

Apprentissage des paramètres de bruit d'un système linéaire gaussien couple

Valérian Nêmesin¹ et Stéphane Derrode²

¹ Institut Fresnel (CNRS UMR 7249), Aix-Marseille Université, 13397 Marseille Cedex 20, France.
valerian.nemesin@centrale-marseille.fr

² LIRIS (CNRS UMR 5205), École Centrale de Lyon, 69134 Écully, France.
stephane.derrode@ec-lyon.fr

Résumé Nous proposons un algorithme non supervisé d'apprentissage *partiel* et *contraint* des paramètres d'un système linéaire gaussien couple, par recherche du maximum de vraisemblance grâce à un algorithme EM (*Expectation-Maximization*). Le terme "partiel" signifie que seule une partie des paramètres est à estimer (l'autre partie étant supposée connue grâce à la physique du problème), tandis que "contraint" signifie que les paramètres restant à estimer peuvent être liés entre-eux (*p. ex.* par un facteur de proportionnalité inconnu). Ici, nous nous restreignons au cas particulier où seule la matrice de covariance des bruits doit être estimée, comme c'est le cas lors de la poursuite d'une cible en traitements vidéo.

Mots clés Système linéaire gaussien, Apprentissage non supervisé, Algorithme EM.

Abstract We propose an algorithm for the *partial* and *constrained* unsupervised learning of parameters involved in a pairwise linear Gaussian system, by means of the EM (*Expectation-Maximization*) principle. The "partial" term means that only a subset of the parameters is unknown and needs to be estimated (the other part is assumed to be known based on a priori knowledge about the studied system), whereas the "constrained" term means that the remaining parameters to be estimated can be self-linked (*e.g.* by an unknown scale factor). Here we limit ourselves to the particular case where only the noise covariance matrix must be estimated, as in the case of target tracking in video processing.

Key words Gaussian state-space systems, Unsupervised parameter estimation, EM algorithm.

1 Introduction

Un problème important en traitement du signal consiste à estimer un ensemble d'états cachés $\mathbf{x} = \{\mathbf{x}_n\}_{n \in [0:N]}$, $\mathbf{x}_n \in \mathbb{R}^{n_x}$, à partir d'observations $\mathbf{y} = \{\mathbf{y}_n\}_{n \in [0:N]}$, $\mathbf{y}_n \in \mathbb{R}^{n_y}$. Dans ce travail, nous nous intéressons au système linéaire gaussien couple (SLGC) suivant

$$\underbrace{\begin{pmatrix} \mathbf{x}_{n+1} \\ \mathbf{y}_n \end{pmatrix}}_{\mathbf{t}_{n+1}} = \underbrace{\begin{pmatrix} \mathbf{F}^{xx} & \mathbf{F}^{xy} \\ \mathbf{F}^{yx} & \mathbf{F}^{yy} \end{pmatrix}}_{\mathbf{F}} \underbrace{\begin{pmatrix} \mathbf{x}_n \\ \mathbf{y}_{n-1} \end{pmatrix}}_{\mathbf{t}_n} + \underbrace{\begin{pmatrix} \boldsymbol{\omega}_{n+1}^x \\ \boldsymbol{\omega}_{n+1}^y \end{pmatrix}}_{\boldsymbol{\omega}_{n+1}}, \quad (1)$$

introduit dans [4] et étendu aux modèles de Markov triplets dans [2], où $\boldsymbol{\omega} = \{\boldsymbol{\omega}_n\}_{n \in [0:N]}$ est un processus gaussien, où tous les $\boldsymbol{\omega}_n \in \mathbb{R}^{n_t}$ sont mutuellement indépendants et indépendants de $\mathbf{t}_0 \sim \mathcal{N}(\hat{\mathbf{t}}_0, \mathbf{Q}_0)$ et où $\boldsymbol{\omega}_n \sim \mathcal{N}(\mathbf{0}_{n_t}, \mathbf{Q})$ avec $\mathbf{Q} = \begin{pmatrix} \mathbf{Q}^{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}^{yy} \end{pmatrix}$ (décorrélacion des bruits d’observation et d’état).

Les matrices de transition \mathbf{F} et de covariance du bruit \mathbf{Q} (supposées indépendantes de n), ainsi que l’état initial, définissent les paramètres du modèle. Le modèle (1) étend le système linéaire gaussien (SLG) conventionnel tout en conservant des algorithmes optimaux exacts et rapides pour le filtrage et le lissage des données [1].

