Adaptive estimation of partially observed nonlinear stochastic systems

Anastasia Papavasiliou

Abstract— **In this paper, we discuss a method for the adaptive estimation of a Markov chain from its noisy partial observations, when the transition probability kernel depends on some unknown parameter. First, we re-write the system so that the parameter becomes part of it. Then, we apply a variation of the Interactive Particle Filter on the new system, in order to compute its optimal filter. The new system has been constructed in such a way that the marginals of its optimal filter are the optimal filter of the original system (up to an error due to the ambiguity in the parameter) and the posterior distribution of the parameter. We show that the error converges to zero, while the bayesian estimator of the parameter converges to the true value.**

Keywords— **nonlinear systems, identification.**

I. Introduction

In stochastic filtering, the goal is to compute the distribution of a stochastic process at any time instant, given some partial information up to that time. This distribution is called 'optimal filter'. The basic model usually consists of a Markov chain X (state variable) and a possibly nonlinear observation Y with observational noise V independent of the signal X. In this case, the optimal filter is determined completely by the observations, the transition probability kernel, the distribution of the noise, and the initial distribution. In practice, though, some of these elements will not be exactly known.

We are interested in computing the optimal filter for a system where the kernel depends on some unknown parameters. We study, instead, an equivalent system where the parameter is part of the state variable: the first component of the state variable is a Markov chain that evolves according to the kernel, whose parameter is set equal to the second component; the second component plays the role of the parameter and does not evolve. As a result, the marginals of the optimal filter of this system are the optimal filter of the original system as a function of the parameter, and the posterior distribution of the parameter. This technique of entering the parameter in the system is quite common in Bayesian statistics. There is an extensive discussion of this technique for partially observed Markov chains in [13]. In this paper, we show that this optimal filter will indeed converge to the true one, i.e. the one corresponding to the true value of the parameter. As a corollary, we get the asymptotic consistency of Bayes' estimator. This result is proven for a finite parameter space, but numerical results suggest that it should also be true for more general spaces.

By entering the parameters in the system, the identification problem becomes equivalent to the problem of asymptotic stability with respect to initial conditions. The study of the asymptotic stability of the optimal filter has been and still is an active area of research. In fact, many of the existing results ([12],[15],[14]) need to be revised, since the recent discovery of a gap in a proof of [12] (see [3] and [4]). The question has been resolved for some cases as, for example, discrete-time ergodic systems ([1]), or systems with finite state space ([3]). Our result can also be seen as an asymptotic stability result for a particular kind of non-ergodic systems.

It remains to find a way to compute the optimal filter. For this, we use a variation of the Interactive Particle Filter algorithm. The basic idea behind this is to approximate the optimal filter by an empirical distribution on some particles that move in such a way, so that the approximation remains true for every time point (see [6] and references within). Thus, applied in this setting, they compute the optimal filter under question, up to some error due to the ambiguity in the parameter, while at the same time they estimate the parameter. Numerical results show that the algorithm works well for low-dimensional compact spaces and can even be used to estimate slowly varying parameters or to detect abrupt changes in the parameters.

The structure of the paper is the following. In Section 2, we define the systems and state the main assumption, which is a form of identifibility condition. In Section 3, we prove the main result. In Section 4, we describe the algorithm for the computation of the optimal filter. In Section 5, we show some numerical results.

II. Definitions and Assumptions

Let E be a Polish space, i.e. a complete separable metric space and let us denote by $\mathcal{B}(E)$ its Borel σ-field. We study the asymptotic behavior of the conditional distribution of a Markov chain X taking values in E, given some noisy partial information, when the kernel depends on an unknown parameter θ. More specifically, we study the optimal filter of the following system, that we will refer to as:

