Filtre de Kalman discret (KF)

On considère un système dynamique stochastique linéaire

$$\left\{\begin{array}{l} x_{k+1} = A_k x_k + w_k \\ y_k = H_k x_k + v_k \end{array}\right. \tag{1.1}$$

Où
- $\mathbf{x}_k \in \mathbb{R}^n$ désigne l'état du système
- $A_k \in \mathcal{M}_{n\times n}(\mathbb{R})$ désigne la matrice de transition relative à la dynamique réelle
- $\mathbf{y}_k \in \mathbb{R}^m$ désigne les mesures relatives à l'état réel $\mathbf{x}$
- $H_k \in \mathcal{M}_{m\times n}(\mathbb{R})$ désigne la matrice de mesure
- $v_k \in \mathbb{R}^m$ désigne l'erreur d'obervation, qu'on suppose $\mathcal{N}(0, R_k)$
- $w_k \in \mathbb{R}^n$ désigne la perturbation du système, qu'on suppose $\mathcal{N}(0, Q_k)$

On suppose de plus que v_k et w_k sont indépendantes.

On mettra en exposant un f (*forecast*) pour les prévisions et un a (*analysis*) pour les analyses.

On postule l'absence d'erreur du modèle pour la prédiction à partir d'une analyse.

$$\mathbf{x}_{k+1}^f = A_k \mathbf{x}_k^a$$

Une estimation de l'état du système à un instant t_k (analyse) est supposément une combinaison linéaire de la prédiction actuelle et des mesures actuelles.

$$\mathbf{x}_k^a = \tilde{L}_k \mathbf{x}_k^f + \tilde{K}_k \mathbf{y}_k \text{ où } \tilde{L}_k \in \mathcal{M}_{n\times n}(\mathbb{R}) \text{ et } \tilde{K}_k \in \mathcal{M}_{n\times m}(\mathbb{R}) \text{ sont des matrices à préciser}$$

On définit formellement les erreurs de prédiction et d'analyse.

$$\mathbf{e}_k^f = \mathbf{x}_k^f - \mathbf{x}_k \quad \mathbf{e}_k^a = \mathbf{x}_k^a - \mathbf{x}_k$$

où $\mathbf{x}_k$ désigne l'état réel du système.

Proposition 1.1 : Si $\tilde{L}_k = I - \tilde{K}_k H_k$ et si $\mathbb{E}\left(\mathbf{e}_k^f\right) = 0$, alors $\mathbb{E}\left(\mathbf{e}_k^a\right) = 0$.

Démonstration

$$\begin{aligned} \mathbb{E}\left(\mathbf{e}_k^a\right) = & \ \mathbb{E}\left(\mathbf{x}_k^a\right) - \mathbf{x}_k \\ = & \ \mathbb{E}\left(\tilde{L}_k \mathbf{x}_k^f + \tilde{K}_k \mathbf{y}_k\right) - \mathbf{x}_k \\ = & \ \mathbb{E}\left((I - \tilde{K}_k H_k)\mathbf{x}_k^f + \tilde{K}_k \mathbf{y}_k\right) - \mathbf{x}_k \\ = & \ \mathbb{E}\left(\mathbf{x}_k^f - \tilde{K}_k H_k \mathbf{x}_k^f + \tilde{K}_k (H_k \mathbf{x}_k + v_k)\right) - \mathbf{x}_k \\ = & \ \mathbb{E}\left((I - \tilde{K}_k H_k)\mathbf{e}_k^f\right) \\ = & \ 0 \end{aligned}$$

□

Définition 1.1 : On appelle $\tilde{K}_k$ la *matrice de gain*. Elle décrit les poids attribués aux observations $\mathbf{y}_k$ et aux mesures $H_k \mathbf{x}_k^f$.

