

Perception et Navigation

ISAE-SupAéro, 3^{ème} année

Philippe Mouyon

ONERA - Centre de Toulouse

Département de Contrôle des Systèmes et Dynamique du vol

Variables et signaux aléatoires

Estimation d'état récursive

Equations du filtrage

Filtrage optimal



Perception et Navigation

Variables et signaux aléatoires

Estimation d'état réursive

Equations du filtrage

Filtrage optimal

Densités de probabilité

Quelques définitions relatives aux densités de probabilité, ou lois de probabilité.
 x et y sont deux variables aléatoires.

Loi conjointe et marginale Définition

$$\underbrace{p(x)}_{\text{Densité marginale de } x} \triangleq \int \underbrace{p(x, y)}_{\text{Densité conjointe de } x \text{ et } y} dy$$

En termes de probabilités
 $P(A) = \sum_i P(A \cap B_i)$
 $\{B_i\}$ = partition disjointe
de l'ensemble des possibles¹

Loi conditionnelle Définition

$$\underbrace{p(x|y)}_{\text{Densité conditionnelle de } x \text{ sachant } y, \text{ ou loi de } x \text{ connaissant } y} \triangleq \frac{p(x, y)}{p(y)}$$

En termes de probabilités
 $P(A|B) = \frac{P(A \cap B)}{P(B)}$

La densité conditionnelle est une densité de probabilité.

1. ici et dans toute la suite

Le conditionnement

Le conditionnement permet de remplacer certains calculs complexes par plusieurs calculs plus simples. En effet, au plus on connaît de choses (le *sachant* que du conditionnement), au moins il y a d'aléatoire, et au plus les calculs sont simples...

Conditionnement Formule des probabilités totales

A partir des définitions de la densité marginale et de la densité conditionnelle,

$$p(x) = \int p(x|y) p(y) dy$$

En termes de probabilités

$$P(A) = \sum_i P(A|B_i) P(B_i)$$

Conditionnement Règle de Bayes

$$p(y|x) = \frac{p(x|y) p(y)}{p(x)}$$

où

$$p(x) = \int p(x|y) p(y) dy$$

En termes de probabilités

$$P(B_i|A) = \frac{P(A|B_i) P(B_i)}{P(A)}$$

où

$$P(A) = \sum_i P(A|B_i) P(B_i)$$

Caractéristiques statistiques

Quelques éléments de calcul de l'espérance, ou moyenne stochastique.

Espérance Définition

$$m = E(x) = \int x p(x) dx$$

Espérance Cas d'une fonction de variable aléatoire $y = \varphi(x)$

On peut calculer $E(y)$ sans expliciter $p(y)$, car $p(y) dy = p(x) dx$.

$$E(\varphi(x)) = \int \varphi(x) p(x) dx$$

Caractéristiques statistiques

Espérance Espérance conditionnelle

$p(x|y)$ étant une densité de probabilité, on peut calculer l'espérance de $x|y$ par

$$E(x|y) = \int x p(x|y) dx$$

Espérance Théorème de l'espérance totale

$E(x|y)$ est une fonction de y . Son espérance est donnée par le théorème de l'espérance totale

$$E(E(x|y)) = E(x)$$

Caractéristiques statistiques

Matrice de covariance, Variance Définition

La matrice de covariance, est une matrice définie non négative.

$$P = \text{cov}(x) \triangleq E \left((x - m) (x - m)^T \right)$$

La variance est un scalaire positif ou nul.

$$\text{var} = \sigma^2 \triangleq \text{trace} \{P\} = E \left((x - m)^T (x - m) \right)$$

Décomposition

$$\begin{aligned} P &= E \left(x x^T \right) - m m^T \\ \text{var} &= E \left(x^T x \right) - m^T m \end{aligned}$$

Le cas gaussien

Pourquoi Une somme de petits effets aléatoires indépendants de même loi, se comporte comme une variable aléatoire gaussienne (loi des grands nombres, théorème central limite).

Loi Gaussienne Définition

Une loi gaussienne (ou normale) est entièrement caractérisée par sa moyenne m et sa matrice de covariance P .

$$\begin{aligned} p(x) &= \mathcal{N}(x - m, P) \\ &= (2\pi \det(P))^{-n/2} \exp \left\{ -1/2 (x - m)' P^{-1} (x - m) \right\} \end{aligned}$$

Loi Gaussienne La magie des gaussiennes

Une combinaison linéaire de gaussienne est une gaussienne.

