \,|Y_{0:t}) \leq \\
&\sup_{t\geq 0} E(|\int f d\hat{\pi}_t^N - \int f d\hat{\pi}_t| \,|Y_{0:t}) + \\
&\quad \sup_{t\geq 0} E(|\int f d\hat{\pi}_t - \int f d\pi_t| \,|Y_{0:t}), \qquad (6)
\end{aligned}$$

where $\hat{\pi}_t$ is defined by

$$\hat{\pi}_{t+1} = \frac{\hat{G}_{Y_t}\hat{F}_t\hat{\pi}_t}{(\hat{G}_{Y_t}\hat{F}_t\hat{\pi}_t)(\mathcal{X})}.$$

By Theorem 3.1 of (Del Moral, 1998) the first term of the right hand side of (6) can be bounded by $\frac{5\exp(2\hat{\gamma}')}{N^{\hat{\alpha}/2}}$.

By means of some contraction arguments and the well known properties of h (G.Birkhoff, 1967) it can be proven that

$$h(\hat{\pi}_t, \pi_t) \leq \frac{2\beta}{1-\tanh(C(K)/4)}$$

Since, by the well known inequality (see (G.Birkhoff, 1967))

$$\|\hat{\pi}_t - \pi_t\|_{TV} \leq \frac{2}{\log 3} h(\hat{\pi}_t, \pi_t)$$

and since $\|\hat{\pi}_t - \pi_t\|_1 \leq \|\hat{\pi}_t - \pi_t\|_{TV}$, we arrive at

$$\left|\int f d\hat{\pi}_t - \int f d\pi_t\right| \leq \|f\|_\infty \|\hat{\pi}_t - \pi_t\|_1 \leq \frac{4\beta\|f\|_\infty}{\log(3)(1-\tanh(C(K)/4))}.$$

Taking expectation of both sides and combining this inequality with that of derived above for the first term of (6) yields the result. □

4. SIMULATION RESULTS

In this section we compare the performance of LS-N-IPS and N-IPS in a simulated scenario. The dynamics we consider is as follows:

$$\begin{aligned}
\hat{X}_{t+1} &= X_t + S_{t+1}\Delta_t + W_t \\
S_{t+1} &= (2B_{t+1} - 1)S_t \\
U_{t+1} &= \chi(|\hat{X}_{t+1}| \leq K) \\
X_{t+1} &= U_{t+1}\hat{X}_{t+1} + (1-U_{t+1})X_t \\
\Delta_{t+1} &= U_{t+1}(X_{t+1} - X_t) + \\
&\qquad (1-U_{t+1})(X_{t+1} - \hat{X}_{t+1}),
\end{aligned}$$

where $W_t \sim \mathcal{N}(0,\sigma)$ are i.i.d. Gaussian random variables, and B_t is a Bernoulli variable with parameter α. The dynamics can be thought to model a "bouncing ball"(with no mass), where the ball is bounced at the points $-K, +K$ and at random time instances when $B_t = 0$.

The observation is $Y_t = X_t + V_t$ where $V_t \sim \mathcal{N}(0,\delta)$ i.i.d. The dynamics is highly non-linear, so linear filters would be useless. In the experiments we used $\alpha = 0.99$ (low probability of bounce), and $\sigma = 10\delta$ (uncertainity of dynamics is higher than that of the observation), $\delta = 0.5, K = 250$.

We tested both N-IPS and LS-N-IPS on this problem. Since we are interested in the performance when the number of particles is small we used $N = 10$. Time to time, both N-IPS and LS-N-IPS loose the object, i.e., $D_t = |\frac{1}{N}\sum_i X_n^{(i)} - X_n)| > \theta$ (we used $\theta = 25$). In order to make a quantitative comparison we ran 10^4 simulations of length $t = 400$ each and for each $T \in \{1, 2, \ldots, 400\}$ estimated the probability of loosing the object at time T for the first time. Results are shown in Figure 1. It should be clear from the figure that LS-N-IPS performs much better than N-IPS. The SEARCH-ONLY algorithm, also shown in the picture, is an LS-N-IPS algorithm with "zero" dynamics, i.e. $\hat{Z}_{t+1}^{(i)} = S_\lambda(X_t^{(i)}, Y_{t+1})$. This algorithm performs much worse than either of the two other ones, underlining the importance of the role of the dynamics in particle filtering.

Figure 2 shows the tracking precision of the same algorithms as a function of the time. More precisely, the figure shows the estimates of $E[D_t|D_t \leq \theta]$ as a function of t, as estimated by computing the averages over the 10^4 runs. The ordering of the

