C. R. Acad. Sci. Paris, t. 325, Série I, p. 653-658, 1997
Probabilités/*Probability Theory*
(Statistique/*Statistics*)

Nonlinear filtering: Interacting particle resolution

Pierre DEL MORAL

Laboratoire de Statistiques et Probabilités, CNRS UMR C55830, Bât. 1R1,
Université Paul-Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex, France.
E-mail : delmoral@cict.fr

Abstract. In this Note, we study interacting particle approximations of discrete time and measure valued dynamical systems. Such systems have arisen in such diverse scientific disciplines as in Propagation of Chaos Theory (*see* [12] and [19]), and in Nonlinear Filtering Theory. The main contribution of this Note is to prove the convergences to the optimal filter of such approximations, yielding what seemed to be the first mathematically well-founded convergence results for such approximations of the nonlinear filtering equations. This new treatment was influenced primarily by the development of genetic algorithms (*see* [16] and [3]), and secondarily by the papers of H. Kunita and L. Stettner, [17] and [18] respectively.

Filtrage non-linéaire par systèmes de particules en interaction

Résumé. *Cette Note présente une méthode de résolution particulaire de systèmes dynamiques à valeurs mesures, basée sur la simulation de systèmes de particules en interaction. Ces techniques permettent d'aborder l'étude de certaines équations étudiées dans la Théorie de la propagation du chaos* (voir [12] *et* [19]), *et tout particulièrement les équations du filtrage non-linéaire. Un des principaux résultats de cette Note concerne la convergence vers le filtre optimal de ces approximations particulaires. Ces résultats constituent, à la connaissance de l'auteur, la première démonstration de convergence de telles approximations des équations du filtrage non-linéaire. Cette nouvelle approche a été influencée par le développement des Algorithmes Génétiques* (voir [16] *et* [3]), *et par les articles de H. Kunita et L. Stettner,* [17] *et* [18] *respectivement.*

Version française abrégée

Les bases du Filtrage Non-Linéaire reposent sur la construction de la probabilité conditionnelle à l'information dont on dispose. À l'exception notable de la situation linéaire gaussienne, les filtres optimaux ne peuvent en général être réalisés par un système dynamique en dimension finie (*voir*

Note présentée par Alain BENSOUSSAN.

0764-4442/97/03250653

[4]). D'autre part, la plupart des mtthodes numeriques basées sur I'utilisation de grilles fixes, sur la linéarisation des coefficients du modèle ou sur la recherche du meilleur estimateur linéaire (au sens variance minimum) n'ont jamais réellement permis de resoudre des problemes de filtrage dans lesquels les modèles sont dans des espaces de grande dimension et/ou presentant de fortes non-linéarités.

Les algorithmes de resolution particulaires sont une adaptation dynamique de la loi des grands nombres et de la règle de Bayes (*voir* [9]). Ils ont montré leur habilité à résoudre de tels problemes et tirent leur popularité de la simplicité de leur mise en ceuvre. Pour un panorama détaillé sur la diversite de ces méthodes, nous renvoyons le lecteur à la série d'articles references (*voir* [6], [7], [8] et [20]).

Lorsque les particules sont independantes conditionnellement aux observations, la convergence de telles approximations a été montrée dans [10]. À l'inverse, la description et l'étude des schémas numeriques heuristiques présentés dans [1], [2], [13] et [14], conduisent à la definition d'une large classe des systèmes de particules, que nous avons appelee *systèmes* de *particules en interaction*, par reference à la terminologie utilisée en Mécanique Statistique.

Les textes de reference à la présente Note sont [11] et [12]. Dans la premiere partie, nous presentons et Ctudions la convergence d'une méthode de resolution particulaire de systèmes dynamiques à temps discret et à valeurs mesures, basée sur la simulation de systèmes de particulcs en interaction. D'un point de vue probabiliste, on modélise l'évolution des particules par une chaîne de Markov sur un espace produit et l'évolution de chaque particule depend de la configuration du système à l'ttape précédente. Le choix de ces transitions est dicté par la dynamique locale du modèle à approcher.

