Interacting Particle Approximations of Non Linear Filtering and Smoothing

Pierre Del Moral

LSP-CNRS-UMR C55830, Bat. 1R1, Univ. P. Sabatier, 118 Route de Narbonne, 31062 Toulouse, France, delmoral@cict.fr

In this short note we briefly introduce the reader to the recently developed interacting particle approximating models of non linear filtering, smoothing and path estimation problems. For a more precise discussion, the interested reader may consult [1, 2] and references therein. We begin with some standard notation and we write

$$\eta K(f) = \int \eta(dx)\ K(x,dy)\ f(y) \quad \text{and} \quad K(f)(x) = \int K(x,dy)\ f(y)$$

for a bounded measurable function $f \in \mathcal{B}_b(E)$, a probability measure $\eta \in \mathcal{P}(E)$ and for a Markov transition $K(x,dy)$ on a measurable space $(E,\mathcal{E})$.

Non linear filtering: Our goal is to introduce a general and abstract non linear filtering model. We let (X,Y) stand for a Markov process taking values in a product space $(E \times F, \mathcal{E} \otimes \mathcal{F})$ with transition kernels of the form

$$\begin{array}{l} \mathbb{P}\left[(X,Y)_n \in d(x,y)_n | (X,Y)_{n-1} = (x,y)_{n-1}\right] \\ = \overline{g}_n(x_n,y_n)\ K_n(x_{n-1},dx_n)\ \theta_n(dy_n) \end{array} \tag{1}$$

where $d(x,y)_n = dx_n \times dy_n$ denotes an infinitesimal neighborhood of the point (x_n,y_n), θ_n is a probability measure on $(F,\mathcal{F})$, $\overline{g}_n : E \times F \to (0,\infty)$ and K_n is a Markov transition on $(E,\mathcal{E})$.

To illustrate this abstract model suppose the observation process is given by $Y_n = H_n(X_n,V_n)$ where random perturbations V_n are i.i.d. random variables (independent of the signal X) taking values in an auxiliary measurable space $(G,\mathcal{G})$ and $H_n : E \times G \to F$. In this situation (1) is met with

$$\mathbb{P}\left(H_n(x,V_n) \in dy\right) = \overline{g}_n(x,y)\ \theta_n(dy)$$

For a more concrete situation assume that $F = \mathcal{V} = \mathbb{R}^d$, $H_n(x,v) = h_n(x)+v$, $h_n : E \to \mathbb{R}^d$ and the law of each V_n has a density $p_n(v)$. In this situation we have $\overline{g}_n(x,y) = p_n(y-h_n(x))/p_n(y)$ and $\theta_n(dy) = p_n(y)dy$. The non linear filtering problem consists in computing the one step predictor η_n and the optimal filter $\widehat{\eta}_n$ defined by the conditional distributions

$$\begin{aligned} \eta_n(dx_n) &= \mathbb{P}(X_n \in dx_n | Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}) \\ \widehat{\eta}_n(dx_n) &= \mathbb{P}(X_n \in dx_n | Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}, Y_n = y_n) \end{aligned} \tag{2}$$

We recall that their time evolution involves a two-step mechanism

$$\begin{aligned}
&\eta_n \xrightarrow{\text{Updating}} \widehat{\eta}_n = \Psi_n(\eta_n) \xrightarrow{\text{Prediction}} \eta_{n+1} = \widehat{\eta}_n K_{n+1} = \Phi_{n+1}(\eta_n)\\
&\text{with} \quad \Psi_n(\eta_n)(dx_n) = \frac{1}{\eta_n(g_n)}\, g_n(x_n)\ \eta_n(dx_n)\,, \quad g_n(x_n) = \overline{g}_n(x_n, y_n)\\
&\text{and} \quad \Phi_{n+1}(\eta) = \Psi_n(\eta) K_{n+1}
\end{aligned}$$

Non linear and measure valued equations: The N-interacting particle systems approximating model associated to a given abstract non linear and measure valued processes

$$\eta_n = \Phi_n(\eta_{n-1}) \quad \text{with} \quad \Phi_n : \mathcal{P}(E) \longrightarrow \mathcal{P}(E)$$

is defined as the Markov chain $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$, $n \geq 0$, on the product space E^N with transition kernels

$$\mathbb{P}(\xi_n \in dz | \xi_{n-1} = x) = \prod_{p=1}^{N} \Phi_n\left(m(x)\right)(dz^p) \quad \text{where} \quad m(x) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x^i}$$