L’estimation non supervisée des paramètres de tels modèles, par le biais de l’algorithme EM (*Expectation-Maximization*) [5], a fait l’objet de plusieurs papiers, cf. [10] pour les SLGs et [1,6] pour les SLGCs. Nous nous intéressons à l’estimation partielle et contrainte des paramètres:

- le terme “partiel” signifie que certains paramètres sont connus a priori, permettant d’intégrer des connaissances spécifiques sur la physique du problème étudié;
- le terme “contraint” signifie que les paramètres restant à estimer peuvent être liés, *p. ex.* le bruit d’observation est deux fois plus fort que le bruit d’état.

Ce type d’algorithmes a été étudié par Holmes [3] dans le cas des SLGs, et très récemment généralisé aux SLGCs dans [7]. Nous étudions ici le cas particulier où la matrice de transition \mathbf{F} est entièrement connue, et où seule une partie de la matrice de covariance du bruit \mathbf{Q} reste à estimer¹.

2 Apprentissage des paramètres de bruit

Cette Partie décrit un algorithme d’apprentissage partiel et contraint des paramètres d’un SLGC, dénommé EMPC-SLGC, selon le principe de l’algorithme EM. L’algorithme EM cherche à approcher le maximum de la fonction de vraisemblance pour estimer les paramètres. C’est un algorithme itératif dans lequel chaque itération² est composée de deux étapes [6], cf. eq. (1):

- L’étape **Espérance** évalue la fonction auxiliaire de vraisemblance qui s’écrit ici

$$g_{\tilde{\mathcal{C}}}(\mathbf{Q}) = -\frac{1}{2} [N + 1] \log |\mathbf{Q}| - \frac{1}{2} \text{tr} \left[\mathbf{Q}^{-1} [\mathbf{F}, -\mathbf{I}] \tilde{\mathcal{C}} [\mathbf{F}, -\mathbf{I}]^T \right], \quad (2)$$

où $\tilde{\mathcal{C}}$ est précisément définie dans [6, Section II]³.

- L’étape **Maximisation** re-estime les paramètres en maximisant (2).

Il n’est pas possible d’imposer n’importe quelle contrainte sur \mathbf{Q} tout en conservant le principe EM strict (*i.e.* formules exactes de ré-estimation des paramètres).

1. Dans la suite, nous ne chercherons pas à estimer les paramètres $\hat{\mathbf{t}}_0$ et \mathbf{Q}_0 de l’état initial \mathbf{t}_0 .
2. L’indice j indiquant l’itération EM est omis intentionnellement pour simplifier les écritures.
3. La fonction $g_{\tilde{\mathcal{C}}}$ ne dépend que de \mathbf{Q} car \mathbf{F} est supposée connue étant donné les hypothèses de travail.

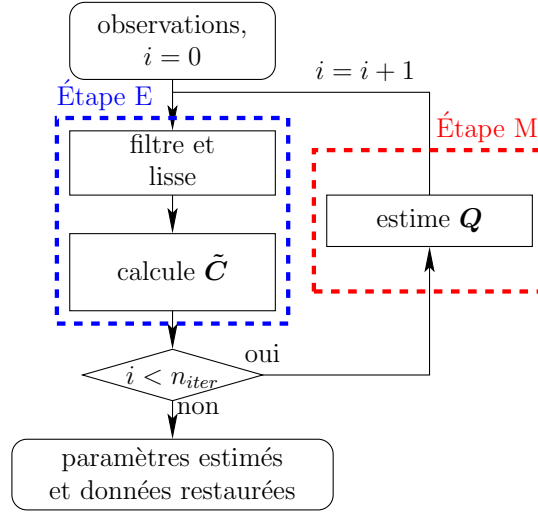


Figure 1: Squelette de l'algorithme EMPC-SLGC.

Nous présentons dans la suite une série de contraintes qui préserve EM et impacte seulement l'étape M de l'algorithme décrit dans la Fig. 1.