On définit les covariances des erreurs de prévision et d'analyse.

$$P_k^f = \mathbb{E}\left(\mathbf{e}_k^f \cdot^{\mathrm{T}} \mathbf{e}_k^f\right) \quad P_k^a = \mathbb{E}\left(\mathbf{e}_k^a \cdot^{\mathrm{T}} \mathbf{e}_k^a\right) \tag{1.2}$$

Proposition 1.2 : Sous les hypothèses de la proposition précédente, on a :

$$\begin{aligned} P_k^a = & \ (I - \tilde{K}_k H_k) \cdot P_k^f \cdot^{\mathrm{T}} (I - \tilde{K}_k H_k) + \tilde{K}_k \cdot R_k \cdot^{\mathrm{T}} \tilde{K}_k \\ P_k^f = & \ A_k \cdot P_k^a \cdot^{\mathrm{T}} A_k - Q_k \end{aligned} \tag{1.3}$$

Démonstration On pose $Cov(\mathbf{x}) = \mathbb{E}\left(\mathbf{x} \cdot^{\mathrm{T}} \mathbf{x}\right)$.

$$\begin{aligned}
P_k^a = & \ \mathbb{E}\left(\mathbf{e}_k^a \cdot^{\mathrm{T}} \mathbf{e}_k^a\right) = Cov(\mathbf{e}_k^a) \\
= & \ Cov\left(\tilde{L}_k \mathbf{x}_k^f + \tilde{K}_k \mathbf{y}_k - \mathbf{x}_k\right) \\
= & \ Cov\left((I - \tilde{K}_k H_k)\mathbf{x}_k^f + \tilde{K}_k \mathbf{y}_k - \mathbf{x}_k\right) \\
= & \ Cov\left((I - \tilde{K}_k H_k)\mathbf{x}_k^f + \tilde{K}_k (H_k \mathbf{x}_k + v_k) - \mathbf{x}_k\right) \\
= & \ (I - \tilde{K}_k H_k) \cdot Cov\left(\mathbf{x}_k^f - \mathbf{x}_k\right) \cdot^{\mathrm{T}} (I - \tilde{K}_k H_k) + Cov\left(\tilde{K}_k v_k\right) \\
= & \ (I - \tilde{K}_k H_k) \cdot P_k^f \cdot^{\mathrm{T}} (I - \tilde{K}_k H_k) + \tilde{K}_k \cdot R_k \cdot^{\mathrm{T}} \tilde{K}_k
\end{aligned}$$

$$\begin{aligned}
P_{k+1}^f = & \ Cov\left(\mathbf{e}_{k+1}^f\right) \\
= & \ Cov\left(\mathbf{x}_{k+1}^f - \mathbf{x}_{k+1}\right) \\
= & \ Cov\left(A_k \mathbf{x}_k^a - A_k \mathbf{x}_k - w_k\right) \\
= & \ A_k \cdot Cov(e_k^a) \cdot^{\mathrm{T}} A_k - Cov\left(w_k\right) \\
= & \ A_k \cdot P_k^a \cdot^{\mathrm{T}} A_k - Q_k
\end{aligned}$$

□

Remarque : Dans un modèle parfait, on aurait $Q_k = 0$.

On contruit maintenant $\mathcal{I}_k^a$, un estimateur de la fiabilité des analyses, en mesurant l'écart entre l'analyse et le véritable état. Soit $B_k \in \mathcal{M}_{n\times n}(\mathbb{R})$ déterministe diagonale définie positive : la matrice d'échelle. On pose :

$$\mathcal{I}_k^a = \mathbb{E}\left({}^{\mathrm{T}}\mathbf{e}_k^a \cdot B_k \cdot \mathbf{e}_k^a\right) = \mathbb{E}\left(\sum_{i=1}^{n} (\mathbf{e}_k^a)_i m_{ii} (\mathbf{e}_k^a)_i\right) = \mathrm{Tr}(B_k P_k^a) \tag{1.4}$$

On souhaite minimiser $\mathcal{I}_k^a$. Cela revient à minimiser P_k^a. On va différencier l'application $\varphi : \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathbb{R}, \tilde{K}_k \mapsto P_k^a$.