Here $dz = dz^1 \times \ldots \times dz^N$ stands for an infinitesimal neighborhood of the point $z = (z^1, \ldots, z^N)$, $x = (x^1, \ldots, x^N) \in E^N$ and δ_x^i denotes the Dirac measure at $x^i \in E$. In other words, given the configuration $(\xi_{n-1}^1, \ldots, \xi_{n-1}^N)$ at time $(n-1)$ the next system consists of N independent and identically distributed particles $(\xi_n^1, \ldots, \xi_n^N)$ with the same law $\Phi_n\left(m(\xi_{n-1})\right)$, that is

$$(\xi_n^1, \ldots, \xi_n^N) \ \text{N i.i.d.} \ \sim \ \Phi_n\left(m(\xi_{n-1})\right)$$

Genetic algorithms: As a particular case, in filtering settings we have that

$$\forall x \in E^N\,, \qquad \Phi_{n+1}\left(m(x)\right) = \sum_{i=1}^{N} \frac{g_n(x^i)}{\sum_{j=1}^{N} g_n(x^j)}\ K_{n+1}(x^i, .)$$

From the above one easily sees that the N-particle model is a genetic type algorithm which evolves according to two separate mechanisms

$$\xi_n \xrightarrow{\text{Selection}} \widehat{\xi}_n \xrightarrow{\text{Mutation}} \xi_{n+1}$$

In the selection stage one updates the positions in accordance with the likelihood function g_n, so that each particle examines the system $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$ and chooses randomly a site ξ_n^i with a probability $g_n(\xi_n^i)/\sum_{j=1}^{N} g_n(\xi_n^j)$. This mechanism is called the selection transition as the particle are selected for reproduction, the most fit individual being more likely to be selected. During the mutation transition each particle evolves randomly according to the same transitions as the signal, that is

$$\mathbb{P}\left(\xi_{n+1} \in dz | \widehat{\xi}_n = x\right) = \prod_{p=1}^{N} K_{n+1}(x^p, dz^p)$$

Non linear path estimation/smoothing: It turns out that our abstract and general formulation contains as a special case, non linear smoothing and path estimation models. To describe precisely this point, let X'_n be a Markov chain on a given measure space $(E', \mathcal{E}')$. We define the random sequences

$$X_n = (X_{p,n})_{p\geq 0} \in E = (E')^{\mathbb{N}} \qquad \text{with} \qquad X_{p,n} = \begin{cases} X'_p & \text{if} \quad p \leq n \\ X'_n & \text{if} \quad p \geq n \end{cases}$$

The key observation is that we have constructed in this way a Markov chain with transitions defined for any test function $f_n : x = (x_p)_{p\geq 0} \in E \mapsto f_n(x) = f(x_0, \ldots, x_n) \in \mathbb{R}$, $f \in \mathcal{B}_b((E')^{n+1})$, by setting

$$K_n(f_n)(x) = \mathbb{E}(f_n(X_n)|X_{n-1} = x) = \int_{E'} f(x_0, \ldots, x_{n-1}, u) \; K'_n(x_{n-1}, du)$$

From the above we easily concludes that the one step predictor and the optimal filter (2) represent the conditional distributions on path space

$$\begin{aligned} \eta_n(f_n) &= \mathbb{E}(f(X'_0, \ldots, X'_n)|Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}) \\ \widehat{\eta}_n(f_n) &= \mathbb{E}(f(X'_0, \ldots, X'_n)|Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}, Y_n = y_n) \end{aligned}$$

Genealogy and historical process: Now genetic algorithms can be revisited somehow. In path space framework the particle approximating model consists of N path-particle

$$\xi^i_n = (\xi^i_{0,n}, \ldots, \xi^i_{n,n}) \in (E')^{n+1} \qquad \text{and} \qquad \widehat{\xi}^i_n = (\widehat{\xi}^i_{0,n}, \ldots, \widehat{\xi}^i_{n,n}) \in (E')^{n+1} \tag{3}$$

During the selection stage and with some obvious abusive notations we see that the N-path/particles $\widehat{\xi}'_n = (\widehat{\xi}^1_n, \ldots, \widehat{\xi}^N_n)$ are chosen independently with the distribution

$$\Psi_n(m(\xi_n)) = \sum_{i=1}^{N} \frac{g_n(\xi^i_{0,n}, \ldots, \xi^i_{n,n})}{\sum_{j=1}^{N} g_n(\xi^j_{0,n}, \ldots, \xi^j_{n,n})} \; \delta_{(\xi^i_{0,n}, \ldots, \xi^i_{n,n})}$$