La matrice de covariance du bruit étant supposée diagonale, nous pouvons écrire $\mathbf{Q} = \mathbf{P}_1 \mathbf{Q}^{xx} \mathbf{P}_1^T + \mathbf{P}_2 \mathbf{Q}^{yy} \mathbf{P}_2^T$, où $\{\mathbf{P}_i\}_{i=1,2}$ sont les projecteurs des blocs \mathbf{Q}^{xx} et \mathbf{Q}^{yy} de \mathbf{Q} ($\mathbf{P}_1^T \mathbf{P}_2 = \mathbf{0}$)⁴. Sous ces hypothèses, l'éq. (2) se décompose selon

$$g_{\tilde{\mathbf{C}}}(\mathbf{Q}) = g_{\tilde{\mathbf{C}}_1}(\mathbf{Q}^{xx}) + g_{\tilde{\mathbf{C}}_2}(\mathbf{Q}^{yy}) \quad \text{avec } \tilde{\mathbf{C}}_i = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_i \end{bmatrix} \tilde{\mathbf{C}} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_i^T \end{bmatrix}, \quad (3)$$

dont les deux sous-fonctions se maximisent séparément. Aussi, dans la suite, on désignera par $g_{\tilde{\mathbf{C}}_i}(\mathbf{Q}^i)$ soit $g_{\tilde{\mathbf{C}}_1}(\mathbf{Q}^{xx})$, soit $g_{\tilde{\mathbf{C}}_2}(\mathbf{Q}^{yy})$ et par $\mathbf{F}^i = \mathbf{P}_i \mathbf{F}$ la ligne de \mathbf{F} associée à $g_{\tilde{\mathbf{C}}_i}(\mathbf{Q}^i)$, c'est-à-dire

- $\mathbf{F}^i = (\mathbf{F}^{xx}, \mathbf{F}^{xy})$ si $g_{\tilde{\mathbf{C}}_i}(\mathbf{Q}^i) = g_{\tilde{\mathbf{C}}_1}(\mathbf{Q}^{xx})$;
- $\mathbf{F}^i = (\mathbf{F}^{yx}, \mathbf{F}^{yy})$ si $g_{\tilde{\mathbf{C}}_i}(\mathbf{Q}^i) = g_{\tilde{\mathbf{C}}_2}(\mathbf{Q}^{yy})$.

On désignera également la dimension de la matrice carrée \mathbf{Q}^i par n_i .

Nous avons réussi à distinguer quatre configurations pour \mathbf{Q}^i , dont les deux dernières sont particulièrement intéressantes :

1. \mathbf{Q}^i est connue, auquel cas $\hat{\mathbf{Q}}^i = \mathbf{Q}^i$, et il n'y a rien à faire.
2. \mathbf{Q}^i est entièrement inconnue, et les eq. de l'algorithme général d'estimation s'appliquent au sous-bloc:

$$\hat{\mathbf{Q}}^i = \frac{1}{N+1} \left[[\mathbf{F}^i, -\mathbf{I}] \tilde{\mathbf{C}}_i [\mathbf{F}^i, -\mathbf{I}]^T \right]. \quad (4)$$

4. Dans le cas où $n_x = n_y = 1$, nous avons $\mathbf{P}_1 = (1, 0)^T$ et $\mathbf{P}_2 = (0, 1)^T$.

3. $\mathbf{Q}^i = \lambda_i \mathbf{Q}_i^0$, où \mathbf{Q}_i^0 désigne une matrice connue et λ_i un scalaire inconnu, qui s'estime par

$$\hat{\lambda}_i = \frac{1}{(N+1)n_i} \text{tr} \left[[\mathbf{Q}_i^0]^{-1} \left[[\mathbf{F}^i, -\mathbf{I}] \tilde{\mathbf{C}}_i [\mathbf{F}^i, -\mathbf{I}]^T \right] \right]. \quad (5)$$

4. $\mathbf{Q}^i = \sum_{j=1}^{n_i} [\mathbf{P}_i^j]^T \mathbf{R}_i \mathbf{P}_i^j$, où $\{\mathbf{P}_i^j\}_{j=1:n_i}$ sont des projecteurs orthogonaux et \mathbf{R}_i est une matrice de covariance à estimer. La matrice $\hat{\mathbf{R}}_i$ s'estime alors par

$$\hat{\mathbf{R}}_i = \frac{1}{(N+1)n_i} \sum_{j=1}^{n_i} \mathbf{P}_i^j [\mathbf{F}^i, -\mathbf{I}] \tilde{\mathbf{C}}_i [\mathbf{F}^i, -\mathbf{I}]^T [\mathbf{P}_i^j]^T. \quad (6)$$