$$\begin{aligned}
\varphi(\tilde{K}_k + h) = & \ (I - (\tilde{K}_k + h)H_k) \cdot P_k^f \cdot^{\mathrm{T}} (I - (\tilde{K}_k + h)H_k) + (\tilde{K}_k + h) \cdot R_k \cdot^{\mathrm{T}} (\tilde{K}_k + h) \\
= & \ P_k^f - \tilde{K}_k \cdot H_k \cdot P_k^f - P_k^f \cdot^{\mathrm{T}} H_k \cdot^{\mathrm{T}} \tilde{K}_k + \tilde{K}_k \cdot H_k \cdot P_k^f \cdot^{\mathrm{T}} H_k^{\mathrm{T}} \tilde{K}_k + \tilde{K}_k \cdot R_k \cdot^{\mathrm{T}} \tilde{K}_k \\
& -h \cdot H_k \cdot P_k^f - P_k^f \cdot^{\mathrm{T}} H_k \cdot^{\mathrm{T}} h + h \cdot H_k \cdot P_k^f \cdot^{\mathrm{T}} H_k \cdot^{\mathrm{T}} \tilde{K}_k \\
& +\tilde{K}_k \cdot H_k \cdot P_k^f \cdot^{\mathrm{T}} H_k \cdot^{\mathrm{T}} h + h \cdot R_k \cdot^{\mathrm{T}} \tilde{K}_k + \tilde{K}_k \cdot R_k \cdot^{\mathrm{T}} h \\
& +h \cdot H_k \cdot P_k^f \cdot^{\mathrm{T}} H_k \cdot^{\mathrm{T}} h + h \cdot R_k \cdot^{\mathrm{T}} h
\end{aligned}$$

d'où

$$d(\varphi)(\tilde{K}_k)(h) = -h \cdot (H_k \cdot P_k^f \cdot (I -^{\mathrm{T}} H_k \cdot^{\mathrm{T}} \tilde{K}_k) - R_k \cdot^{\mathrm{T}} \tilde{K}_k) - ((I - \tilde{K}_k \cdot H_k) P_k^f \cdot^{\mathrm{T}} H_k - \tilde{K}_k \cdot R_k) \cdot^{\mathrm{T}} h$$

On cherche $\tilde{K}_k$ optimale, c'est à dire telle que $d(\varphi)(\tilde{K}_k) = 0$. C'est le cas lorsque :

$$H_k \cdot^{\mathrm{T}} P_k^f \cdot (I -^{\mathrm{T}} H_k \cdot^{\mathrm{T}} \tilde{K}_k) -^{\mathrm{T}} R_k \cdot^{\mathrm{T}} \tilde{K}_k = 0$$

donc lorsque :

$$\tilde{K}_k = P_k^f \cdot^{\mathrm{T}} H_k \cdot \left(H_k \cdot P_k^f \cdot^{\mathrm{T}} H_k + R_k\right)^{-1} := K_k. \tag{1.5}$$

Cette matrice $K_k \in \mathcal{M}_{n\times m}(\mathbb{R})$ est appelée le *gain de Kalman*. En l'injectant dans l'expression de P_k^a, on trouve

$$\begin{aligned}
P_k^a = & \ (I - K_k \cdot H_k) \cdot P_k^f \cdot^{\mathrm{T}} (I - K_k \cdot H_k) + K_k \cdot R_k \cdot^{\mathrm{T}} K_k \\
= & \ P_k^f - P_k^f \cdot^{\mathrm{T}} H_k \cdot (H_k \cdot P_k^f \cdot^{\mathrm{T}} H_k + R_k)^{-1} \cdot H_k \cdot P_k^f \\
& -P_k^f \cdot^{\mathrm{T}} H_k \cdot (H_k \cdot^{\mathrm{T}} P_k^f \cdot^{\mathrm{T}} H_k +^{\mathrm{T}} R_k)^{-1} \cdot H_k \cdot^{\mathrm{T}} P_k^f \\
& +P_k^f \cdot^{\mathrm{T}} H_k \cdot (H_k \cdot^{\mathrm{T}} P_k^f \cdot^{\mathrm{T}} H_k +^{\mathrm{T}} R_k)^{-1} \cdot H_k \cdot^{\mathrm{T}} P_k^f \\
= & \ P_k^f \cdot (I -^{\mathrm{T}} H_k \cdot^{\mathrm{T}} K_k) \\
= & \ (I - H_k \cdot K_k) \cdot P_k^f
\end{aligned}$$

La dernière égalité ayant lieu grâce au caractère symétrique de P_k^f.