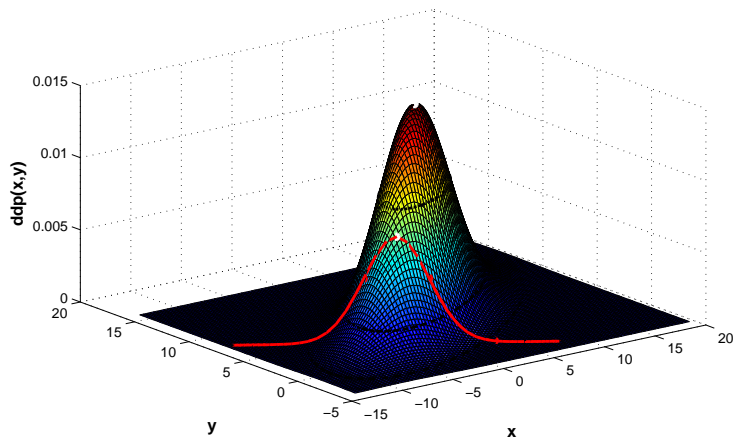
Les lois marginales d'une gaussienne sont des gaussiennes.

La loi conditionnelle d'une gaussienne par une gaussienne est gaussienne.

Le produit de gaussiennes est une gaussienne.

Le cas gaussien

Exemple



Gaussienne 2D
Conditionnement par la mesure
 $z = 1 \cdot x + 2 \cdot y = 3$

Perception et Navigation

Variables et signaux aléatoires

Estimation d'état récursive

Equations du filtrage

Filtrage optimal

Estimation d'état récursive

Contexte Systèmes dynamiques

On considère un système dont le comportement au cours du temps est décrit par :

x_{k+1}	$=$	$f_k(x_k, v_k)$	Equation d'état
z_k	$=$	$h_k(x_k, w_k)$	Equation de mesure

- x est le vecteur d'état
- v est le vecteur de bruit d'état
- z est le vecteur de mesure
- w est le vecteur de bruit de mesure

L'ensemble des mesures connues à l'instant k est

$$Z_k = \{z_i, i = 1, \dots, k\}$$

Estimation d'état récursive

Estimation Objectif

- Estimation d'état

On cherche à construire un estimateur d'état, c'est-à-dire un algorithme qui délivre une estimation \hat{x} de l'état x du système, à partir de la connaissance des mesures z .

- Estimation récursive

L'estimation est dite récursive quand le calcul de \hat{x}_{k+1} utilise la connaissance de \hat{x}_k et de z_{k+1} seulement.

Estimation d'état récursive

Estimation Approche Bayésienne

C'est un cadre de travail dans lequel toutes les inconnues sont modélisées comme des grandeurs aléatoires.

On représente les **grandeurs inconnues**
par des **densités de probabilités connues** :

- x_0 comme une variable aléatoire de densité de probabilité $p(x_0)$.
- v_k comme un signal aléatoire de densité de probabilité $p_v(v_k)$.
- w_k comme un signal aléatoire de densité de probabilité $p_w(w_k)$.

$\implies x_k$ et z_k sont des signaux aléatoires.

Estimation d'état récursive

Démarche générale Calcul récursif de $p(x_k|Z_k)$

Dans le cadre Bayésien, on établit les **équations du filtrage**.

Elles montrent comment calculer récursivement $p(x_k|Z_k)$, en partant de la condition initiale connue $p(x_0|Z_0) = p(x_0)$.

Démarche générale Estimation d'état

A partir de la densité de probabilité $p(x_k|Z_k)$ on peut calculer différentes estimations de l'état :

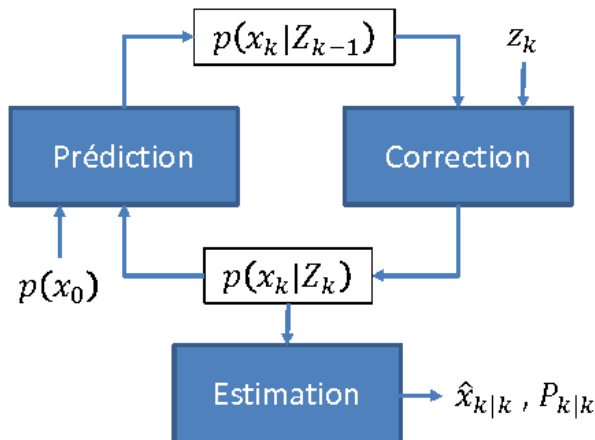
Estimation à variance minimale (MV)

$$\hat{x}_{k|k}^{MV} \triangleq E(x_k|Z_k) = \int x_k p(x_k|Z_k) dx_k$$

Estimation au sens du maximum a posteriori (MAP)

$$\hat{x}_{k|k}^{MAP} \triangleq \arg_{x_k} \max p(x_k|Z_k)$$

Estimation d'état récursive



N.B.