Dans la seconde partie, nous motivons l'approche précédente par son application au problème du filtrage non-lidaire. On verra que le cadre general décrit dans la premiere partie permet de mieux comprendre la nature des interactions entre particules.

Notre point de depart sera la formulation markovienne du filtre optimal, décrite par H. Kunita dans [17] et L. Stettner dans [18]. Dans ce cadre, la principale difficulté reside dans l'infini-dimensionalitt de l'espace d'état du système dynamique réalisant le tiltre optimal.

L'ultime partie de cette etude consiste à appliquer les schtmas numeriques développés dans la premiere partie. Les algorithmes stochastiques qui en découlent appartiennent à la classe des algorithmes génétiques introduits par R. Cerf dans [3]. Plus précisément, l'évolution des particules se decompose en deux étapes : 1) une Ctape de mutation/prediction pendant laquelle chaque particule Cvolue de façon independante selon les mêmes probabilités de transitions que celle du signal observe : 2) une Ctape de sélection/ré-actualisation pendant laquelle chaque particule choisit aléatoirement une position parmi celles de la configuration précédente en fonction de l'observation reçue et des vraisemblances de chacun de ces sites.

1. Measure valued processes

Let $(E, \beta(E))$ be a locally compact and separable metric space, endowed with a Bore1 a-field, state space. Denote by P(E) the space of all probability measures on E with the weak topology. The aim of this work is the design of a stochastic particle system approach for the computation of a general discrete time and measure valued dynamical system given by:

$$\eta_n = \phi(n, \eta_{n-1}) \qquad \forall n \geq 1, \qquad \eta_0 = \eta, \tag{1}$$

where $\eta \in$ P(E) and $\phi(n, .\ .)$: P(E) $\rightarrow$ P(E), $n \geq 1$, are continuous functions. Such systems have arisen in such diverse scientific disciplines as physics (see [19], and the references given there), nonlinear economic modelling and signal processing (see [17] and [18]). Solving (1) is in general an

enormous task as it is nonlinear and usually involves integrations over the whole space E. To obtain a computationnally feasible solution, some kind of approximation is needed. Our major motivation is from advanced signal processing applications in engineering and particularly in optimal nonlinear filtering problems. Recall that this consists in computing the conditional distribution of internal states in dynamical systems, when partial observations are made, and random perturbations are present in the dynamics as well as in the sensors. With the notable exception of the linear-Gaussian situation or wider classes of models (*see* [5], and the references given there), optimal filters have no finitely recursive solution (*see* [4]). Although Kalman filtering is a popular tool in handling estimation problems, its optimality heavily depends on linearity. When used for nonlinear filtering (Extended Kalman Filter), its performance relies on and is limited by the linearizations performed on the concerned model. Nevertheless, we will see that the dynamics structure of such conditional distributions can be viewed, under mild assumptions, as a special case of (1).

Let us introduce some notations. By $\mathcal{C}_b(E)$ we denote the subspace of all bounded continuous functions. Let $\mu \in \mathcal{P}(E)$, $f \in \mathcal{C}_b(E)$, and let K be a Markov kernel. We will use the standard notations $Kf(x) = \int K(x, dy)\, f(y)$ and $\mu f = \int f(x)\, \mu(dx)$.

A transition probability kernel K on E is said to be *Feller* if $K(\mathcal{C}_b(E)) \subset \mathcal{C}_b(E)$. It is well known that $\mathcal{P}(E)$ with the topology of weak convergence can be considered as a separable metric space. By $\beta(\mathcal{P}(E))$ we denote the σ-algebra of Borel subsets in $\mathcal{P}(E)$, by $\mathcal{C}_b(\mathcal{P}(E))$ the subspace of all bounded continuous functions, and by $\mathcal{P}(\mathcal{P}(E))$ the collection of all probability measures on $(\mathcal{P}(E), \beta(\mathcal{P}(E)))$ with the induced weak topology. As usually, for an $F \in \mathcal{C}_b(\mathcal{P}(E))$, M a Feller transition on $\mathcal{P}(E)$, and $\Phi \in \mathcal{P}(\mathcal{P}(E))$, we write $MF(\mu) = \int M(\mu, d\nu)\, F(\nu)$ and $\Phi F = \int F(\mu)\, \Phi(d\mu)$.