On aboutit finalement à la relation :

$$\mathbf{x}_k^a = \mathbf{x}_k^f + K_k(\mathbf{y}_k - H_k\mathbf{x}_k^f). \tag{1.6}$$

La procédure s'itère dans le temps suivant les étapes :

Analyse
$$\left|\begin{array}{l} K_k = P_k^f \cdot H_k \cdot (H_k \cdot P_k^f \cdot^{\mathrm{T}} H_k + R_k)^{-1} \\ \mathbf{x}_k^a = \mathbf{x}_k^f + K_k \cdot (\mathbf{y}_k - H_k \cdot \mathbf{x}_k^f) \\ P_k^a = (I - K_k \cdot H_k) \cdot P_k^f \end{array}\right.$$

Prédiction
$$\left|\begin{array}{l} \mathbf{x}_{k+1}^f = A_k \cdot \mathbf{x}_k^a \\ P_{k+1}^f = A_k \cdot P_k^a \cdot^{\mathrm{T}} A_k + Q_k \end{array}\right.$$

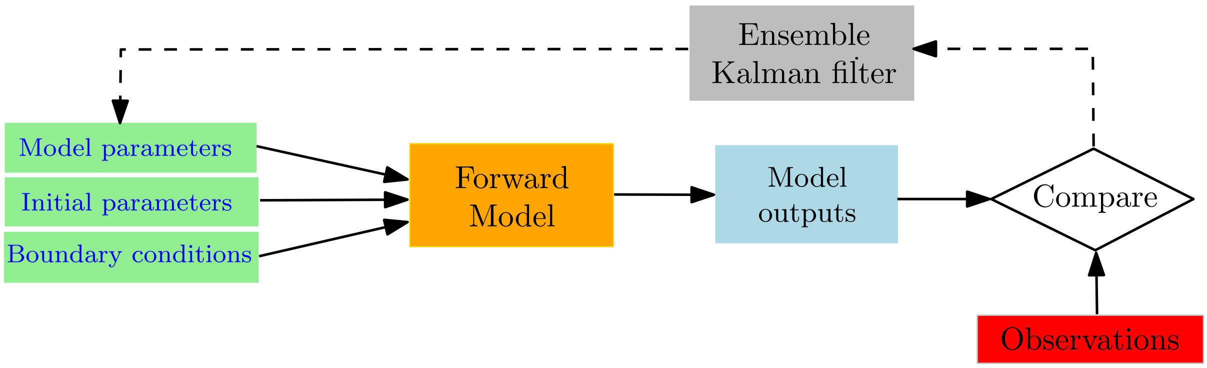

Figure 1 – Fonctionnement général du KF

Exemple : chute d'un mobile

On considère la chute d'un mobile soumis à une accélération de la pesanteur g au cours d'un intervalle de temps $[0,T]$ discrétisé avec un pas dt. On note sa position, sa vitesse et son accélération au temps $k.dt$ respectivement x_k, x'_k et x''_k. Le vecteur d'état du système à un instant $k.dt$ est :

$$\mathbf{x}_k = \begin{pmatrix} x_k \\ x'_k \\ x''_k \end{pmatrix}.$$

La physique nous renseigne sur la matrice de transition A d'un instant au suivant :

$$A = \begin{pmatrix} 1 & dt & dt^2 \\ 0 & 1 & dt \\ 0 & 0 & 1 \end{pmatrix}$$

de façon à ce que la relation de récurrence suivante ait lieu :

$$\mathbf{x}_{k+1} = A \cdot \mathbf{x}_k + \mathbf{w}_k$$

où $\mathbf{w}_k \sim \mathcal{N}(0, Q)$ pour une matrice de covariance Q fixée. En supposant que les bruits affectant la position, la vitesse et l'accélération sont indépendants, on trouve :

$$Q = \begin{pmatrix} (w_k)^2 & 0 & 0 \\ 0 & (w'_k)^2 & 0 \\ 0 & 0 & (w''_k)^2 \end{pmatrix}.$$

On relève à chaque instant l'état du système en commettant une erreur $\mathbf{v}_k$:

$$\mathbf{y}_k = H \cdot \mathbf{x}_k + \mathbf{v}_k = \mathbf{x}_k + \mathbf{v}_k$$

en prenant $H = I_3$, où $\mathbf{v}_k \sim \mathcal{N}(0, R)$ pour une matrice de covariance diagonale fixée R (on suppose les erreurs de mesure indépendantes).

En considérant le pas de temps dt suffisamment petit, il est raisonnable d'initialiser `Pf=`$P_1^f = R$.