System 1: Let $\{X_n\}$ be a homogeneous Markov chain taking values in $(E, \mathcal{B}(E))$. Let μ be its initial distribution and K_θ its transition probability kernel depending on a parameter $\theta \in \Theta$. Furthermore, we assume that for each $\theta \in \Theta$, K_θ is Feller and it satisfies the ergodicity conditions

$$\mu K_\theta^n \xrightarrow{w} \mu_\theta \quad and \quad \limsup_{n\to\infty} \mu_\theta |K_\theta^n f - \mu_\theta f| = 0, \ \forall f \in \mathcal{C}_b(E) \tag{1}$$

for some probability measure μ_θ. The observation process

Applied Physics and Applied Mathematics department, Columbia University, New York, NY 10027, USA. E-mail: pp2102@columbia.edu.

is defined by

$$Y_n = h(X_n) + V_n,$$

where V_n are i.i.d. $\mathbb{R}^p$-valued random variables independent of X, whose density function g is nowhere-vanishing and continuous, and $h : E \to \mathbb{R}^p$ is a bounded continuous function.

In practice, the parameter space Θ is usually Euclidean. More generally, we assume that it is a Polish space. We rewrite the system, so that the parameter becomes part of a Markov chain, whose transition probability kernel is now completely known. We will refer to the new system as:

System 2: Suppose now, that $\{\tilde{X}_n = (X_n, \theta_n)\}$ is an $E \times \Theta$-valued homogeneous Markov chain, with transition probability

$$\tilde{K}((x', \theta'), dx \otimes d\theta) = K_\theta(x', dx) \otimes \delta_{\theta'}(d\theta).$$

and initial distribution $\mu \otimes u$. The observation process is defined as in System 1, i.e.

$$Y_n = \tilde{h}(\tilde{X}) + V_n,$$

where $\tilde{h}(\tilde{x}) = h(x)$ and $\tilde{x} = (x, \theta)$.

We denote by P_θ the law of the Markov chain and by Q_θ the law of the observation process, i.e. $P_\theta = \mathcal{L}_\mu(X)$ and $Q_\theta = \mathcal{L}_\mu(Y)$. We use P^n_θ and Q^n_θ for their restrictions to the σ-algebras $\sigma(X_0, X_1, \ldots, X_n)$ and $\sigma(Y_1, \ldots, Y_n)$ respectively.

We denote by $\Psi^\theta_n(\mu)$ and $\Phi_n(\mu \otimes u)$ the optimal filters for Systems 1 and 2, with initial distributions μ and $\mu \otimes u$, respectively. Clearly, if $u = \delta_\alpha$, then $\Phi_n(\mu \otimes u)(dx) = \Psi^\alpha_n(\mu)(dx)$. Let $f \in \mathcal{C}_b(E \times \Theta)$. Then,

$$\Phi_n(\mu \otimes u)(f) \quad = \quad \mathbf{E}[f(X_n, \theta_n) | Y_n, \ldots, Y_1] \quad =$$

$$= \frac{\int_\Theta \int_{E^\infty} f(x_n, \theta) \prod_{k=1}^n g(y_k - h(x_k)) P_\theta(dx) u(d\theta)}{\int_\Theta \int_{E^\infty} \prod_{k=1}^n g(y_k - h(x_k)) P_\theta(dx) u(d\theta)}.$$

We set

$$\eta^\theta_n f = \int_{E^\infty} f(x_n, \theta) \prod_{k=1}^n g(y_k - h(x_k)) P_\theta(dx).$$

Then,

$$Q^n_u = \int_\Theta Q^n_\theta u(d\theta) = u(\eta^\theta_n \mathbf{1}),$$

and

$$\Phi_n(\mu \otimes u)(f) = \frac{u(\eta^\theta_n f)}{u(\eta^\theta_n \mathbf{1})}. \tag{2}$$

Our goal is to find under which conditions the optimal filter of System 2 is stable with respect to the initial distribution and compute the optimal filter recursively. First, we need to make some assumptions:

We define an equivalence relation on the parameter space as follows:

$$\alpha \sim \beta \Leftrightarrow \mu_\alpha \circ h^{-1} = \mu_\beta \circ h^{-1} \tag{3}$$

Recall that μ_θ in defined in (1). We assume that there is no pair of equivalent points in the parameter space. Otherwise, it is impossible to tell them apart by looking at the observations. A trivial example is when h is constant. Problems can also arise when h is symmetric.

