

## Compréhension des données

**Article DPF (Partie 2) :** L'idée ici est de passer par le principe de transport optimal afin d'améliorer l'étape de reparamétrisation. En minimisant la distance Wasserstein-2 entre 2 mesures de probabilités, on arrive à obtenir le plan de transport optimal (T.O.) qui permet de passer de l'une à l'autre et vice-versa.

Il est alors possible d'utiliser cela avec des poids et des mesures de probabilités atomiques pour trouver une matrice  $P$  qui minimise la somme des normes au carré des atomes, tout en vérifiant que la somme des termes d'une ligne ou d'une colonne quelconque donne le poids correspondant. On peut également raisonner par le problème dual de maximisation.

Le "resampling" à partir du transport repose à la fois sur le poids des particules mais aussi sur leur position. On a des approximations particulières avec nos mesures de probabilités atomiques, et le plan de T.O. prend la forme d'une carte déterministe qui peut être vue comme la transformation d'ensemble (T.E.). Cette approche nécessite de résoudre le problème linéaire vu plus haut à un coût élevé et la T.E. résultante n'est pas différentiable.

**Partie 3 :** L'utilisation de T.O. avec régularisation de l'entropie (R.E.) permet de créer une matrice de transport différentiable et moins "cher" en termes de complexité. On utilise donc une autre version de la distance Wasserstein-2, et en la minimisant on retombe de nouveau sur un plan de transport optimal. On peut également passer par le problème dual pour récupérer la matrice de transport régularisée.

Le dual peut être maximisé grâce à l'algorithme de Sinkhorn, qui donnent des solutions différentiables par rapport aux paramètres d'entrée. Les dérivées du plan de T.O. sont donc accessibles en combinant ces solutions avec la matrice de transport régularisée. En utilisant un nouvel algorithme fourni par l'article avec le T.O. avec R.E., on obtient une T.E. qui est bien différentiable calculée à complexité  $O(N^2)$ .

Ainsi la technique pour un filtrage particulière différentiable consiste à utiliser l'algorithme standard pour un filtrage particulière, mais en utilisant l'astuce de reparamétrisation vu dans la partie 2, tout en y implémentant la T.E. différentiable

## Compréhension du filtrage particulaire (Partie 1) :

### I) Modèle d'états d'espace

Ils présentent typiquement des équations telles que :

$$\begin{aligned}\mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \\ \mathbf{y}_k &= \mathbf{g}(\mathbf{x}_k, \mathbf{v}_k)\end{aligned}$$

où  $\mathbf{x}_k$  est le vecteur d'état au temps  $k$ ,  $\mathbf{u}_k$  est le vecteur de contrôle d'entrée au temps  $k$ ,  $\mathbf{w}_k$  est le vecteur du processus de perturbation au temps  $k$ ,  $\mathbf{y}_k$  est le vecteur d'observation de sortie au temps  $k$  et  $\mathbf{v}_k$  est le vecteur de mesure du bruit. On ne peut pas directement observer  $\mathbf{x}_k$  malheureusement, on fait des estimations à partir des  $\mathbf{y}_k$ . Les  $\mathbf{w}_k$  et  $\mathbf{v}_k$  sont inconnus, mais en revanche on connaît parfaitement  $\mathbf{u}_k$ . Pour faciliter la compréhension, on travaillera ici sur ce modèle :

$$\begin{aligned}\mathbf{x}_k &= A\mathbf{x}_{k-1} + B\mathbf{u}_{k-1} + \mathbf{w}_{k-1} \\ \mathbf{y}_k &= C\mathbf{x}_k + \mathbf{v}_k\end{aligned}$$

On va considérer que  $\mathbf{w}_k$  est un vecteur aléatoire gaussien de moyenne 0 et de matrice de covariance  $Q$ , et  $\mathbf{v}_k$  l'est également mais avec une matrice de covariance  $R$ . Comme  $\mathbf{w}_k$  est aléatoire pour tout  $k$ ,  $\mathbf{x}_k$  le sera aussi. On aura donc des probabilités liées au prochain état :  $p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_k)$ .