Les matrices impliquées dans nos calculs ont des expressions relativement simples grâce à l'hypothèse $H = I_3$, qui revient à supposer que l'on mesure directement les grandeurs qui nous intéressent.

```
génération des états réels x_k bruités par w_k de covariance Q
génération des mesures y_k bruitées par v_k de covariance R

Pf = R ;
xf(1) = y(1);

pour k de 2 à T/dt
    K = Pf * inv(Pf + R);
    xa = xf(:,i-1) + K * (y(:,i) - xf(:,i-1));
    Pa = (I - K) * Pf;

    xf(:,i) = A * xa;
    Pf = A * Pa * A' + Q;
fin
```

La figure 1 illustre les résultats obtenus pour la position et la vitesse d'un mobile dans le cas où :

— $g = -9.81 m.s^{-2}$,
— $dt = 0.2s$,
— $T = 5s$,
— $\mathbf{x}_1 =^{\mathrm{T}} \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$,
— $w_k = 1$, $w'_k = 0.5$, $w''_k = 0.25$,
— $v_k = 5$, $v'_k = 5$, $v''_k = 10$.

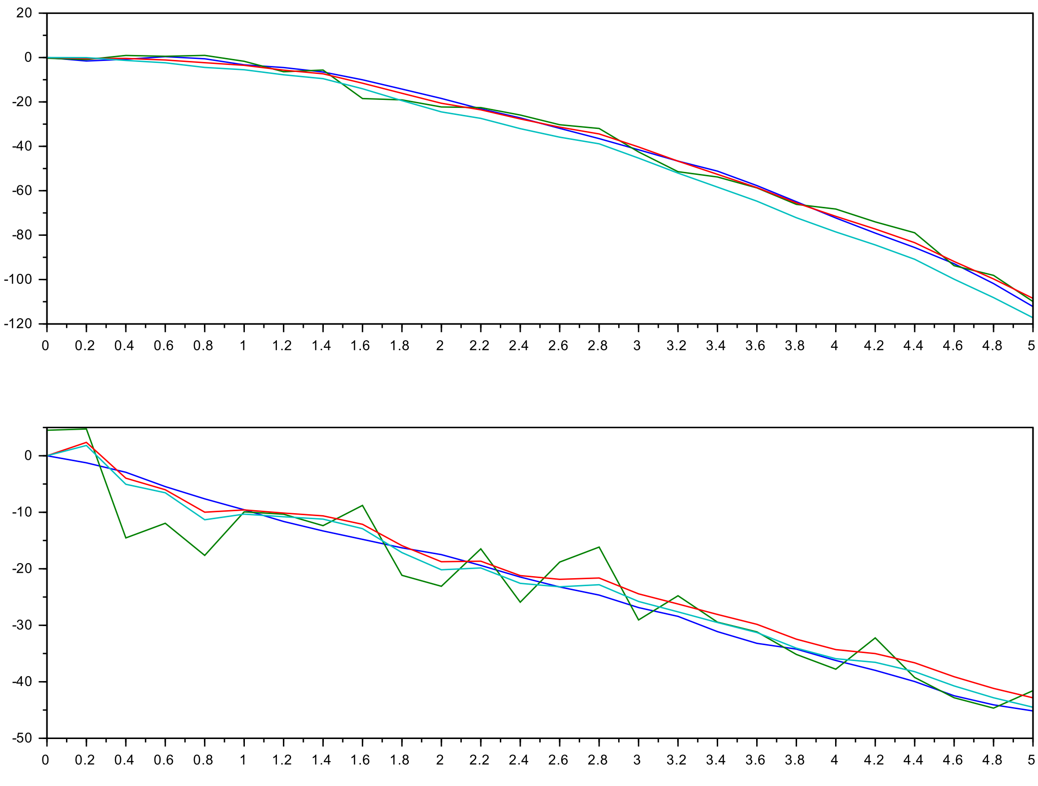

FIGURE 2 – En bleu : $\mathbf{x}$. ; en vert : $\mathbf{y}$; en rouge : $\mathbf{x}^a$; en cyan : $\mathbf{x}^f$.

Filtre de Kalman à ensemble (EnKF) - dynamique explicite

Ce filtre fonctionne sur le meme principe que le KF, à ceci près qu'on évalue P_k^f à partir d'un ensemble $\{\mathbf{x}_k^{fi} | i = 1, ..., q_{ens}\}$ où $q_{ens} > 1$ est un entier. Dans le cas où la dynamique du système est décrite par une fonction linéaire ou par une fonction très régulière, on préfèrera le KF appliqué à cette fonction ou à la fonction linéarisée pour des raisons de coût. En revanche, l'EnKF dévoile ses performances dans le cas des processus non-linéaires. En toute généralité :

$$\begin{cases} \mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k \\ \mathbf{y}_k = h(\mathbf{x}_k) + \mathbf{v}_k \end{cases} \tag{3.1}$$

où les notations employées sont les mêmes que précédemment, et f désigne la fonction décrivant la dynamique du système et $\mathbf{u}_k$ est un paramètre du système.

