Modèle corrigé par les données : contrôle optimal et filtre aux moindres carrés

Florian De Vuyst

CMLA UMR 8536 – ENS CACHAN

20 janvier 2014

Hypothèses et motivation

- On possède un modèle d'évolution (a priori faux)
- On connaît partiellement la donnée initiale (incertitude)
- On possède "quelques" observations réelles (série temporelle en un point par exemple)

- On souhaite recaler la solution numérique au cours du temps en assimilant les données
- On adopte une démarche "déterministe"
- Revient à rechercher un observateur d'état (state observer)

Objectifs d'application

- Simulation au niveau système $\Rightarrow$ Recalage à partir de données expérimentales
- Simulation système "embarquée" $\Rightarrow$ Recalage temps réel à partir de flux de données
- Modèles EDP $\Rightarrow$ recalage à partir de mesures expérimentales

Cadre habituel d'analyse

- Filtres de Kalman-Bucy (xKF)
- Formulation probabiliste
- Probabilités conditionnelles aux mesures
- Prédiction + filtrage
- Hypothèses sur les variables aléatoires

Équivalence moindres carrés – max. de vraisemblance avec loi normale

- Variables aléatoires $\boldsymbol{d}$ de moyenne nulle, matrice de covariance normalisée à l'identité
- Sorties : $\boldsymbol{y} = C\boldsymbol{d}$
- Observations : $\tilde{\boldsymbol{y}}$
- Sorties $\boldsymbol{z} = H\boldsymbol{d}$ à estimer.

$$\min_{\boldsymbol{d}} \left\{ \frac{1}{2}\|\boldsymbol{d}\|^2 \ : \quad \boldsymbol{y} = C\boldsymbol{d} \right\}$$

Lagrangien :

$$\mathscr{L}(\boldsymbol{d}, \lambda) = \frac{1}{2}\|\boldsymbol{d}\|^2 + \lambda^T (C\boldsymbol{d} - \boldsymbol{y})$$

Estimateur

Conditions d'optimalité du 1er ordre (Euler-Lagrange) :

$$\partial_d \mathscr{L} = \hat{\boldsymbol{d}} + C^T \hat{\boldsymbol{\lambda}} = 0,$$

$$\partial_\lambda \mathscr{L} = C\hat{\boldsymbol{d}} - \tilde{\boldsymbol{y}} = 0$$

$\Rightarrow\ \hat{\boldsymbol{\lambda}} = -(CC^T)^{-1}\boldsymbol{y}$
$\Rightarrow\ \hat{\boldsymbol{d}} = -C^T\hat{\boldsymbol{\lambda}} = C^T(CC^T)^{-1}\boldsymbol{y}$

$$\Rightarrow\ \hat{\boldsymbol{z}} = H\hat{\boldsymbol{d}} = H\,C^T(CC^T)^{-1}\boldsymbol{y}$$

(supposant C de rang maximal). Identique à l'estimateur du maximum de vraisemblance en hypothèse de loi normale.

Extension au cas dynamique (linéaire)

$$\begin{cases} \dot{\boldsymbol{x}} = A\boldsymbol{x} + D\boldsymbol{d}_1, \quad t \in (0,T], \\ \boldsymbol{y} = C\boldsymbol{x} + \boldsymbol{d}_2, \\ \boldsymbol{z} = H\boldsymbol{x} \\ \boldsymbol{x}(0) = \boldsymbol{x}^0. \end{cases}$$

d_1 est vu comme une perturbation (éxogène) du modèle.
d_2 est vu comme une perturbation de mesure ("bruit").
x^0 est un état initial non (totalement) observable.

Hypothèses de régularité

Notations :

$$\|f\|^2_{L^2(A,\mathbb{R}^n)} = \int_A \|f(t)\|^2_{\mathbb{R}^n}\,dt$$

On suppose que $d_1 \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^d)$, $d_2 \in L^2_{loc}(\mathbb{R}_+, \mathbb{R}^y)$.
Alors on a les solutions

$$y(t) = Ce^{At}x(0) + \int_0^T Ce^{A(t-\tau)}Gd_1(\tau)\,d\tau + d_2(t),$$

$$z(t) = He^{At}x(0) + \int_0^T He^{A(t-\tau)}Gd_1(\tau)\,d\tau.$$

Q : Pour un $T > 0$, comment estimer $\hat{z}(T)$ de $z(T)$ à partir des observations $\tilde{y}(t)$, $0 \le t \le T$?