2. Interacting particle systems

The particle system under study will be a Markov chain with state space $E^{\otimes N} = E \times \cdots \times E$, where $N \geq 1$ is the size of the system. The N-tuple of elements of E, *i.e.* the points of the set $E^{\otimes N}$, are called *particle systems*, and will be mostly denoted by the letters x and z. With $\nu \in \mathcal{P}(E)$ we associate the N-fold product $\nu^{\otimes N} \stackrel{\text{def}}{=} \nu \times \cdots \times \nu \in \mathcal{P}(E^N)$ of the measure ν, and we note $m^N(x) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{j=1}^N \delta_{x^j}$ the empirical measure associated to the point $x = (x^1, \ldots, x^N)$. Let us consider an N-particle system, which is a Markov chain $(\zeta_n)_{n \geq 0}$ on $E^{\otimes N}$ generated by the transition probability kernels $(\mathcal{K}_n^N)_{n \geq 1}$, given by:

$$(2) \qquad \forall f \in \mathcal{C}_b(E^{\otimes N}), \quad \forall x \in E^{\otimes N} \quad \forall n \geq 1, \qquad \mathcal{K}_n^N f(x) = \int_{E^{\otimes N}} \phi^{\otimes N}(n, m^N(x))(dz)\, f(z).$$

We assume the initial distribution is the N-fold product $\eta^{\otimes N}$ of the initial measure η of the dynamical system (1). By $(\Omega, (F_n)_{n \geq 0}, P)$ we denote the canonical space for the N-particle system $(\zeta_n)_{n \geq 0}$. It is clear from the construction above that $\zeta_n = (\zeta_n^1, \ldots, \zeta_n^N)$ can be viewed as a system of N particles with nonlinear interaction function $\phi(n, m^N(\zeta_{n-1}))$. The algorithm constructed in this way will be called an *interacting particle approximation* of (1). The terminology *interacting* is intended to emphasized that the particles are not independent and the evolution of an individual particle is no longer Markovian. Nevertheless, the system of particles is Markovian. Much more is true, the above description enables us to consider the particle density profiles $\eta_n^N \stackrel{\text{def}}{=} m^N(\zeta_n)$ as a time inhomogeneous Markov process $(\eta_n^N)_{n \geq 0}$ on $\mathcal{P}(E)$, with initial distribution $M_0^N \stackrel{\text{def}}{=} M_0 C_N$, and generated by the transition probability kernels $M_n^N \stackrel{\text{def}}{=} M_n C_N$, $n \geq 1$, defined by:

$$(3) \qquad M_n F(\nu) = F(\phi(n, \nu)), \qquad C_N F(\nu) = \int_{E^{\otimes N}} F(m^N(x))\, \nu^{\otimes N}(dx)$$

for every $F \in \mathcal{C}_b(\mathcal{P}(E))$, $n \geq 0$, and $\nu \in \mathcal{P}(E)$, with the convention $\phi(0, \nu) = \eta$ for all $\nu \in \mathcal{P}(E)$.

The crucial question is of course whether η_n^N converges to η_n as N is growing. We show hereafter a slightly more general result.