Au premier pas, on se munit d'un ensemble d'estimations initiales $\mathbf{X}_1 = \{\mathbf{x}_1^{fi} | i = 1, ..., q_{ens}\}$, et on suppose à chaque pas qu'on dispose d'une telle famille $\mathbf{X}_k$. On définit la prédiction moyenne :

$$\overline{\mathbf{x}_k^f} = \frac{1}{q_{ens}} \sum_{i=1}^{q_{ens}} \mathbf{x}_k^{fi}. \tag{3.2}$$

La matrice de covariance de l'erreur de prédiction est alors :

$$P_k^f = \frac{1}{q_{ens} - 1} \sum_{i=1}^{q_{ens}} (\mathbf{x}_k^{fi} - \overline{\mathbf{x}_k^f}) \cdot^{\mathrm{T}} (\mathbf{x}_k^{fi} - \overline{\mathbf{x}_k^f}). \tag{3.3}$$

On bruite q_{ens} fois la mesure de façon à produire une famille $\{\mathbf{y}_k^i = \mathbf{y}_k + \mathbf{e}_k^i | i = 1, ..., q_{ens}\}$ où chaque $\mathbf{e}_k^i$ est gaussien de taille m et de faible variance.

Pour calculer le gain de Kalman, il est nécessaire de construire la matrice R_k. Par indépendance des bruits, elle doit être diagonale. On trouve dans la littérature :

$$R_k = \operatorname{diag}\left(\frac{1}{q_{ens} - 1} E \cdot^{\mathrm{T}} E\right) \tag{3.4}$$

où $E = (\mathbf{e}_k^1, ..., \mathbf{e}_k^{q_{ens}}) \in \mathcal{M}_{m \times q_{ens}}(\mathbb{R})$.

Le gain de Kalman s'exprime de la même façon que dans le cas précédent ; toutefois, H_k désigne l'approximation linéaire de h au voisinage de $\mathbf{x}_k$, ce qui peut être délicat à calculer. D'après Houtekamer et Mitchell si $\overline{h(\mathbf{x}_k^f)} = h(\overline{\mathbf{x}_k^f})$ et si $\|\mathbf{x}_k^{fi} - \overline{\mathbf{x}_k^f}\|$ est assez petit pour tout i, alors les approximations suivantes sont raisonnables :

$$P_{\mathbf{x}\mathbf{y}_k}^f = P_k^f \cdot^{\mathrm{T}} H_k \simeq \frac{1}{q_{ens} - 1} \sum_{i=1}^{q_{ens}} (\mathbf{x}_k^{fi} - \overline{\mathbf{x}_k^f}) \cdot^{\mathrm{T}} (h(\mathbf{x}_k^{fi}) - h(\overline{\mathbf{x}_k^f})) \tag{3.5}$$

$$P_{\mathbf{y}\mathbf{y}_k}^f = H \cdot P_k^f \cdot^{\mathrm{T}} H + R_k \simeq \frac{1}{q_{ens} - 1} \sum_{i=1}^{q_{ens}} (h(\mathbf{x}_k^{fi}) - \overline{h(\mathbf{x}_k^f)}) \cdot^{\mathrm{T}} (h(\mathbf{x}_k^{fi}) - \overline{h(\mathbf{x}_k^f)}) + R_k. \tag{3.6}$$

Il suffit alors de remarquer que :

$$K_k = P_{xy_k}^f \cdot \left(P_{yy_k}^f\right)^{-1}. \tag{3.7}$$

En somme, on applique l'algorithme suivant :

```
pour k de 2 à T/dt
    pour i de 1 à qens
        xf(i) = f(xa(i),uk) + w(i);
    fin
    Pxy = formule (3.5);
    Pyy = formule (3.6);
    K = Pxy*inv(Pyy);
    pour i de 1 à qens
        y(i) = yk + e(i);
        xa(i) = xf(i) + K*(y(i) -  h(xf(i)))
    fin
fin
```