This assumption is sufficient for the probability distributions of the observation process corresponding to different parameters, to be mutually singular, i.e. if $\alpha \not\sim \beta$, then $Q_\alpha \perp Q_\beta$ and $P_\alpha \perp P_\beta$. The proof is a straight forward application of Birkhoff's ergodic theorem on the ergodic chain (X, V).

To summarize, our main assumption will be the following:

Assumption 1: From now on, we assume that there is no identifiability problem, meaning that $\alpha \neq \beta$ implies $\alpha \not\sim \beta$ and thus $Q_\alpha \perp Q_\beta$.

We can, now, state the

III. Main result

In this section, we prove the asymptotic stability of System 2, with respect to the initial conditions, when Θ is a discrete space. To prove exponential rate of convergence, we will need the following condition, which is necessary for the Large Deviation Principle of the Markov chains to hold.

Condition 1: There exists a measurable function U mapping E into $[0, \infty)$ and having the following properties:

(a) $\inf_{x \in E}\{U(x) - \log \int_E e^{U(y)} K(x, dy)\} > -\infty$.

(b) For each $M < \infty$, the level set

$$Z(M) := \{x \in E : U(x) - \log \int_E e^{U(y)} K(x, dy) \leq M\}$$

is a relatively compact subset of E.

(c) U is bounded above on every compact subset of E.

Theorem III.1: Let α be the true parameter of System 2 (meaning that Q_α is the law of Y) and $u = \frac{1}{2}\delta_\alpha + \frac{1}{2}\delta_\beta$. Then,

$$\lim_{n \to \infty} \mathbb{E}_{Q_\alpha} |\Phi_n(\mu \otimes u)(f) - \Phi_n(\mu \otimes \delta_\alpha)(f)| = 0. \tag{4}$$

If, in addition, Condition 1 holds for K_α and K_β, then

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E}_{Q_\alpha} |\Phi_n(\mu \otimes u)(f) - \Phi_n(\mu \otimes \delta_\alpha)(f)| \leq -I < 0, \tag{5}$$

for some positive number I to be defined in the proof.

Proof: If $\beta = \alpha$ there is nothing to show, so we consider the case where $\beta \neq \alpha$. We rewrite the optimal filters using (2). Then, $\forall f \in \mathcal{C}_b(E)$,

$$\mathbb{E}_{Q_\alpha} |\Phi_n(\mu \otimes u)(f) - \Phi_n(\mu \otimes \delta_\alpha)(f)| =$$
$$= \int_{(\mathbb{R}^p)^n} |\frac{u(\eta^\theta_n f)}{u(\eta^\theta_n \mathbf{1})} - \frac{\eta^\alpha_n f}{\eta^\alpha_n \mathbf{1}}| Q^n_\alpha(dy^n) =$$

$$= \int_{(\mathbb{R}^p)^n} \left|\frac{\eta_n^\alpha f + \eta_n^\beta f}{\eta_n^\alpha \mathbf{1} + \eta_n^\beta \mathbf{1}} - \frac{\eta_n^\alpha f}{\eta_n^\alpha \mathbf{1}}\right| Q_\alpha^n(dy^n) =$$

$$= \int_{(\mathbb{R}^p)^n} \left|\frac{\eta_n^\alpha f \eta_n^\beta \mathbf{1} - \eta_n^\beta f \eta_n^\alpha \mathbf{1}}{\eta_n^\alpha \mathbf{1} \eta_n^\beta \mathbf{1}}\right| \left(\frac{\eta_n^\beta \mathbf{1}}{\eta_n^\alpha \mathbf{1} + \eta_n^\beta \mathbf{1}}\right) Q_\alpha^n(dy^n) =$$

$$= \int_{(\mathbb{R}^p)^n} |\Psi_n^\alpha(\mu)(f) - \Psi_n^\beta(\mu)(f)| \left(\frac{\eta_n^\beta \mathbf{1}}{\eta_n^\alpha \mathbf{1} + \eta_n^\beta \mathbf{1}}\right) Q_\alpha^n(dy^n) \leq$$

$$\leq 2\|f\|_\infty \int_{(\mathbb{R}^p)^n} \frac{dQ_\beta^n}{dQ_\alpha^n + dQ_\beta^n}(y) Q_\alpha^n(dy^n).$$