$Z_{k+1} = Z_k \cup z_{k+1}$.

En l'absence de mesure, $p(x_k | Z_k) = p(x_k | Z_{k-1})$.

Estimation d'état récursive

Erreur d'estimation Caractéristiques

L'erreur d'estimation est : $e_k = x_k - \hat{x}_k$

Ses principales caractéristiques statistiques sont :

- Le biais $b_k = E(e_k)$, i.e. l'erreur moyenne
- La covariance $P_k = \text{cov}(e_k) = E((e_k - b_k)(e_k - b_k)^T)$
- La variance $\sigma_k^2 = \text{trace}(P_k)$

Qualification d'un estimateur

On peut qualifier globalement la précision de l'estimation en utilisant l'Erreur Quadratique Moyenne (EQM) :

$$\begin{aligned} EQM_k &= E((x_k - \hat{x}_k)^T (x_k - \hat{x}_k)) \\ &= b_k^2 + \sigma_k^2 \end{aligned}$$

On demande souvent à un estimateur d'être non biaisé (au moins asymptotiquement, c'est-à-dire quand $k \rightarrow \infty$).

Les estimateurs à variance minimale sont recherchés, parfois à tort...

Estimation d'état récursive

L'estimation à variance minimale Caractéristiques

L'estimation à variance minimale est non biaisée.

$$\begin{aligned} b &= E(x_k - \hat{x}_k) \\ &= E(x_k) - E(\hat{x}_k) && \text{par linéarité de } E \\ &= E(x_k) - E(E(x_k|Z_k)) && \text{par définition de } \hat{x} \\ &= E(x_k) - E(x_k) && \text{par théorème de l'espérance totale} \\ &= 0 \end{aligned}$$

Elle minimise donc aussi l'erreur quadratique moyenne.

Le caractère non biaisé implique aussi que $\text{cov}(\hat{x}_k) = \text{cov}(e_k) = P_k$.

La matrice P_k renseigne donc sur la précision de l'estimation, mais aussi directement sur \hat{x}_k .

Perception et Navigation

Variables et signaux aléatoires

Estimation d'état récursive

Equations du filtrage

Filtrage optimal

Equations du filtrage

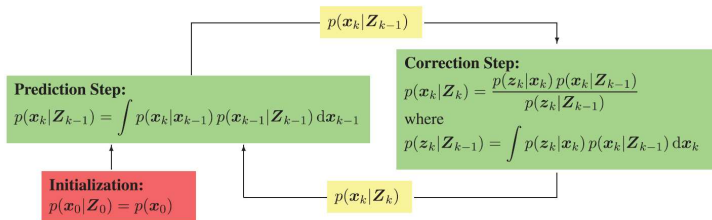
Contexte

L'estimation d'état peut être faite si on connaît la densité de probabilité de l'état, conditionnelle aux mesures passées, $p(x_k|Z_k)$. Il y a plusieurs possibilités :

- Minimiser la variance : $\hat{x}_{k|k}^{MV} \triangleq E(x_k|Z_k) = \int x_k p(x_k|Z_k) dx_k$
- Maximiser la ddp a posteriori : $\hat{x}_{k|k}^{MAP} \triangleq \arg_{x_k} \max p(x_k|Z_k)$

Les équations du filtrage montrent que le calcul de $p(x_k|Z_k)$ peut être conduit de manière récursive en deux étapes :

- **Prédiction** : Calcul de $p(x_k|Z_{k-1})$ à partir de $p(x_{k-1}|Z_{k-1})$.
- **Correction** : Calcul de $p(x_k|Z_k)$ à partir de $p(x_k|Z_{k-1})$ et de z_k .



Equations du filtrage

En prédiction Position du problème

On connaît $p(x_{k-1}|Z_{k-1})$.