The mutation stage simply consists in extending each path

$$\widehat{\xi}^i_n = (\widehat{\xi}^i_{0,n}, \ldots, \widehat{\xi}^i_{n,n}) \longrightarrow \xi^i_{n+1} = ((\widehat{\xi}^i_{0,n}, \ldots, \widehat{\xi}^i_{n,n}), \xi^i_{n+1,n+1})$$

with an elementary move $\widehat{\xi}^i_{n,n} \to \xi^i_{n+1,n+1}$ with transition $K'_{n+1}(\widehat{\xi}^i_{n,n}, .)$. If we use a graphical representation we easily see that the set of all individuals and vertices defined formally by setting $\xi^i_{0,n} \longrightarrow \ldots \longrightarrow \xi^i_{n,n}$ represents the complete genealogy of the population $\xi_{n,n} = (\xi^i_{n,n})_{1\leq i\leq N}$. If the functions $g_n(x_0, \ldots, x_n) = g_n(x_n)$ only depend on the n-coordinate then the end-points configurations $\xi_{n,n}$, $\widehat{\xi}_{n,n}$ also form a selection/mutation genetic algorithm and the previously defined path-space model represents its historical process.

$\mathbb{L}_p$-uniform estimates: For any $n \geq 0$ and $p \geq 1$ there exists some finite constants $b(p)$ et $c(n)$ such that for any $f \in \mathcal{B}_b(E)$, $\|f\| \leq 1$,

$$\mathbb{E}(|\eta_n^N(f) - \eta_n(f)|^p)^{1/p} \leq \frac{1}{\sqrt{N}}\, b(p)\; c(n)\,, \quad \text{with} \quad \eta_n^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_n^i} \tag{4}$$

If we use the following regularity and mixing conditions

$$(\mathcal{H}):\ \frac{g_n(x)}{g_n(y)} \geq \epsilon_n(g) > 0\,,\ K_n(x,.) \sim K_n(y,.) \text{ with } \frac{dK_n(x,.)}{dK_n(y,.)} \geq \epsilon_n(k) > 0$$

with $\epsilon(g) = \inf_n \epsilon_n(g) > 0$ and $\epsilon(k) = \inf_n \epsilon_n(k) > 0$ then we have that

$$\sup_{n\geq 0} \mathbb{E}(|\eta_n^N(f) - \eta_n(f)|^p)^{1/p} \leq \frac{1}{\sqrt{N}}\, b(p)\ \epsilon^{-2}(g)\epsilon^{-4}(k)$$

Increasing propagation of chaos: For any $0 \leq q \leq N$ we define

$$\mathbb{P}_{[0,n]}^{(N,q)} = \text{Law}(\xi_{[0,n]}^1, \ldots, \xi_{[0,n]}^q) \qquad \mathbb{P}_n^{(N,q)} = \text{Law}(\xi_n^1, \ldots, \xi_n^q)$$

with $\xi_{[0,n]}^i = (\xi_0^i, \ldots, \xi_n^i)$. Increasing propagation of chaos estimates allow to calibrate the independence degree between particles. For instance there exists some finite constant $c(n)$ such that

$$\|\mathbb{P}_n^{(N,q)} - \eta_n^{\otimes q}\|_{tv} \leq \frac{\exp\left(c(n)q\right)}{N} \quad \text{with} \quad \eta_n^{\otimes q} = \underbrace{\eta_n \otimes \ldots \otimes \eta_n}_{\text{q-times}} \tag{5}$$

In addition, if $K_n(x,.) \sim \eta_n$ with $\sup_x dK_n(x,.)/d\eta_n \in \mathbb{L}_2(\eta_n)$ then we have the entropy estimate

$$\text{Ent}\left(\mathbb{P}_{[0,n]}^{(N,q)} | \eta_{[0,n]}^{\otimes q}\right) \leq \frac{c(n)q}{N} \tag{6}$$

If $(\mathcal{H})$ is met with $\inf_n \epsilon_n(g) > 0$ and $\inf_n \epsilon_n(k) > 0$ then (6) holds with $c(n) = c.n$, $c < \infty$ In this situation, we obtain an increasing propagation of chaos for particle block size $q = q(N)$ and time horizon $n = n(N)$

$$\lim_{N\to\infty} n(N)q(N)/N = 0 \implies \lim_{N\to\infty} \text{Ent}\left(\mathbb{P}_{[0,n(N)]}^{(N,q(N))} | \eta_{[0,n(N)]}^{\otimes q(N)}\right) = 0$$

References

[1] P. Del Moral, L. Miclo. Genealogies and increasing propagation of chaos for Feynman-Kac and genetic models. *The Annals of Applied Probability*, to appear 2001.

[2] P. Del Moral, L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. *Séminaire de probabilité XXXIV, Lecture Notes in Mathematics No. 1729, Eds. J. Azéma, M. Emery, M. Ledoux, M. Yor, Springer, pp. 1-145, 2000.*