By Assumption 1, $Q_\alpha \perp Q_\beta$, so we can find sets $A^n \in \mathcal{B}((\mathbb{R}^p)^n)$ such that

$$Q_\alpha^n(cA^n) \to 0, \text{ and } Q_\beta^n(A^n) \to 0,$$

where by cA we denote the complement of a set A. (Take, for example, $A^n = \{y^n : \frac{1}{n} \sum_{k=0}^{n-1} f(y_k) \in [\nu_\alpha f - \frac{1}{n},\ \nu_\alpha f + \frac{1}{n}]\}$, where f is chosen so that $\nu_\alpha f \neq \nu_\beta f$.) Thus,

$$\int_{(\mathbb{R}^p)^n} \frac{dQ_\beta^n}{dQ_\alpha^n + dQ_\beta^n}(y) Q_\alpha^n(dy^n) =$$

$$= \int_{cA^n} \frac{dQ_\beta^n}{dQ_\alpha^n + dQ_\beta^n}(y) Q_\alpha^n(dy^n) + \int_{A^n} \frac{dQ_\alpha^n}{dQ_\alpha^n + dQ_\beta^n}(y) Q_\beta^n(dy^n)$$

$$\leq Q_\alpha^n(cA^n) + Q_\beta^n(A^n) \to 0$$

and this completes the proof of (4).

Now, let's assume that Condition 1 holds for K_α and K_β. Then, the upper bound for the large deviation principle holds (for a proof, see [9]), i.e. if $\mathcal{P}(E)$ is the space of Borel probability measures on E, equipped with the weak topology, then for any closed $A \subseteq \mathcal{P}(E)$,

$$\limsup_{n\to\infty} \frac{1}{n} \log P_\gamma^n\{x : \frac{1}{n}\sum_{k=1}^n \delta_{x_k} \in A\} \leq$$

$$\leq -I_\gamma(A) = -\inf_{\mu \in A} I_\gamma(\mu),$$

for $\gamma = \alpha, \beta$, where

$$I_\gamma(\mu) = \inf_{\{K \in \mathcal{T} : \mu K = \mu\}} \int_E \int_E \log \frac{dK(x.\cdot)}{dK_\gamma(x,\cdot)}(y) K(x,dy)\mu(dx),$$

and $\mathcal{T}$ is the set of all transition probability kernels. The Large Deviation Principle (LDP) also holds for $\{V_n\}$ with rate function

$$\tilde{I}(\nu) = \inf_{\{\nu' \in \mathcal{P}(\mathbb{R}^p)\}} \int_{\mathbb{R}^p} \log \frac{d\nu'}{d\nu}(x)\nu'(dx)$$

(Sanov's theorem). By the Contraction Principle, the LDP transfers to the measures Q_α^n and Q_β^n with rate functions J_γ, given by

$$J_\gamma(\nu) = \inf\{I(\mu) : (\mu \circ h^{-1}) * g = \nu\},\ \gamma = \alpha, \beta.$$

So, for any closed set $B \subseteq \mathcal{P}(\mathbb{R}^p)$ and $\gamma = \alpha, \beta$

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_\gamma^n\{y : \frac{1}{n}\sum_{k=1}^n \delta_{y_k} \in B\} \leq$$

$$\leq -J_\gamma(B) = -\inf_{\nu \in B} J_\gamma(\nu).$$

Using Jensen's inequality, we can show that $I_\gamma(\mu) = 0 \Leftrightarrow \mu = \mu_\gamma$ (see [7]) and thus, $J_\gamma(\nu) = 0 \Leftrightarrow \nu = \nu_\gamma$, for $\gamma = \alpha, \beta$. Since $\mathcal{P}(\mathbb{R}^p)$ equipped with the weak topology is a Polish space, it is also Hausdorff. Thus, there exists an $\eta > 0$ such that