Minimisation d'une fonctionnelle d'incertitude

$$\min_{\substack{d_1 \in L^2((0,T),\mathbb{R}^d),\\ d_2 \in L^2((0,T),\mathbb{R}^y),\\ x(0)\in\mathbb{R}^n}} \quad \mathscr{U} := \frac{1}{2}\left\{ \|x(0)\|_\Gamma^2 + \|d_1\|^2_{L^2((0,T),\mathbb{R}^d)} + \|d_2\|^2_{L^2((0,T),\mathbb{R}^y)} \right\}$$

avec $x(t)$ solution des EDO, $y = Cx + d_2 = \tilde{y}$, $z = Hx$.

On note $(d_1^\star, d_2^\star, x(0)^\star)$ l'argument minimum.
$\Rightarrow$ On est amené à choisir $\hat{z}(T) = z^\star(T)$, sortie la plus vraisemblable, qui minimise la mesure d'incertitude.

Q : existe-t-il un dispositif récursif permettant de calculer $\hat{z}$ de façon efficace, pour tous les T ?

Lemme préparatoire

Lemme. On considère

$$\dot{x} = Ax + Gd_1, \quad y = Cx + d_2,$$

et

$$\frac{d\hat{x}}{dt} = A\hat{x} + \Sigma C^T (y - C\hat{x}),$$

$$\frac{d\Sigma}{dt} = GG^T + A\Sigma + \Sigma A^T - \Sigma C^T C\Sigma \quad \text{(Riccati).}$$

Alors si $d_1, d_2 \in L^2$, Σ sym. non singulière pour $0 \leq t \leq T$, on a

$$\|x(0) - \hat{x}(0)\|^2_{\Sigma(0)^{-1}} + \|d_1\|^2_{L^2((0,T),\mathbb{R}^d)} + \|d_2\|^2_{L^2((0,T),\mathbb{R}^y)} =$$

$$= \|x(T) - \hat{x}(T)\|^2_{\Sigma(T)^{-1}} + \|d_1 - G^T\Sigma(.)^{-1}(x - \hat{x})\|^2_{L^2(\dots)} + \|y - C\hat{x}\|^2_{L^2(\dots)}$$

Preuve du Lemme

Démonstration.

$$\frac{d}{dt}\left[(x-\hat{x})\Sigma^{-1}(x-\hat{x})\right] = ... := RHS$$

Rem :

$$\frac{d\Sigma^{-1}}{dt} = -\Sigma^{-1}\frac{d\Sigma}{dt}\Sigma^{-1}.$$

On trouve

$$RHS = \|d_1(t)\|^2 + \|d_2(t)\|^2 - \|d_1 - G^T\Sigma^{-1}(x-\hat{x})\|^2(t) - \|y - C\hat{x}\|^2(t)$$

On termine par une intégration en temps sur $(0, T)$. $\square$

Lemme 2

Application : on considère $\Sigma(0) = \Gamma^{-1}$, Γ spd.

Lemme 2. Soit Σ matrice solution de

$$\frac{d\Sigma}{dt} = GG^T + A\Sigma + \Sigma A^T - \Sigma C^T C \Sigma,$$
$$\Sigma(0) = \Gamma^{-1}.$$

Alors Σ sdp pour tout $t \in [0, T]$. Soit $d_1, d_2 \in L^2$ et soit $\hat{x} = \hat{x}(., y, \Sigma)$ solution

$$\frac{d\hat{x}}{dt} = A\hat{x} + \Sigma C^T (y - C\hat{x}),$$
$$\hat{x}(0) = 0.$$

Alors

$$\mathscr{U} = \|x(0)\|^2_\Gamma + \|d_1\|^2_{L^2((0,T),\mathbb{R}^d)} + \|d_2\|^2_{L^2((0,T),\mathbb{R}^y)} =$$

$$= \|x(T) - \hat{x}(T)\|^2_{\Sigma(T)^{-1}} + \|d_1 - G^T \Sigma(.)^{-1}(x - \hat{x})\|^2_{L^2(\dots)} + \|y - C\hat{x}\|^2_{L^2(\dots)}.$$