THEOREM 1. *Let us suppose that E is compact. Let $(M_n)_{n\geq 1}$ denote a series of time-inhomogeneous and Feller Markov transitions on P(E), $M_0 \in \mathcal{P}(\mathcal{P}(E))$, and let $M_n^N = M_n C_N$, $n \geq 0$. Using these notations, we have:*

$$\lim_{N\to+\infty} M_0^N M_1^N \cdots M_n^N F = M_0 M_1 \cdots M_n F \qquad \forall M_0 \in \mathcal{P}(\mathcal{P}(E)), \qquad \forall F \in \mathcal{C}_b(\mathcal{P}(E)). \tag{4}$$

More generally, (4) holds when E is locally compact and $(M_n)_{n\geq 1}$ is such that $M_n(\mathcal{U}(\mathcal{P}(E)) \subset \mathcal{U}(\mathcal{P}(E))$, where $\mathcal{U}(\mathcal{P}(E))$ denotes the space of all bounded and uniformly continuous functions F : P(E) $\to \mathbb{R}$.

Our next objective is to introduce an additional assumption with regard to the functions $\phi(n, \ldots)$ which enables us to develop some useful theorem. In a little while, we will see one way in which this result may be applied in nonlinear filtering problems.

THEOREM 2. *Suppose that for every $f \in \mathcal{C}_b(E)$, $\nu \in$ P(E), and $n \geq 1$, there exist some constant $C_n(\nu, f)$ and a finite set of bounded functions $\mathcal{H}_n(\nu, f)$ such that*

$$\forall \mu \in P(E) \qquad |\phi(n,\nu)f - \phi(n,\mu)f| \leq C_n(\nu, f) \sum_{h\in\mathcal{H}_n(\nu,f)} |\nu h \quad \mu h|. \tag{5}$$

Then for every $f \in \mathcal{C}_b(E)$ and $n \geq 1$, there exists $A_n(f) < +\infty$ such thut

$$E(|\eta_n^N f - \eta_n f|^2) \leq \frac{A_n(f)}{N}. \tag{6}$$

3. Measure valued processes associated to the nonlinear filtering problem

The object of this section is to introduce the filtering model in such a way that the techniques of section 3 can be applied. The basic model for the general Nonlinear Filtering problem consists in a time inhomogeneous Markov process X and a nonlinear observation Y with observation noise V. Namely, let (X, Y) be the Markov process taking values in $S \times \mathbb{R}^d$, $d \geq 1$, and defined by the system:

$$X = (X_n)_{n\geq 0}, \qquad Y_n = h_n(X_n) + V_n, \qquad n \geq 1, \tag{7}$$

where S is a locally compact and separable metric space, $h_n : S \to \mathbb{R}^d$, $d \geq 1$, are Bore1 measurable functions, and V_n are independent random variables with continuous and positive density g_n with respect to Lebesgue measure. The signal process X that we consider is assumed to be a non-inhomogeneous and S-valued Markov process with Feller transition probability kernel K,,, $n \geq 1$, and initial probability measure ν on S. We will assume the observation noise V and X are independent. The interacting particle systems approach developed hereafter can be seen as a nonlinear filtering method which discards linearizations. More precisely, these techniques use the nonlinear system model itself in order to solve the filtering problem. The evolution of this material may be seen quite directly through the following chain of papers [I], [2], [13], [14], and [15]. Nevertheless, in these papers this method is applied as an heuristic approximation to specific models, its general nature is not emphasized and experimental simulations are the only guides for handling concrete problems.

A natural idea in our settings is to study the distribution of the pair process (X,, . X_{n+1}) conditionally to the observations up to time n. Namely.

$$\eta_n(f) \stackrel{\text{def}}{=} E(f(X_n, X_{n+1})/Y_1, \ldots, Y_n) \qquad \forall f \in \mathcal{C}_b(S^2), \qquad \forall n \geq 0. \tag{8}$$

The following result gives a dynamical system formulation of the conditional distribution $(\eta_n)_{n\geq 0}$.