3 Illustrations

Cette Partie propose d'illustrer l'algorithme d'estimation partielle et contrainte des paramètres de bruit d'un SLGC à l'aide d'une expérience construite à partir de données simulées. Considérons un système où deux capteurs identiques ($n_{\mathbf{y}} = 2$) mesurent une même grandeur physique ($n_{\mathbf{x}} = 1$). Pour l'expérience, la matrice de transition est fixée à

$$\mathbf{F} = \begin{bmatrix} 0.5 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.2 \\ 0.1 & 0.2 & 0.5 \end{bmatrix}. \quad (7)$$

Il est légitime d'imposer la même matrice de covariance du bruit de mesure pour les deux capteurs. Ainsi, en supposant que les bruits d'état et de mesure sont décorrélés, et que les bruits des deux capteurs le sont aussi, nous avons fixé la matrice de covariance du bruit à

$$\mathbf{Q} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \quad (8)$$

Il est alors possible de simuler des jeux de données complets (*i.e.* \mathbf{x} et \mathbf{y}) grâce au système (1). L'objectif est alors de restaurer les états, supposés inconnus, à partir des observations \mathbf{y} , de la matrice \mathbf{F} , et des paramètres de \mathbf{Q} estimés par EM selon les quatre configurations suivantes:

- **Filtre f1** : Les covariances de \mathbf{Q} sont toutes supposées inconnues. Cela correspond à une configuration de type 2, selon la numérotation de la Partie précédente.
- **Filtre f2** : La matrice \mathbf{Q} est décomposée en deux blocs, $\mathbf{Q}^1 = \mathbf{Q}^{xx}$, $\mathbf{Q}^2 = \mathbf{Q}^{yy}$, et chacun d'entre eux est estimé selon une configuration de type 2 également (*i.e.* aucune contrainte n'est imposée aux blocs).

- **Filtre f3** : Le bloc $\mathbf{Q}^1 = \mathbf{Q}^{xx}$ est estimé selon une configuration type 2, mais le bloc $\mathbf{Q}^2 = \mathbf{Q}^{yy}$ est estimé à l'aide d'une contrainte de type 4 qui impose que l'estimation des covariances des bruits des deux capteurs soit réalisée conjointement, de manière à garantir qu'elles seront identiques. Au regard de l'éq. (6), nous avons $\mathbf{Q}^2 = \begin{bmatrix} \mathbf{Q}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_s \end{bmatrix}$ avec $\mathbf{P}_2^1 = [\mathbf{I}, \mathbf{0}]$, $\mathbf{P}_2^2 = [\mathbf{0}, \mathbf{I}]$ et $\mathbf{R}_2 = \mathbf{Q}_s$.
- **Filtre f4** : On suppose \mathbf{Q} parfaitement connu (aucune estimation réalisée). C'est le filtre de référence qui permet de mesurer la pertinence des filtres précédents. Pour les filtres f1 et f2, nous avons: \mathbf{F}^1 formée par la 1^{re} ligne de \mathbf{F} , et \mathbf{F}^2 formée par les deux autres lignes.

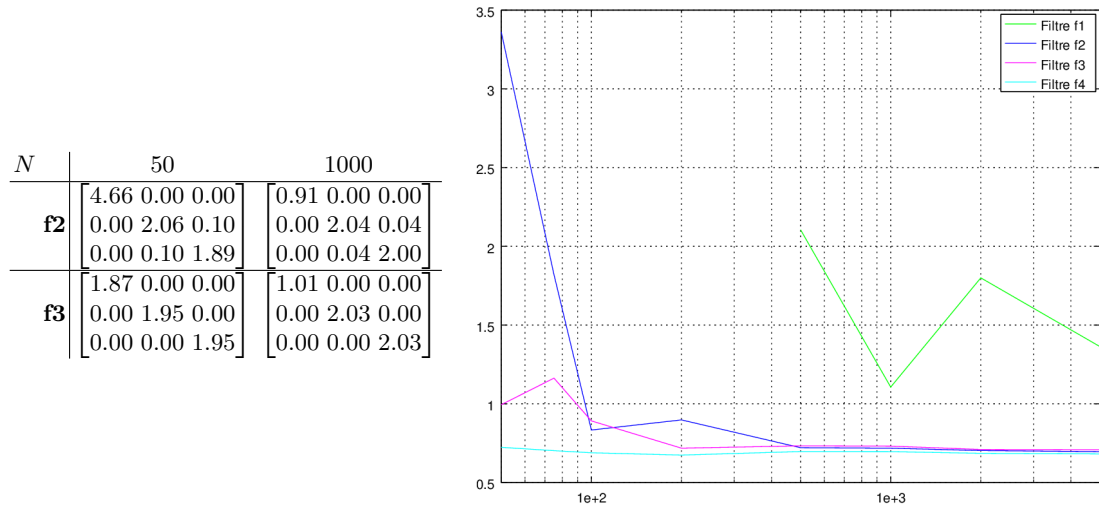


Figure 2: Gauche: Estimées des matrices de covariance de bruit par différentes méthodes d'apprentissage. Les résultats sont à comparer à la matrice de l'éq. (8). Droite: EQM des différents filtres en fonction de la taille de l'échantillon.