Appl. : filtre aux moindres carrés

Considérons maintenant

$$\frac{d\hat{x}}{dt} = A\hat{x} + \Sigma C^T(\boxed{\tilde{y}} - C\hat{x}),$$

$$\hat{x}(0) = 0.$$

$\hat{x}$ est alors une fonction de $\tilde{y}$.
A condition que $x(0)$ conduise à $y = \tilde{y}$, on a la borne inférieure pour $\mathscr{U}$ (dépendant de $\tilde{y}$) :

$$\begin{aligned} \mathscr{U} &= \|x(T) - \hat{x}(T)\|^2_{\Sigma(T)^{-1}} + \|d_1 - G^T\Sigma(.)^{-1}(x - \hat{x})\|^2_{L^2(\dots)} + \|\tilde{y} - C\hat{x}\|^2_{L^2(\dots)} \\ &\geq \|\tilde{y} - C\hat{x}\|^2_{L^2(\dots)} = m(\tilde{y}). \end{aligned}$$

Q : la borne inf peut-elle être atteinte pour un certain $x(0)$, d_1, d_2 ?

Recherche d'un minimum

Objectif. Trouver un triplet $(x(0), d_1, d_2)$ et la trajectoire associée $x(t)$ tels que

1. $\dot{x} = Ax + Gd_1$, $y = Cx + d_2$, $x(t=0) = x(0)$;
2. $y(t) = \tilde{y}(t)$,
3. $x(T) = \hat{x}(T)$,
4. $d_1(t) = G^T \Sigma(t)(x(t) - \hat{x}(t)),\ 0 \leq t \leq T$.

Si c'est possible, alors, pour le triplet trouvé,

$$\mathscr{U}(x(0), d_1, d_2) = \|\tilde{y} - C\hat{x}\|^2_{L^2(\dots)}.$$

On récupère en plus la mesure de l'incertitude minimale via l'estimateur $\hat{x}$ (estimateur à posteriori).

Théorème

Théorème. Soit $\tilde{y} \in L^2_{loc}(\mathbb{R}, \mathbb{R}^y)$. Soit Σ l'unique solution de

$$\frac{d\Sigma}{dt} = GG^T + A\Sigma + \Sigma A^T - \Sigma C^T C\Sigma,$$
$$\Sigma(0) = \Gamma^{-1}.$$

Alors le filtre optimal aux moindre carrés est donné par

$$\frac{d\hat{x}}{dt} = A\hat{x} + \Sigma C^T(\tilde{y} - C\hat{x}),$$

$$\hat{x}(0) = 0,$$

$$\hat{z} = H\hat{x}$$

et on a la mesure d'incertitude

$$\mathscr{U} = \|\tilde{y} - C\hat{x}\|^2_{L^2(\dots)}.$$

Démonstration

Démonstration.

- On considère Σ solution des éq. de Riccati (forward) ;
- On considère les équations (forward) de l'estimateur $\hat{x}$;
- On considère de plus le système supplémentaire (rétrograde) :

$$\frac{d}{dt}\tilde{x} = A\tilde{x} - GG^T\Sigma^{-1}(\tilde{x} - \hat{x}), \quad t \in [0, T),$$

$$\tilde{x}(T) = \hat{x}(T).$$

On considère

$$d_1^\star(t) = G^T\Sigma^{-1}(t)(\tilde{x}(t) - \hat{x}(t)), \quad 0 \leq t \leq T,$$

$$d_2^\star(t) = \tilde{y}(t) - C\tilde{x}(t) \quad \text{et } x(0)^\star = \tilde{x}(0).$$

$\square$

Démonstration (suite)

On vérifie que

$$\frac{d\tilde{x}}{dt} = A\tilde{x} + Gd_1^\star(t),$$

$$y(t) = C\tilde{x} + d_2^\star(t) = \tilde{y}(t),$$

$$\tilde{x}(0) = x(0)^\star$$

avec notamment $\tilde{x}(T) = \hat{x}(T)$. CQFD.