LEMMA (*see* [17] and [18]). – *Given the observations* $Y = y$, *the distributions* (8) *are solution of the* $\mathcal{P}(S^2)$*-valued dynamical system*

$$\begin{cases} \eta_n = \phi_n(y_n, \eta_{n-1}), & n \geq 1, \\ \eta_0 = \nu \times K_1, \end{cases} \tag{9}$$

where $\nu \times K_1$ *is the distribution of the pair variable* (X_0, X_1), *and* $\phi_n(y_n, \ldots) : \mathcal{P}(S^2) \longrightarrow \mathcal{P}(S^2)$, $n \geq 1$, $y_n \in \mathbb{R}^d$, *are the continuous function given by:*

$$\forall n \geq 1, \quad \forall \mu \in \mathcal{P}(S^2), \quad \forall f \in \mathcal{C}_b(S^2), \qquad \phi_n(y_n, \mu)f = \frac{\mu(\overline{g}_n\, T_n f)}{\mu(\overline{g}_n)},$$

with $\overline{g}_n(x_0, x_1) = g_n(y_n - h_n(x_1))$ *and* $T_n f(x_0, x_1) = \int f(x_1, x_2)\, K_{n+1}(x_1, dx_2)$ *for all* $x_0, x_1 \in S$.

When considering the nonlinear filtering equation (9), we can prove that condition (5) is satisfied. The dynamics of the corresponding interacting particle system depends on the measurements $Y = y$, and they are modeled by a time-inhomogeneous Markov chain $(\zeta_n)_{n\geq 0}$ with state space $(S^2)^{\otimes N}$, $N \geq 1$, with initial distribution $\eta_0^{\otimes N}$, and generated by the familly of transition probability kernels $(\mathcal{K}^N_{n,y_n})_{n\geq 1}$ given by:

$$\begin{cases} \forall f \in \mathcal{C}_b(E^{\otimes N}), \quad \forall x \in E^{\otimes N}, \quad \forall n \geq 1, \\ \mathcal{K}^N_{n,y_n} f(x) = \int_{E^{\otimes N}} \phi_n^{\otimes N}(y_n, m^N(x))(dz)\, f(z). \end{cases} \tag{10}$$

The approximation of the conditional distribution η_n by the empirical measure $\eta_n^N \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_n^i}$ is guaranteed by theorem 2. To be more precise, it is convenient to introduce additional notations. Let us set $\zeta_n \overset{\text{def}}{=} (\widehat{\xi}_n, \xi_{n+1}) \in S^{\otimes N} \times S^{\otimes N}$, $n \geq 0$. In the special case where the state space S is finite, we have from (10):

$$P_{[y]}(\xi_n = x / \widehat{\xi}_{n-1} = z) = \prod_{p=1}^N K_n(z^p, x^p),$$

$$P_{[y]}(\widehat{\xi}_n = x / \xi_n = z) = \prod_{p=1}^N \sum_{i=1}^N \frac{g_n(y_n - h_n(z^i))}{\sum_{j=1}^N g_n(y_n - h_n(z^j))}\, 1_{z^i}(x^p),$$

where y_n is the observation data at the time n, $x = (x^1, \ldots, x^N)$ and $z = (z^1, \ldots, z^N) \in S^{\otimes N}$. Probabilistically, and in a more precise language, we may describe its evolution in time as follows:

1. At the time $n = 0$, the particle system consists of N independent random particles $\widehat{\xi}_0^1, \ldots, \widehat{\xi}_0^N$ with common law ν.
2. At the time $n \geq 1$, the transition $\widehat{\xi}_{n-1} \rightsquigarrow \widehat{\xi}_n$ is decomposed in two mechanisms $\widehat{\xi}_{n-1} \rightsquigarrow \xi_n$ and $\xi_n \rightsquigarrow \widehat{\xi}_n$

(a) **Prediction/Mutation:** Before the updating mechanism, each particle evolves independently of each other according the transition probability kernel of the signal process.

(b) **Updating/Selection:** When the observation $Y_n = y_n$ is received, each particle *examines* the system of particles $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$ and chooses randomly a site ξ_n^i with probability

$$\frac{g_n(y_n - h_n(\xi_n^i))}{\sum_{j=1}^N g_n(y_n - h_n(\xi_n^j))}.$$

What is important is that each particle interacts selectively with the system in accordance with the observations data. Roughly speaking, the updating mechanism stabilizes the particles' motion around certain values of the real signal which are determined by the observations thus providing a well behave adaptative grid.