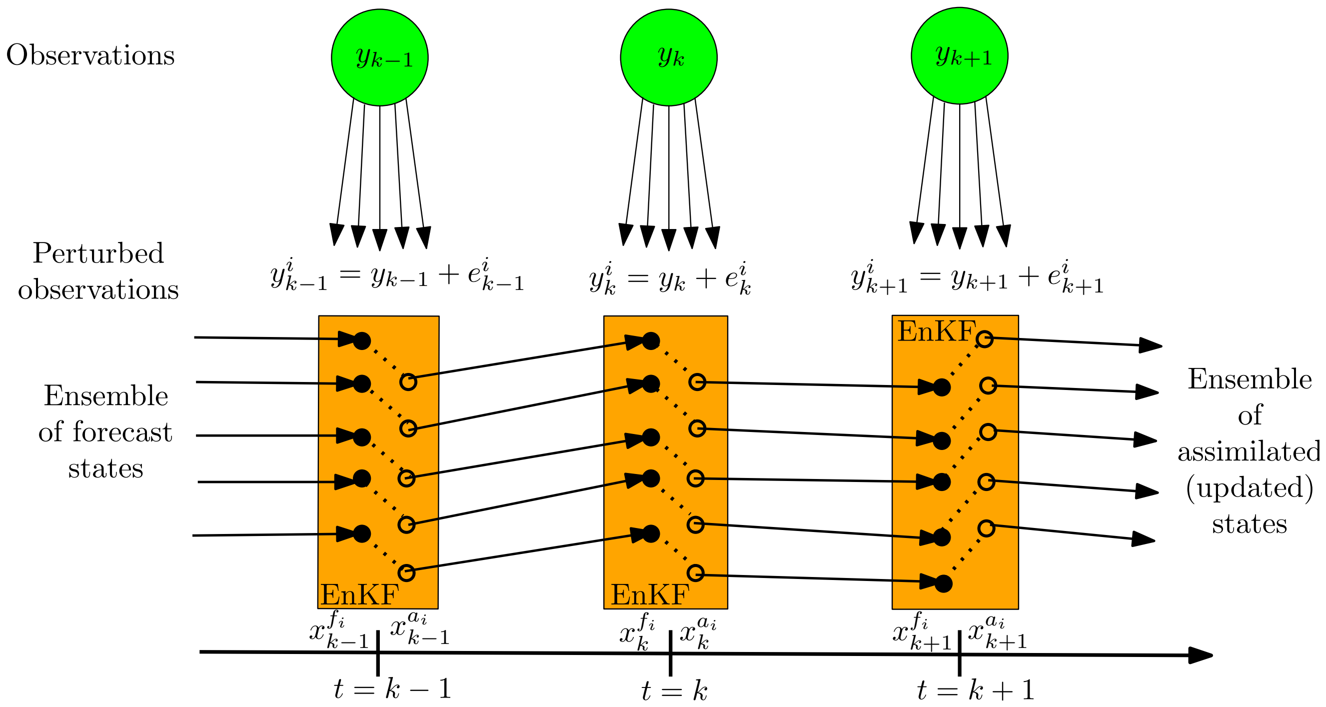

Figure 3 – Fonctionnement du EnKF

Filtre de Kalman à ensemble - dynamique implicite

Il arrive souvent que l'utilisateur n'ait pas accès directement à la fonction f relative à la dynamique du système. Dans ce cas, on peut se contenter d'admettre que x^f suit une marche aléatoire :

$$x^f_{k+1} = x^f_k + \delta_k \tag{4.1}$$

où $\delta_k \sim \mathcal{N}(0, D_k)$. En reprenant l'exemple précédent, on condsidère que la marche aléatoire concerne indépendamment chaque composante de $\mathbf{x}$. La matrice D_k est alors diagonale. Dans le but d'approcher la dynamique réelle, choisissons :

$$D_k(1,1) = (x_k - x_{k-1})^2 = (x'_0 dt + g(2k-1)dt^2)^2$$
$$D_k(2,2) = (x'_k - x'_{k-1})^2 = (gdt)^2$$

La matrice D_k devant être définie positive, on choisit un $\varepsilon > 0$ très petit et on pose :

$$D_k(3,3) = \varepsilon.$$

Dans l'exemple, nous avons arbitrairement pris $\varepsilon = dt$.