On ne dispose pas encore de la mesure à l'instant k .

On veut calculer la densité de probabilité a priori : $p(x_k|Z_{k-1})$.

Démarche

- La formule $p(x) = \int p(x|y) p(y) dy$ est appliquée à la loi de x_k conditionnelle aux mesures Z_{k-1} .

$$p(x_k|Z_{k-1}) = \int p(x_k|x_{k-1}, Z_{k-1}) p(x_{k-1}|Z_{k-1}) dx_{k-1}$$

- On suppose que le système a un comportement Markovien, c'est-à-dire que la connaissance de x_{k-1} résume entièrement le comportement passé. La connaissance de Z_{k-1} n'apporte pas d'information supplémentaire à celle de x_{k-1} .

$$\text{Comportement Markovien} \implies p(x_k|Z_{k-1}, x_{k-1}) = p(x_k|x_{k-1})$$

- On fait ainsi apparaître la probabilité de transition $p(x_k|x_{k-1})$.

Equations du filtrage

Equation de prédiction

En prédiction (c'est-à-dire en l'absence de mesure), la densité de probabilité conditionnelle aux mesures évolue selon l'équation de Chapman-Kolmogorov

$$p(x_k | Z_{k-1}) = \int \underbrace{p(x_k | x_{k-1})}_{\text{Probabilité de transition}} p(x_{k-1} | Z_{k-1}) dx_{k-1}$$

Remarque

Si la dynamique vérifie $x_{k+1} = f_k(x_k) + v_k$, avec $v_k \sim p_v(v_k)$, alors la probabilité de transition est $p(x_{k+1} | x_k) = p_v(x_{k+1} - f_k(x_k))$.

Equations du filtrage

En correction Position du problème

On connaît $p(x_k|Z_{k-1})$.

On acquiert la mesure z_k à l'instant k .

On veut calculer la densité de probabilité a posteriori : $p(x_k|Z_k)$.

Démarche

- La règle de Bayes $p(y|x) = p(x|y) p(y) / p(x)$, avec $p(x) = \int p(x|y) p(y) dy$ est appliquée à la densité de x_k conditionnelle aux mesures Z_{k-1} .

$$p(x_k|z_k, Z_{k-1}) = \frac{p(z_k|x_k, Z_{k-1}) p(x_k|Z_{k-1})}{p(z_k|Z_{k-1})}$$

où $p(z_k|Z_{k-1}) = \int p(z_k|x_k, Z_{k-1}) p(x_k|Z_{k-1}) dx_k$

- On fait apparaître : $Z_k = \{Z_{k-1}, z_k\}$
- Et l'hypothèse Markovienne donne : $p(z_k|x_k, Z_{k-1}) = p(z_k|x_k)$, qui est la vraisemblance de la mesure.

Equations du filtrage

Equation de correction

En correction (c'est-à-dire lors de l'acquisition d'une mesure), la densité de probabilité conditionnelle aux mesures évolue selon l'équation

$$p(x_k|Z_k) = \frac{\overbrace{p(z_k|x_k)}^{\text{Vraisemblance de la mesure}} p(x_k|Z_{k-1})}{p(z_k|Z_{k-1})}$$

où $p(z_k|Z_{k-1}) = \int p(z_k|x_k) p(x_k|Z_{k-1}) dx_k$

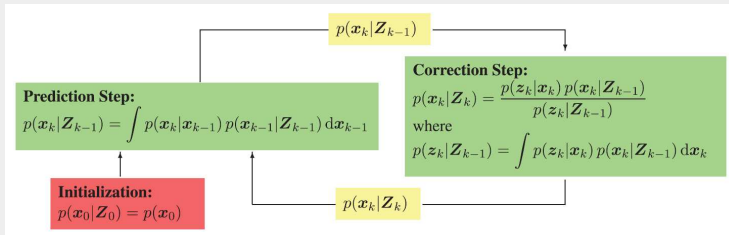
Remarque

Si la mesure vérifie $z_k = h_k(x_k) + w_k$, avec $w_k \sim p_w(w_k)$, alors la vraisemblance est $p(z_k|x_k) = p_w(z_k - h_k(x_k))$.