La Fig. 2 reporte les matrices $\hat{\mathbf{Q}}$ estimées par les filtres **f2** et **f3**, pour deux tailles de signaux ($N = 50, 1000$). Ces résultats sont des moyennes de 100 expériences indépendantes, le nombre d'itérations de EM étant fixé à 300 dans chacun des cas⁵, avec une initialisation à $\mathbf{Q} = \mathbf{I}_3$. Les courbes d'EQM représentées dans la figure de droite montre l'intérêt de l'estimation contrainte (filtre **f3**) vis à vis de **f2** surtout lorsque la taille de l'échantillon est petite. Il est aussi intéressant de constater que le filtre non contraint est incapable d'estimer correctement les paramètres lorsque l'échantillon est trop petit, et que même dans le cas de grands échantillons,

5. Notre expérience montre que 300 itérations sont toujours suffisantes pour obtenir la convergence des estimations, mais un critère adaptatif basé sur les fluctuations relative des paramètres d'une itération à l'autre serait plus adéquat.

l'estimation obtenue n'est pas en moyenne satisfaisante. Nous pensons que ce comportement provient d'un piégeage par un maximum local de la vraisemblance, que le filtre **f3** évite du fait de son espace plus réduit de paramètres : ceci reste à vérifier.

4 Conclusion

Nous avons présenté un algorithme EM pour l'estimation des paramètres dans un système linéaire gaussien couple. L'originalité de l'approche réside dans les différentes possibilités qu'il offre de contraindre l'estimation de la matrice de covariance du bruit, permettant, pour une application donnée, d'intégrer des informations a priori sur les capteurs ou sur le système lui-même. Le cas typique d'application est lorsque les observations proviennent de deux capteurs que l'on sait identiques sans connaître pour autant précisément la variance du bruit. Des algorithmes d'estimation partielle et contrainte plus généraux ont été établis dans [7,8], qui permette de contraindre également l'estimation de la matrice de transition.

Notons également que les algorithmes utilisés ici sont des versions robustes des algorithmes d'estimation originaux dans la mesure où ils ne cherchent pas à estimer directement les matrices de covariance mais une de leur racine-carrée, par la propagation d'une racine des matrices de covariance nécessaires pour le filtrage, le lissage et la maximisation de la fonction auxiliaire de vraisemblance [9,6].

References

1. B. Ait-el-Fquih and F. Desbouvries. Unsupervised signal restoration in partially observed Markov chains. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'06)*, Toulouse, France, 2006.
2. F. Desbouvries and W. Pieczynski. Triplet Markov models and Kalman filtering. *Comptes Rendus de l'Académie des Sciences - Mathématique - Série I*, 336(8):667–670, 2003.
3. E.E. Holmes. Derivation of the EM algorithm for constrained and unconstrained MARSS models. Technical report, Northwest Fisheries Science Center, 2013.
4. R.S. Lipster and A.N. Shiryaev. *Statistics of Random Processes: I. General theory*. Springer, 2nd edition, 2001.
5. G.J. McLachlan and T. Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, New York, Second Edition, 2008.
6. V. Nêmesin and S. Derrode. Robust blind pairwise Kalman algorithms using QR decompositions. *IEEE Trans. on Signal Processing*, 61(1):5–9, 2013.
7. V. Nêmesin and S. Derrode. Robust partial-learning in linear Gaussian systems. Technical report, To be published in *IEEE Trans. on Automatic Control*, 2015.
8. Valérien Nêmesin. *Apprentissage non-supervisé dans les modèles linéaires gaussiens. Application à la biométrie dynamique de l'iris*. PhD thesis, École Centrale Marseille, soutenue le 30/09/2014.
9. A. Bryson P. Kaminski and S. Schmidt. Discrete square-root filtering: a survey of current techniques. *IEEE Trans. on Automatic Control*, 16(6):727–736, December 1971.
10. R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *J. of Time Series Analysis*, 3(4):253–264, 1982.