La fonction de distribution pour la transition d'état est une distribution normale à pour moyenne  $A\mathbf{x}_{k-1} + B\mathbf{u}_{k-1}$  et  $Q$  pour matrice de covariance. On a donc :

$$\begin{aligned}p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_{k-1}) &= \\ &= (2\pi)^{-\frac{n}{2}} \det(Q)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_k - A\mathbf{x}_{k-1} - B\mathbf{u}_{k-1})^T Q^{-1} (\mathbf{x}_k - A\mathbf{x}_{k-1} - B\mathbf{u}_{k-1})}\end{aligned}$$

De manière similaire, pour  $\mathbf{y}_k$  on aura  $p(\mathbf{y}_k | \mathbf{x}_k)$  et en utilisant l'équation de départ et les informations sur  $\mathbf{v}_k$  on a :

$$p(\mathbf{y}_k | \mathbf{x}_k) = (2\pi)^{-\frac{r}{2}} \det(R)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{y}_k - C\mathbf{x}_k)^T R^{-1}(\mathbf{y}_k - C\mathbf{x}_k)}$$

On travaille ici dans un modèle d'états d'espace markovien, c'est à dire qu'on a l'équation suivante vérifiée :

$$p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \dots, \mathbf{x}_0) = p(\mathbf{y}_k | \mathbf{x}_k)$$

## II) Formulation du problème

Le but est d'estimer l'état  $\mathbf{x}_k$ ,  $\mathbf{y}_k$  à partir des données  $\mathbf{x}_0$  à  $\mathbf{x}_{k-1}$ ,  $\mathbf{y}_0$  à  $\mathbf{y}_k$  et  $\mathbf{u}_0$  à  $\mathbf{u}_{k-1}$  en général dans ce genre de problèmes. Cependant, les filtres particuliers quant à eux, visent à donner une loi de distribution pour les  $\mathbf{x}_k$  dépendant des  $\mathbf{y}_k$  et  $\mathbf{u}_{k-1}$ . On veut donc avoir  $p(\mathbf{x}_k | \mathbf{y}_{0:k}, \mathbf{u}_{0:k-1})$ . Dans ce genre d'études, les particules sont des états  $\mathbf{x}_k^{(i)}$  de poids  $w_k^{(i)}$ .

En connaissant les particules, on a :

$$p(\mathbf{x}_k | \mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}) \approx \sum_{i=1}^N w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)})$$

Voici donc la formulation des 2 problèmes proprement rédigée :

**ESTIMATION PROBLEM 1 - GENERAL ESTIMATION PROBLEM FOR THE POSTERIOR DISTRIBUTION:** By using the information about

1. State-space model of the system (14).
2. Transition probability  $p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_{k-1})$ .
3. Measurement probability  $p(\mathbf{y}_k | \mathbf{x}_k)$ .
4. Observed outputs  $\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k$  and applied control inputs  $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k-1}$ .
5. Initial guess of the particles

$$\{(\mathbf{x}_0^{(i)}, w_0^{(i)}) | i = 1, 2, 3, \dots, N\} \quad (22)$$

At every time step  $k$ , estimate the set of particles (weights and state samples)

$$\{(\mathbf{x}_k^{(i)}, w_k^{(i)}) | i = 1, 2, 3, \dots, N\} \quad (23)$$

that approximate the posterior distribution of the state  $\mathbf{x}_k$

$$p(\mathbf{x}_k | \mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}) \approx \sum_{i=1}^N w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}) \quad (24)$$

**ESTIMATION PROBLEM 2- RECURSIVE ESTIMATION PROBLEM FOR THE POSTERIOR DISTRIBUTION:** By using the information about

1. State-space model of the system (14).
2. Transition probability  $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1})$ .
3. Measurement probability  $p(\mathbf{y}_k|\mathbf{x}_k)$ .
4. Observed outputs  $\mathbf{y}_k$  and the control input  $\mathbf{u}_{k-1}$  applied at the previous time step  $k - 1$ .
5. The set of particles computed at the time step  $k - 1$

$$\{(\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}) | i = 1, 2, 3, \dots, N\} \quad (25)$$

**At the time step  $k$ , compute the set of particles**

$$\{(\mathbf{x}_k^{(i)}, w_k^{(i)}) | i = 1, 2, 3, \dots, N\} \quad (26)$$

**That recursively approximates the posterior distribution**

$$p(\mathbf{x}_k|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}) \approx \sum_{i=1}^N w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}) \quad (27)$$

**Partie 2 :** Il faut déjà résoudre cette équation :

$$E[\beta(\mathbf{x}_{0:k})|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}] = \int_{\mathbf{x}_{0:k} \in \bar{S}} \beta(\mathbf{x}_{0:k}) p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}) d\mathbf{x}_{0:k}$$

Le souci est que nous n'avons pas  $p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1})$ , mais on peut utiliser l'important sampling avec :

$$E[\beta(\mathbf{x}_{0:k})|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}] \approx \sum_{i=1}^N \tilde{w}_k^{(i)} \beta(\mathbf{x}_{0:k}^{(i)}) \quad (18)$$

where  $\mathbf{x}_{0:k}^{(i)}$  is a sample of the state sequence  $\mathbf{x}_{0:k}$  sampled from the smoothed proposal (importance) density  $q(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1})$ , and the weight is defined by

$$\tilde{w}_k^{(i)} = \frac{p(\mathbf{x}_{0:k}^{(i)}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1})}{q(\mathbf{x}_{0:k}^{(i)}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1})} \quad (19)$$