Work partly supported by CEC Contract No ERB-FMRX-CT96-0075 and INTAS-RFBR 95-0091.

Note remise le 13 juin 1997, acceptée le 16 juin 1997.

References

[1] **Carvalho H., Del Moral P., Monin A. and Salut G., 1997.** Optimal nonlinear filtering in GPS/INS Integration, *IEEE trans. on Aerospace and electronic systems*, Vol. 33, No. 3, pp. 835-850.

[2] **Carvalho H., 1995.** Filtrage optimal non-linéaire du signal GPS NAVSTAR en racalage de centrales de navigation, *Thèse* de l'École Nationale Supérieure de l'Aéronautique et de l'Espace.

[3] **Cerf R., 1994.** Une théorie asymptotique des Algorithmes Génétiques, *Université Montpellier* II, Sciences et techniques du Languedoc.

[4] **Chaleyat-Maurel M. and Michel D., 1983.** Des résultats de non-existence de filtres de dimension finie, *C. R. Acad. Sci. Paris*, Série I, t. 296.

[5] **Crisan D., 1996.** The Nonlinear Filtering Problem, *Ph.D. Disseration*, University of Edinburgh.

[6] **Crisan D. and Lyons T. J., 1996.** Nonlinear Filtering And Measure Valued Processes, Imperial College London, *preprint.*

[7] **Crisan D. and Lyons T. J., 1996.** Convergence of a Branching Particle Method to the Solution of the Zakai Equation, Imperial College London, *preprint.*

[8] **Crisan D., Gaines J. and Lyons T. J., 1996.** A Particle Approximation of the Solution of the Kushner-Stratonovitch Equation, Imperial College London, *preprint.*

[9] **Del Moral P. and Salut G., 1995.** Filtrage non-linéaire : résolution particulaire à la Monte-Carlo, *C. R. Acad. Sci. Paris*, Série I, t. 320.

[10] **Del Moral P., 1995.** Nonlinear filtering using random particles, *Theor. Prob. Appl.*, Vol. 40, (4).

[11] **Del Moral P., 1996.** Nonlinear Filtering: Interacting particle resolution, *Markov Processes and Related Fields*, Vol. 2, pp. 555-579.

[12] **Del Moral P.** Measure Valued Processe and Interacting Particle Systems. Application to Nonlinear Filtering Problems, *Publications du Laboratoire de Statistiques et Probabilités*, Université Paul Sabatier, No 15-96.

[13] **Del Moral P. and Salut G., 1997.** Particle interpretation of nonlinear filtering and optimization, *Russian J. Math. Phy.* in press.

[14] **Del Moral P., Noyer J. C. and Salut G., 1995.** Résolution particulaire et traitement non-linéaire du signal : Application Radar/Sonar, *Traitement du signal.*

[15] **Gordon N. J., Salmon D. J. and Smith A. I. M., 1993.** Novel Approach to non-linear/non-Gaussian Bayesian state estimation, *IEE*

[16] **Holland J. H., 1975.** *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor.

[17] **Kunita H., 1971.** Asymptotic Behavior of Nonlinear Filtering Errors of Markov Processes, *J. Multivariate Anal.*, Vol. 1, No 4, p. 365-393.

[18] **Stettner L., 1989.** *On Invariant Measures of Filtering Processes*, Stochastic Differential Systems, Lecture Notes in Control and Information Sciences, 126, Springer Verlag.

[19] **Sznitman A. S., 1989.** Topics in propagation of chaos, *École d'été de Probabilités de Saint-Flour* XIX-1989, Lecture Notes in Mathematics, No 1464.

[20] **Van Dootingh M., Viel F., Rakotopara D. and Gauthier J. P., 1991.** Coupling of nonlinear control with a stochastic filter for state estimation : Application on a free radical polimerization reactor, *I.F.A.C. International Symposium* ADCHEM'91, Toulouse, France, October 14-15.