$$B(\nu_\alpha, \eta) \cap B(\nu_\beta, \eta) = \emptyset.$$

We have seen that, for any $A^n \in \mathcal{B}((\mathbb{R}^p)^n)$ and for any $n \geq 0$,

$$\mathbb{E}_{Q_\alpha}|\Phi_n(\mu \otimes u)(f) - \Phi_n(\mu \otimes \delta_\alpha)(f)| \leq C(f)[Q_\alpha^n(cA^n) + Q_\beta^n(A^n)].$$

Choose $A^n = \{y : \frac{1}{n}\sum_{j=1}^n \delta_{y_j} \in B_\eta\}$, where $B_\eta = B(\nu_\alpha, \eta)$. Then,

$$\mathbb{E}_{Q_\alpha}|\Phi_n(\mu \otimes u)(f) - \Phi_n(\mu \otimes \delta_\alpha)(f)| \leq C(f)[Q_\alpha^n(cB_\eta) + Q_\beta^n(B_\eta)].$$

By taking the limit of the logarithm of the above expectation over n, as $n \to \infty$, the constants disappear and we are left with

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{Q_\alpha}|\Phi_n(\mu \otimes u)(f) - \Phi_n(\mu \otimes \delta_\alpha)(f)| \leq$$

$$\leq -\min\{J_\alpha(cB_\eta), J_\beta(\bar{B}_\eta)\},$$

which, by the choice of B_η, will be strictly negative($\bar{B}_\eta$ being the closure of B_η). Thus, (5) holds, with $I = \sup_{\eta > 0}\{J_\alpha(cB_\eta), J_\beta(\bar{B}_\eta) : B(\nu_\alpha, \eta) \cap B(\nu_\beta, \eta) = \emptyset\}$. ■

Remark III.2: The above result can be generalized to any $u = \sum_{j=1}^N w_j \delta_{\alpha_j}$, where α_1 is the true parameter, $w_1 > 0$ and $\sum_{j=1}^N w_j = 1$. In fact, (4) holds for any parameter space, provided that the prior distribution u has a positive mass on the true value. The proof is similar.

IV. The Algorithm

In this section, we describe a variation of the Interactive Particle Filter for the computation of the optimal filter. The Interactive Particle Filter (IPF) was originally suggested in [10] and [11], independently. The basic idea is to approximate the optimal filter by an empirical measure on particles that move in a way that imitates the dynamics of the optimal filter. For a complete and rigorous analysis, see [6].

Many different variations of the IPF have been proposed, adapted to the case where we include in the system the unknown parameters and thus, there are non-dynamic components. See, for example, [13] for a historical perspective, or [16] and [8] for more recent results. Below, we describe the simplest form of the algorithm, since it better corresponds to the theoretical results that we discussed previously.

Algorithm: Interacting Particle Filter (IPF)
1. At time $n = 0$:
The particle system consists of $N = M_1 \times M_2$ i.i.d. particles in $E \times \Theta$,

$$\begin{array}{ccc}
(\xi_0^1, \theta_0^1) & \dots & (\xi_0^{M_1}, \theta_0^1) \\
\vdots & \dots & \vdots \\
(\xi_0^{M_1(M_2-1)+1}, \theta_0^{M_2}) & \dots & (\xi_0^N, \theta_0^{M_2})
\end{array}$$

with common law $\mu \otimes u$. That is, we pick M_2 particles in Θ according to u and we map each of them with M_1 particles in E, simulated from μ.

2. At time $n \geq 1$:

(Mutation): The first components of the particles, ξ_{n-1}^i, $1 \leq i \leq N$, evolve according to the probability kernel $K_{\theta_{n-1}^i}$ of the signal, i.e. we simulate particle $\hat{\xi}_n^i$ from the distribution $K_{\theta_{n-1}^i}(\xi_{n-1}^i, \cdot)$. The second component remains the same.