Aspect remarquable de la méthode

Naïvement, on calculerait $\hat{z}(t)$ à partir de

$$\dot{x}(t) = Ax(t) + Gd_1^\star(t),$$
$$x(0) = x(0)^\star = \tilde{x}(0) \quad \text{(backward)}$$
$$\hat{z}(T) = Hx(T)$$

+ répéter l'opération pour chaque $T \geq 0$. En fait, il y a juste à intégrer l'équation en $\hat{x}$ pour obtenir le filtre optimal

$$\dot{\hat{x}}(t) = A\hat{x}(t) + \Sigma C^T(\tilde{y} - C\hat{x}),$$
$$\hat{x}(0) = 0,$$
$$\hat{z}(T) = H\hat{x}(T).$$

$\Rightarrow$ Approche récursive ! (purement forward)

Variantes

On peut bien sûr considérer d'autres fonctionnelles de mesure d'incertitude, par exemple

$$\mathscr{U} = \|x(0) - a\|_{\Gamma}^2 + \|d_1\|_{L^2((0,T),\mathbb{R}^d)}^2 + \|d_2\|_{L^2((0,T),\mathbb{R}^y)}^2$$

NB : a est souvent appelé un a priori sur $x(0)$. La matrice (sdp) Γ représente la matrice de covariance sur la donnée initiale par rapport à la moyenne a.

Équation de l'erreur

D'un coté,

$$\dot{x} = Ax + d_1,\ y = Cx,\ x(t=0) = x(0).$$

De l'autre coté,

$$\dot{\hat{x}} = A\hat{x} + \Sigma C^T(\tilde{y} - C\hat{x}),\ \hat{x}(0) = 0$$

et $y = \tilde{y}$. Considérant $e = \hat{x} - x$, on a

$$\dot{e} = Ae - \Sigma C^T Ce - d_1 = (A - \Sigma C^T C)e - d_1.$$
$$e(0) = x(0).$$

Filtrage en temps infini

Mesure d'incertitude :

$$\mathscr{U} = \|d_1\|^2_{L^2((-\infty,T),\mathbb{R}^d)} + \|d_2\|^2_{L^2((-\infty,T),\mathbb{R}^y)}$$

Théorème. On suppose A Hurwitz. Soit $\tilde{y}(.) \in L^2$ une observation. Soit Σ_∞ la solution de l'équation de Riccati algébrique

$$GG^T + A\Sigma + \Sigma A^T - \Sigma C^T C \Sigma = 0.$$

Alors le filtre optimal aux moindres carrés infini en temps est donné par l'unique solution $\hat{x}$ de

$$\frac{d\hat{x}}{dt} = A\hat{x} + \Sigma_\infty C^T(\tilde{y} - C\hat{x}),$$

et $\hat{z} = H\hat{x}$. La matrice $(A - \Sigma_\infty C^T C)$ est Hurwitz, par conséquent $\hat{x} \in L^2$ et le filtre est bien défini.

Illustration sur un ex. scalaire simple

Considérons

$$\dot{x} = d_1,$$

$$y = x + d_2.$$

Mesure d'incertitude : $\mathscr{U} = \rho^2 \|d_1\|_{L^2}^2 + \|d_2\|_{L^2}^2$ (poids $\rho > 0$).
Équation algébrique de Riccati :

$$\Sigma_\infty^2 - \frac{1}{\rho^2} = 0 \;\rightarrow\; \Sigma_\infty = \frac{1}{\rho}.$$

$\Rightarrow\; \frac{d}{dt}\hat{x} = \frac{1}{\rho}(\tilde{y} - \hat{x})$. On trouve la moyenne exponentielle

$$\hat{x}(T) = \int_0^\infty \frac{1}{\rho}\, e^{-t/\rho}\, \tilde{y}(T-t)\, dt.$$

Rem : ρ homogène à un temps $\leftrightarrow$ temps de relaxation.

Extension au cas non linéaire

Beaucoup moins clair et plus heuristique. On général, on est que sous-optimal. Par exemple, pour

$$\dot{x} = f(x) + d_1,$$

$$y = Cx,$$

on cherche un observateur d'état solution d'un système du type

$$\dot{\hat{x}} = f(\hat{x}) + K(\tilde{y} - C\hat{x}).$$

Rem : en supposant f Lipschitzienne, on peut obtenir des inégalités différentielles sur $e = \hat{x} - x$ ...

En version probabiliste : EKF (Extended Kalman Filter), UKF (Unscented Kalman Filter), ...

Références

- J.C. Willems, Deterministic least squares filtering, J. of Econometrics, 118, 341–373 (2004).
- E. Trélat, Contrôle optimal. Théorie et applications, Vuibert, Mathématiques concrètes, 2è édition (2008).
- P. Chu, Inversion Algorithm Based on the Unscented Kalman Filter for Inverse Heat Conduction Problems, Advanced Materials Research, MEMS ans Mechanics, Vol 705, 474–482 (2013).