Ensuite en partant de cette équation :

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}) = \gamma \cdot p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1})p(\mathbf{x}_{0:k-1}|\mathbf{y}_{0:k-1}, \mathbf{u}_{0:k-2})$$

On peut avoir :

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}) = \gamma \cdot p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1})p(\mathbf{x}_{0:k-1}|\mathbf{y}_{0:k-1}, \mathbf{u}_{0:k-2}) \quad (30)$$

The expression (30) is important since it establishes a recursive relation for the probability density  $p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1})$ . Namely, we can observe that the density  $p(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1})$ , depends on its previous value at the time step  $k-1$ , measurement density, and the state transition density. This recursive relation is very important for deriving the recursive expression for the importance weight  $w_k$ .

On cherche maintenant à avoir une forme récursive de la “smoothed proposal density”  $q$ . On trouve :

$$q(\mathbf{x}_{0:k}|\mathbf{y}_{0:k}, \mathbf{u}_{0:k-1}) = q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1})q(\mathbf{x}_{0:k-1}|\mathbf{y}_{0:k-1}, \mathbf{u}_{0:k-2})$$

On va maintenant définir une relation récursive pour le poids :

$$w_k \propto \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1})}{q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1})} \cdot w_{k-1}$$

où :

$$\tilde{w}_{k-1} = \frac{p(\mathbf{x}_{0:k-1}|\mathbf{y}_{0:k-1}, \mathbf{u}_{0:k-2})}{q(\mathbf{x}_{0:k-1}|\mathbf{y}_{0:k-1}, \mathbf{u}_{0:k-2})}$$

Nous devons connaître  $q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1})$ , pour cela on a :

One of the simplest possible particle filters, called the bootstrap particle filter or the Sequential Importance Resampling (SIR) filter selects this function as the state transition density

$$q(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) \quad (42)$$

L'algorithme du filtre est donc le suivant :

- **STEP 1:** Using the state samples  $\mathbf{x}_{k-1}^{(i)}, i = 1, 2, \dots, N$ , computed at the previous step  $k - 1$ , at the time step  $k$  generate  $N$  state samples  $\mathbf{x}_k^{(i)}, i = 1, 2, \dots, N$ , from the state transition probability with the density of  $p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_{k-1})$ . That is,
$$\mathbf{x}_k^{(i)} \sim p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{u}_{k-1}) \quad (46)$$

- **STEP 2:** By using the observed output  $\mathbf{y}_k$ , and the weight  $w_{k-1}^{(i)}$  computed at the previous time step  $k - 1$ , calculate the intermediate weight update, denoted by  $\hat{w}_k^{(i)}$ , by using the density function of the measurement probability

$$\hat{w}_k^{(i)} = p(\mathbf{y}_k|\mathbf{x}_k^{(i)}) \cdot w_{k-1}^{(i)}, \quad i = 1, 2, 3, \dots, N \quad (47)$$

After all the intermediate weights  $\hat{w}_k^{(i)}, i = 1, 2, 3, \dots, N$  are computed, compute the normalized weights  $w_k^{(i)}$  as follows

$$\begin{aligned} w_k^{(i)} &= \frac{1}{\bar{w}} \hat{w}_k^{(i)}, \\ \bar{w} &= \sum_{i=1}^N \hat{w}_k^{(i)} \end{aligned} \quad (48)$$

After STEP 1 and STEP 2, we computed the set of particles for the time step  $k$

$$\{(\mathbf{x}_k^{(i)}, w_k^{(i)}) | i = 1, 2, 3, \dots, N\} \quad (49)$$

- **STEP 3 (RESAMPLING):** This is a resampling step that is performed only if the condition given below is satisfied.

Calculate the constant  $N_{eff}$

$$N_{eff} = \frac{1}{\sum_{i=1}^N (w_k^{(i)})^2} \quad (50)$$

If  $N_{eff} < N/3$ , then generate a new set of  $N$  particles from the computed set of particles

$$\{(\mathbf{x}_k^{(i)}, w_k^{(i)}) | i = 1, 2, 3, \dots, N\}^{\text{NEW}} = \text{RESAMPLE}\left(\{(\mathbf{x}_k^{(i)}, w_k^{(i)}) | i = 1, 2, 3, \dots, N\}^{\text{OLD}}\right) \quad (51)$$

by using a resampling method (see the comments below). In (51), the "OLD" set of particles is the set of particles computed after STEP 1 and STEP 2, that is, the set given in (49).

La 3ème étape, le resampling, permet de contrer l'effet de dégénérescence de cet algorithme qui peut amener à obtenir des poids trop proches de 0.