(Selection): Particles $(\hat{\xi}_n^1, \theta_{n-1}^1), \dots, (\hat{\xi}_n^N, \theta_{n-1}^N)$ are resampled with weights

$$g(y_n - h(\hat{\xi}_n^1)), \dots, g(y_n - h(\hat{\xi}_n^N)).$$

In some cases, it is necessary to add some noise to the parameter in the mutation step. That is, the particles are 'mutated' to $(\hat{\xi}_n^1, \hat{\theta}_n^1), \dots, (\hat{\xi}_n^N, \hat{\theta}_n^N)$, where $\hat{\theta}_n^i = \theta_{n-1}^i + \sigma \epsilon_n^i$ and $\{\epsilon_n^i\}_{i=1}^n$ are i.i.d for every n. For example, if the parameter space is not compact and we have no prior information on the parameters, then the number of particles needed to satisfactorily cover the space is too large. Thus, we add some noise to the parameters so that they can wonder around the space and eventually visit the true value (for $dim(\Theta) \leq 2$). As a result, we lose something in the approximation but gain in the complexity of the algorithm.

Another case where we need to add noise is when the parameters are not really constant. If the parameters are slowly varying, we approximate their path by allowing the parameters to move along. Similarly, if there are jumps in the parameters and the time between the jumps is large enough, the particles will be able to detect the jump and eventually estimate the new value of the parameter. We cannot yet prove rigorous results in this direction, but numerical simulations as well as the fact that the rate of convergence is 'almost exponentially fast', in the sense of (5), suggest that this could be possible.

V. Numerical Examples

We run the algorithm on a system coming from the field of financial mathematics. Our purpose is not to study this system in terms of its importance in finance, but merely to use it as an example of a nonlinear system with unknown parameters. The reason for our choice is the availability of real data (stock prices), that plays the role of the observations in the system. More specifically, σ is the stochastic volatility of a stock and the observations are the log returns of the stock's price. The state evolution is given by

$$\begin{aligned}
\sigma_n &= \sigma_0 + e^{-\lambda \Delta t}(\sigma_{n-1} - \sigma_0) + \sqrt{\tfrac{\tau}{\rho^2}(1 - e^{-2\lambda \Delta t})}\epsilon_n \\
Y_n &= 1 + (\mu - \tfrac{1}{2}\sigma_{n-1}^2 \rho^2)\Delta t + \sigma_{n-1}\rho\sqrt{\Delta t}\epsilon_n'
\end{aligned} \tag{6}$$

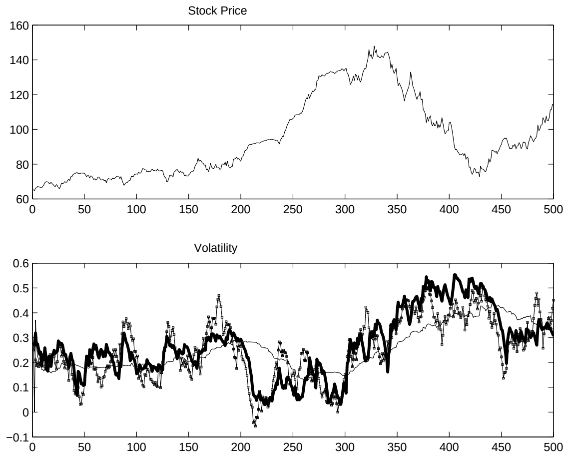

Fig. 1. We plot the simulated stock price, for $(\Delta t, \mu, \sigma_0, \lambda, \tau, \rho) = (1/252, .1, .25, 1, .5, -.5)$. Below, we plot the simulated stochastic volatility that corresponds to that stock price (the thick line), together with the estimated volatility by the IPF (line with squares) and the moving average of the volatility over a window of 60 days(thin line)

and the observations by

$$X_n = Y_n(1 - \frac{1}{2}\sigma_n^2(1 - \rho^2)\Delta t) + Y_n \sigma_n \sqrt{\Delta t(1 - \rho^2)}\tilde{\epsilon}_n, \tag{7}$$

where $\{\epsilon_n, \epsilon_n', \hat{\epsilon}_n \; ; \; n \geq 0\}$ are independent standard Gaussian random variables. The parameter Δt is known, while σ_0 and μ are estimated from the data (as the mean and standard deviation of the observations). The parameters $\lambda \in \mathbb{R}_+$, $\tau \in \mathbb{R}_+$ and $\rho \in [-1, 1]$ are included in the system. We add some noise to λ and τ, since their state space is not compact.