L'algorithme devient :

```
pour k de 2 à T/dt
    Dk = diag( (y1(2)*dt + g*(2*k-1)*dt)^2 , (g*dt)^2 , dt);
    pour i de 1 à qens
        xf(i) = xa(i) + delta;
        yk(i) = yk + e;
    fin
    Pxy = formule (3.5);
    R = formule (3.4);
    Pyy = formule (3.6);
    K = Pxy*inv(Pyy);
    pour i de 1 à qens
        xa(i) = xf(i) + K*(y(i) - h(xf(i)));
    fin
fin
```

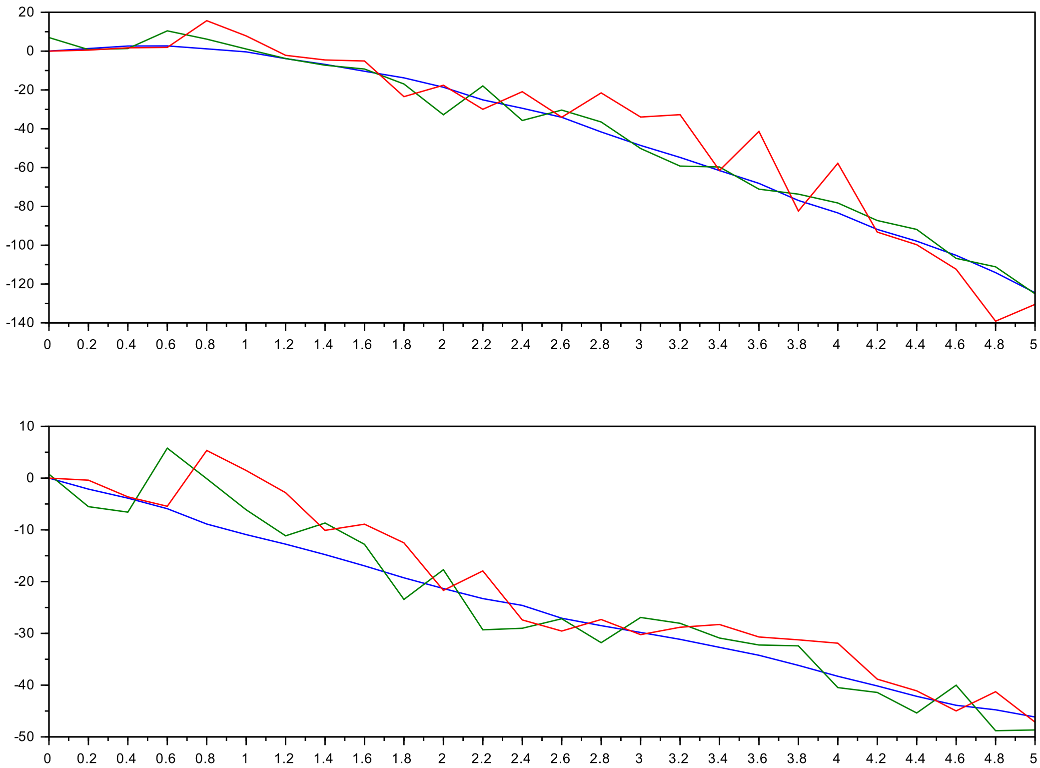

FIGURE 4 – En bleu : $\mathbf{x}$; en vert : $\mathbf{y}$; en rouge : $\overline{\mathbf{x}^f}$.

On voit que la fonction dynamique f n'intervient à aucun moment. La prédiction est évidemment moins précise que dans le cas discret, mais cette méthode permet de produire une prédiction à propos d'un système dont on ne connait *a priori* pas la dynamique. Sur le même exemple que précédemment, on obtient les résultats représentés dans la figure 2.

Ce type de filtre de Kalman permet en particulier d'estimer les paramètres d'un système. Dans sa thèse, R. Lal obtient d'excellentes estimations de la viscosité, de la vitesse d'advection et des conditions initiales dans le cadre de l'équation d'advection-diffusion avec $q_ens = 4$, $T = 100$, $dt = 0,15$, à partir d'une simulation *in silico*.

Bibliographie

[Lal17] R. Lal *Data assimilation and uncertainty quantification in cardiovascular biomechanics*, thèse (2017)

[TC94] R. Todling, S.C. Cohn *Suboptimal schemes for atmospheric data assimilation based on the Kalman filter*, Data Assimilation Office, NASA/Goddard Space Flight Center, Greenbelt, Maryland (1994)