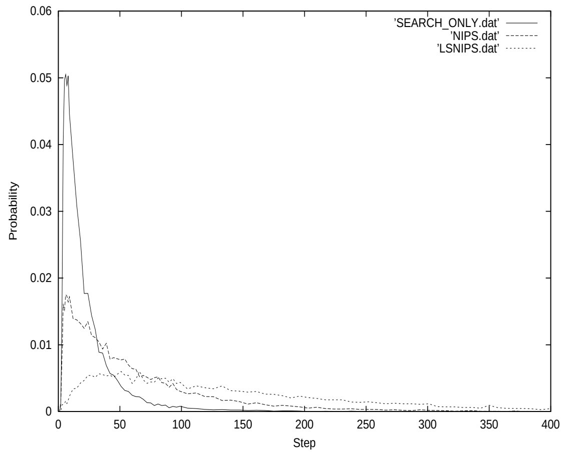

Fig. 1. *Probability of loosing the object as a function of time.*

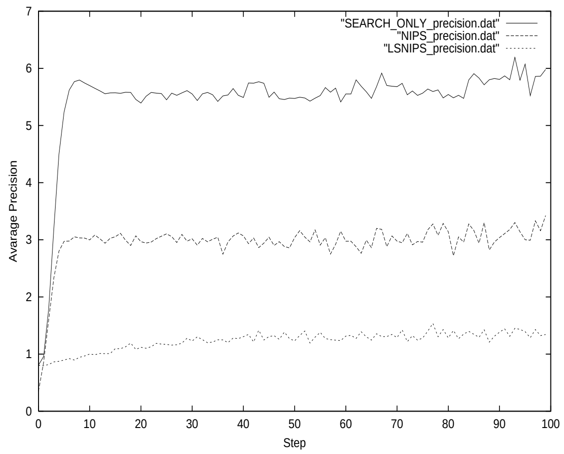

Fig. 2. *Tracking precision as a function of time.*

algorithms is the same as for the previous measurements: LS-N-IPS still performs significantly better than the other two algorithms.

5. RESULTS IN VISUAL TRACKING

The proposed algorithm was tested in a vision based tracking problem. Results are presented in the form of typical image sequences.

Figure 3 shows a typical image sequence of tracking a hand in clutter. The number of particles was chosen to be 100. This allows a tracking speed of 7 frame/seconds on a Pentium III 533 MHz computer without any further optimizations. Note that the speed of the algorithm is independent of the size of the image since the low level image processing involves only local computations. The local search procedure was implemented as a combination of greedy search and an LMS algorithm. For more details see (Torma and Szepesvári, 2000). Occasionally, the tracker looses the object, but it recovers soon in all cases. (The tracker quickly finds the object even when the initial position is randomized.) The image sequence is recorded with a frame rate of 30 frames/second.

Fig. 3. *Tracking hand in clutter with* 100 *particles. Black contours show particles having high observation likelihoods, whilst the white contour shows the actual estimate of the hand.*

The picture shows every 15th frame. N-IPS was unable to achieve a similar tracking performance even when the number of particles was increased above 1000 in which case the tracking speed became unacceptable.

6. RELATED WORK

LS-N-IPS can be thought of as a variance reduction technique. In the literature, many variance reduction techniques have been developed, here we mention only two of them.

The first method is Sequential Importance Sampling with Resampling (SIR) (Doucet, 1998). The design parameter of this method is the so-called "proposal distribution" used to generate the predictions of the particle positions. This should be designed such that the overall variance is reduced. Typically, sampling from the proposal is accomplished by a general purpose, randomized sampling procedure (e.g. MCMC), requiring the evaluation of the observation density at many points. Further, the running times of these sampling procedures are random - a disadvantage in tasks where guaranteed, real-time performance is required.

Another method, meant to avoid the problems of SIR, is the auxiliary variable method (AVM) by Pitt and Shephard (Pitt and Shephard, 1999). Unfortunately, this method requires the calculation of $O(R+N)$ observation likelihoods, where R, a design parameter of the method, should typically be chosen to be much larger than N.

Both SIR (with a carefully designed proposal distribution) and AVM were tested in the simulated scenario and were found to perform significantly worse than LS-N-IPS. Results of these experiments are not reported here due to the lack of space.

7. CONCLUSIONS

In this article a modification of the N-IPS algorithm was proposed. The modified algorithm is applicable when a local search operation maximizing the observation density function can be implemented. The algorithm was claimed to perform better for small values of N than N-IPS and when the observation model is reliable. A theorem proving the uniform stability of the proposed algorithm was given. Experiments confirmed that the proposed modification does improve over the quality of N-IPS. Given equivalent running times, the new algorithm comes with an increased precision, and given equivalent precisions, the new algorithm has a shorter running time than that of N-IPS.

Future work will include the application of the new algorithm in a number of other practical cases, such as speech processing and control, as well as a more thourough theoretical analysis with a focus on weakening the undesirably strong positivity assumptions of the proven stability theorem.

8. REFERENCES

Del Moral, Pierre (1998). A uniform convergence theorem for the numerical solving of the nonlinear filtering problem. *Journal of Applied Probability* **35**, 873–884.

Doucet, Arnaud (1998). On sequential simulation based methods for Bayesian filtering. *Statistics and Computing* **10**(3), 197–208.

G.Birkhoff (1967). *Lattice Theory.* Am. Math. Soc.

Isard, Michael and Andrew Blake (1998). CONDENSATION – conditional density propagation for visual tracking. *International Journal Computer Vision* **29**, 5–28.

Pitt, Michael K. and Neil Shephard (1999). Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association* **94**, 590–9.

Torma, Péter and Csaba Szepesvári (2000). Vision based object tracking using LS-N-IPS. (submitted).