In order to test the algorithm, we first simulate the path of the price of an imaginary stock and its volatility, using (6) and (7). Then, we apply the IPF, after we estimated the parameters μ and σ_0 from the data.

In particular, we choose the following values for our parameters

$$(\Delta t, \mu, \sigma_0, \lambda, \tau, \rho) = (1/252, .1, .25, 1, .5, -.5),$$

so as to mimic the properties of real data. We run the Interactive Particle Filters with 2^{15} particles, adding noise (of standard deviation .01) to the parameters λ and τ. For prior distributions, we choose a uniform distribution in $[-1, 1]$ for ρ and a $\chi^2(2)$ distribution for τ and λ.

In Figure 1, we plot the simulated price and volatility, along with the estimated volatility by the IPF, and the moving average of the volatility over a window of 60 days. In general, the estimated volatility seems to follow the actual one quite closely. In Figure 2, we plot the estimates of the parameters λ, τ and ρ over time.

Next, we run the IPF on real data, some Industrial Index of Dow Jones over the span of almost nine years (starting 1/1/1992). In this case, we would expect the parameters to vary slowly and even jump from time to time. In Figure 3, we plot the estimates of the parameters over time.

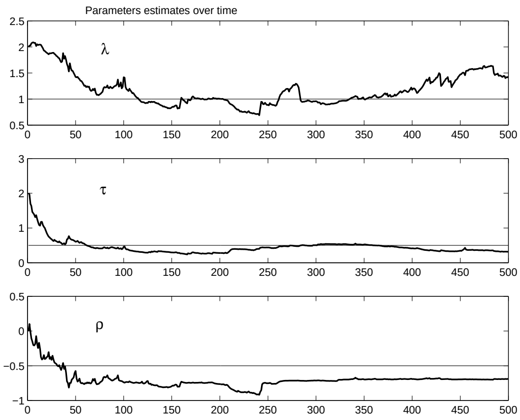

Fig. 2. We plot the mean of the particles corresponding to the parameters λ, τ and ρ over time, for the IPF applied to the system (6)-(7). The straight lines are the actual values. The parameters μ and σ_0 where estimated by the mean and the standard deviation of the observations ($\mu = .2810, \sigma_0 = .2804$).

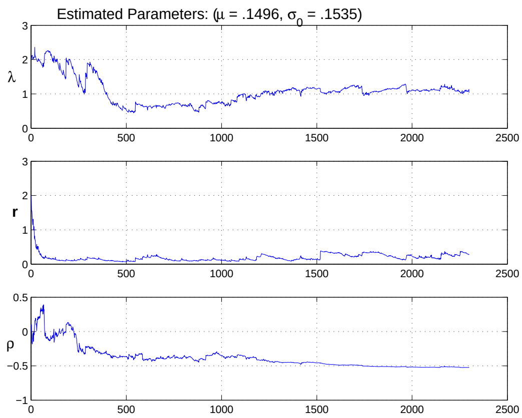

Fig. 3. Parameter Estimates (Industrial Index of Dow Jones). We plot the mean of the particles corresponding to the parameters λ, τ and ρ over time, for the IPF applied to the system (6)-(7).

Parameter ρ, which represents the correlation between the stock price and its volatility, seems remarkably stabe (this was still the case when we added some noise in its mutation step). On the other hand, τ and λ seem stable over some periods but we also notice some jumps. For example, around 550 (January, 1994)), 1500(January, 1998) and 1700(October 1998), we observe a sudden change for both λ and τ. These observations seem to agree with historical and economic facts.