Equations du filtrage

Récapitulation

- Calcul récursif de la ddp de l'état conditionnelle aux mesures :



- Estimation de l'état : plusieurs possibilités
 - Minimiser la variance : $\hat{x}_{k|k}^{MV} \triangleq E(x_k|Z_k) = \int x_k p(x_k|Z_k) dx_k$
 - Maximiser la ddp a posteriori : $\hat{x}_{k|k}^{MAP} \triangleq \arg_{x_k} \max p(x_k|Z_k)$

Perception et Navigation

Variables et signaux aléatoires

Estimation d'état récursive

Equations du filtrage

Filtrage optimal

Résolution des équations du filtrage

Résolution exacte

Il existe quelques cas particuliers pour lesquels on sait résoudre exactement les équations du filtrage. Les deux plus connus sont :

- **Les systèmes linéaires à bruits (et conditions initiales) gaussiens**
⇒ Densité gaussienne : filtre de Kalman (KF)
- **Les systèmes à espace d'état discret**

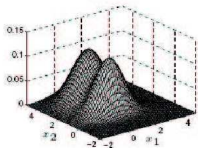
Résolution approchée

Plusieurs techniques ont été développées pour résoudre approximativement les équations du filtrage :

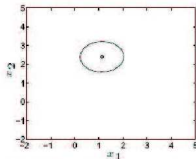
- Approximation gaussienne : filtre de Kalman étendu (EKF).
- Approximation gaussienne : filtre de Kalman sans parfum (UKF).
- Approximation multi-gaussienne : filtre de Kalman multi-gaussien
- Approximation point-par-point : filtre particulaire

Résolution des équations du filtrage

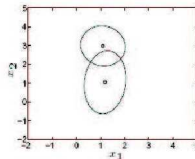
Approximations d'une densité de probabilité



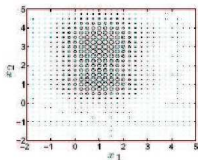
(a) True pdf.



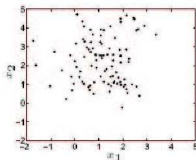
(b) Gaussian approximation.



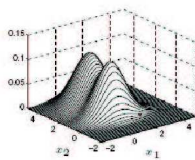
(c) Gaussian sum approximation.



(d) Grid-based approximation.



(e) Particle approximation.



(f) PF posterior pdf (waterfall view).

Filtre de Kalman

Le cas linéaire gaussien

Le comportement du système est décrit par :

$$\begin{cases} x_{k+1} &= F x_k + v_k & \text{avec } v_k \sim \mathcal{N}(0, Q) \\ z_k &= H x_k + w_k & \text{avec } w_k \sim \mathcal{N}(0, R) \end{cases}$$

avec $x_0 \sim \mathcal{N}(m_0, P_0)$, et x_0 , v_k et w_k indépendants.

- Linéarité

⇒ **Caractère gaussien conservé au cours du temps**

- Forme paramétrique des ddp

$$p(x_k | Z_{k-1}) = \mathcal{N}(x_k - m_{k|k-1}, P_{k|k-1})$$

$$p(x_k | Z_k) = \mathcal{N}(x_k - m_{k|k}, P_{k|k})$$

⇒ **Suffit de calculer l'évolution des paramètres**

- Densité gaussienne

⇒ **Tous les estimateurs coïncident**, $\hat{x}_k^{MV} = \hat{x}_k^{MAP} = m_{k|k}$

Le filtre de Kalman correspond à cette solution.

Les ingrédients

- Matrice de transition : $p(x_{k+1} | x_k) = \mathcal{N}(x_{k+1} - F x_k, Q)$
- Vraisemblance des mesures : $p(z_k | x_k) = \mathcal{N}(z_k - H x_k, R)$

Les équations

Filtre de Kalman

Prédiction

$$\hat{x}_{k+1|k} = F \hat{x}_{k|k}$$

$$P_{k+1|k} = F P_{k|k} F^T + Q$$

Correction

$$\nu_{k+1} = z_{k+1} - H \hat{x}_{k+1|k}$$

$$S_{k+1} = H P_{k+1|k} H^T + R$$

$$K_{k+1} = P_{k+1|k} H^T S_{k+1}^{-1}$$

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1} \nu_{k+1}$$

$$P_{k+1|k+1} = P_{k+1|k} - K_{k+1} H P_{k+1|k}$$

Remarques

Comme $\hat{x} = m$, on propage directement \hat{x} au lieu de m .

Le vecteur ν_k est l'innovation, écart entre mesure et mesure prédite.