VI. Conclusions

There is mathematical and numerical evidence that the IPF applied to a system where we also include the parameters can be successfully used for to adaptively estimation of the system. This method also seems applicable to the case where the parameters slowly vary or have jumps. In fact, the use of the IPF to detect jumps in the parameters has already been proposed in [5] and [2]. In both cases, though, the new value of the parameter is not estimated.

We also have to note that this method is not applicable to systems of large dimensions. In that case, the algorithm bacomes computationally intractable.

Acknowledgments

The author wishes to thank her phD advisor, René Carmona, for his help and support. It is his insights and suggestions that inspired this work.

References

[1] R. Atar and O. Zeitouni *Exponential Stability for nonlinear filtering*, Ann. Inst. Henri Poincaré, 33(6):697-725, 1997.
[2] B. Azimi-Sadjadi and P.S. Krishnaprasad, *Change Detection for Nonlinear Systems: A Particle Filtering Approach*, To appear in the proceedings of 2002 American Control Conference, ACC2002.
[3] P. Baxendale, P. Chigansky and R. Lipster, *Asymptotic stability of the Wonham filter: ergodic and nonergodic signals*, submitted.
[4] A. Budhiraja, *Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter*, preprint.
[5] F. Cérou, F. LeGland and N.J. Newton, *Stochastic Particle Methods for Linear Tangent Filtering Equations*, Optimal Control and PDE's - Innovations and Applications, in honor of Alain Bensoussan's 60th anniversary, Jos-Luis Menaldi, Edmundo Rofman and Agns Sulem, editors, pp. 231-240, IOS Press, Amsterdam, 2001.
[6] P. Del Moral and L. Miclo, *Branching and Interactive Particle Systems Approximations of Feynman-Kac Formulae with Application to Non-Linear Filtering*, In J. Azéma, M. Émery, M. Ledoux, and M. Yor, editors, Séminaire de Probabilités XXXIV, volume 1729 of Lecture Notes in Mathematics, pages 1-145. Springer-Verlag, Berlin, 2000.
[7] J.-D. Deuschel and D.W. Strook, *Large Deviations*, AMS Chelsea Publishing, Provedence RI, 2001.
[8] C. Andrieu and A. Doucet, *Particle Filtering for Partially Observed Gaussian State Space Models.* J. Royal Statist. Soc. B (methodological), to appear, 2002.
[9] P. Dupuis and R.S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*, John Wiley & Sons, Inc., New York, 1997.
[10] N.J. Gordon, D.J. Salmon and A.F.M. Smith, *Novel approach to nonlinear/non Gaussian Bayesian state estimation*, IEE Proceedings F, 140:107-113, 1993.
[11] G. Kitagawa, *Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models*, Journal on Computational and Graphical Statistics, 5(1):1-25, 1996.
[12] H. Kunita, *Asymptotic behavior of the nonlinear filtering errors of Markov processes*, J. Multivariate Anal., 1:365-393, 1971.
[13] J. Liu and M. West, *Combined parameter and state estimation in simulation-based filtering.* Sequential Monte Carlo in Practice, A. Doucet, N. de Freitas and N. Gordon, Eds., pp.197-223. Springer-Verlag, 2001.
[14] D. Ocone and E. Pardoux, *Asymptotic stability of the optimal filter with respect to its initial condition*, SIAM Journal on Control and Optimization, 34:226-243, 1996.
[15] L. Stettner, *On Invariant Measures of Filtering Processes*, In K. Helmes and N. Kohlmann, editors, Stochastic Differential Systems, Proc. 4th Bad Honnef Conf., Lecture Notes in Control and Inform. Sci., pages 279-292, 1989.
[16] G. Storvik, *Particle Filters for state space models with the presence of unknown static parameters.* IEEE Trans. on Signal Processing, vol. 50, no. 2, pp. 281-289, 2002.

Anastasia Papavasiliou is an assistant professor at the department of Applied Physics and Applied Mathematics of Columbia University. In 2002, she received a phD in Applied Mathematics from Princeton University. She received her BS (Mathematics) in 1997 from the Aristotelian University of Thessaloniki, Greece.