La matrice S_k est la covariance de ν_k .

La matrice de covariance P reste symétrique définie non négative.

Filtre de Kalman étendu

Le cas presque linéaire gaussien

Le comportement du système est décrit par :

$$\begin{cases} x_{k+1} &= f(x_k, v_k) & \text{avec } v_k \sim \mathcal{N}(0, Q) \\ z_k &= h(x_k, w_k) & \text{avec } w_k \sim \mathcal{N}(0, R) \end{cases}$$

avec $x_0 \sim \mathcal{N}(m_0, P_0)$, et x_0 , v_k et w_k indépendants.

Pour des non linéarités pas trop fortes, la densité de l'état reste approximativement gaussienne.

On développe alors un estimateur appelé *Filtre de Kalman Etendu*.

Pour propager la matrice P , il utilise la linéarisation de f et h évaluée autour de la dernière estimation d'état disponible.

Filtre de Kalman étendu

Les équations

Filtre de Kalman étendu

Prédiction

$$\begin{aligned} \hat{x}_{k+1|k} &= f(\hat{x}_{k|k}, 0) \\ P_{k+1|k} &= F_k P_{k|k} F_k^T + G_k Q G_k^T \end{aligned} \quad \left| \quad \begin{aligned} F_k &= \frac{\partial f}{\partial x}(\hat{x}_{k|k}, 0) \\ G_k &= \frac{\partial f}{\partial v}(\hat{x}_{k|k}, 0) \end{aligned} \right.$$

Correction

$$\begin{aligned} \nu_{k+1} &= z_{k+1} - h(\hat{x}_{k+1|k}, 0) \\ S_{k+1} &= H_{k+1} P_{k+1|k} H_{k+1}^T + E_{k+1} R E_{k+1}^T \\ K_{k+1} &= P_{k+1|k} H_{k+1}^T S_{k+1}^{-1} \\ \hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + K_{k+1} \nu_{k+1} \\ P_{k+1|k+1} &= P_{k+1|k} - K_{k+1} H_{k+1} P_{k+1|k} \end{aligned} \quad \left| \quad \begin{aligned} H_{k+1} &= \frac{\partial h}{\partial x}(\hat{x}_{k+1|k}, 0) \\ E_{k+1} &= \frac{\partial h}{\partial w}(\hat{x}_{k+1|k}, 0) \end{aligned} \right.$$

Remarques

La covariance n'est pas linéaire : $\text{cov}(G v_k) = G \text{cov}(v_k) G^T$

Réglage

- **P_0 , la taille de l'incertitude initiale**

Si elle est grande, le filtre utilisera beaucoup les premières mesures pour recalculer rapidement la condition initiale \hat{x}_0 .

- **Q , la taille du bruit d'état indiquée au filtre**

Si elle est grande, cela empêche P de diminuer. Le filtre utilisera le plus possible les mesures. Le bruit sera transmis à l'estimé, mais l'estimé sera insensible aux imprécisions de modélisation.

- **R , la taille du bruit de mesure indiquée au filtre**

Si elle est grande, le filtre utilisera le moins possible les mesures. Le bruit sera peu transmis à l'estimé, mais l'estimé sera sensible aux imprécisions du modèle.

- **Optimalité**

Le filtre n'est optimal que pour le réglage correspondant à la réalité. Mais la réalité n'est pas toujours bien représentée par le modèle... Il est donc recommandé de jouer avec les paramètres pour obtenir le résultat cherché.

Mise en oeuvre

Pour que ça marche

- **Le système doit être observable !**
Le choix des capteurs, et du modèle doivent être faits pour assurer l'observabilité. Sinon la matrice P diverge...
- **P doit rester symétrique définie non négative**
C'est théoriquement vrai par construction. Mais numériquement cela n'est pas toujours évident, et peut entraîner la divergence du filtre.
- **P doit représenter correctement le domaine d'incertitude**
Une sous-estimation est souvent cause de divergence du filtre.

Du bonus

- **Fonctionnement asynchrone**
A chaque instant, on corrige avec les mesures disponibles.
C'est-à-dire que la dimension de z peut varier au cours du temps.
- **Traitement par composante**
L'étape de correction peut être décomposée, en traitant les composantes de z une par une. Pour chaque composante ν est scalaire et S aussi.
L'inversion de S devient une simple division, numériquement plus simple à gérer.