

Une méthode générique pour approcher des lois a posteriori sous contraintes égalité

Asma RABAOUI¹, Audrey GIREMUS²

¹Institut Fresnel - UMR 6133 CNRS - Université Aix-Marseille, Campus de St. Jérôme, 13013 Marseille, France

²Université Bordeaux 1 - UMR CNRS 5218- IMS - 351 cours de la Libération, 33405 Talence Cedex, France
asma.rabaoui@fresnel.fr, audrey.giremus@ims-bordeaux.fr

Résumé – Cette communication propose une méthode pour approcher une loi a posteriori dans le cas où une partie des paramètres inconnus doit vérifier des contraintes égalité. À notre connaissance, il n’a pas été apporté de solution générique à cette question. Selon le contexte, les approches développées reposent soit sur le choix de lois a priori spécifiques, soit se limitent au calcul de certains paramètres de la loi a posteriori tels que son maximum. En alternative, nous présentons une modélisation du problème applicable quelles que soient les contraintes considérées. Elle repose sur la définition d’une loi a priori dépendant des contraintes et d’un paramètre dit de température dont la valeur ajuste le respect des contraintes. Sur la base de cette formulation, l’approximation de la loi a posteriori peut être obtenue en utilisant un échantillonneur séquentiel de Monte Carlo. Notre approche a été testée sur différents types de problèmes d’estimation et de contraintes. Dans les scénarios considérés, la valeur des contraintes en les estimés converge au fil des itérations de l’échantillonneur et il s’ensuit une diminution de l’erreur moyenne d’estimation.

Abstract – This communication presents a method for approximating a posterior distribution when a part of the unknown parameters must satisfy a set of equality constraints. To our knowledge, no generic solution has been developed to solve this problem. Depending on the context, the proposed approaches either consist in choosing a convenient prior distribution that enforces the constraints or are restricted to the computations of some parameters of the target distribution such as its maximum. As an alternative, we present in this paper a modeling of the problem applicable whatever the constraints. Based upon this formulation, the a posteriori distribution can be approximated by Sequential Monte Carlo samplers. The proposed methodology has been tested on various estimation issues and constraints. For the considered scenarios, the value of the constraints computed at the estimates converges with the iterations and the mean square error subsequently decreases.

1 Position du problème

Dans cette communication, nous nous intéressons au problème de l’estimation dans un cadre bayésien d’un ensemble de paramètres d’intérêt à partir d’observations indirectes et bruitées. L’objectif est donc d’approcher la distribution dite a posteriori qui fournit une représentation exhaustive de l’information disponible sur les paramètres inconnus. Connaissant cette distribution, il est alors possible de calculer des estimateurs ponctuels tels que la moyenne ou le maximum a posteriori (MAP), ainsi que des marges de confiance sur ces estimateurs. Cependant, nous considérons le cas difficile où une partie des paramètres inconnus doit satisfaire des contraintes égalité, comme cela apparaît dans différents contextes. On peut citer par exemple l’imagerie hyperspectrale [1] où un pixel est modélisé comme la somme de plusieurs contributeurs caractérisés par leurs abondances respectives. Non seulement ces abondances doivent être positives mais leur somme doit aussi être égale à 1. Un second exemple est l’optimisation de procédés industriels. En pétrochimie, cette dernière est fondée sur l’estimation de grandeurs physiques, telles que l’enthalpie, la température ou l’entropie, qui sont reliées entre elles par des équations de la physique fondamentale devant être strictement vérifiées. On parle alors de réconciliation de données [2]. Un dernier

exemple est l’estimation de l’attitude d’un véhicule spatial à partir de données gyrométriques. Celle-ci est en effet classiquement représentée par un quaternion dit unitaire dont la norme doit être égale à 1 [4].

Une première solution consiste à introduire une loi a priori dont le support est restreint à l’espace des solutions admissibles défini par les contraintes. Ainsi, en imagerie hyperspectrale, Eches et al. modélisent le vecteur des abondances par une loi de Dirichlet [3]. Selon l’expression de la vraisemblance, soit la loi a posteriori admet une expression analytique, soit il est nécessaire d’avoir recours à approximations numériques telles que les méthodes variationnelles ou les méthodes de Monte Carlo par chaînes de Markov (MCMC) comme l’algorithme de Métropolis-Hastings ou l’échantillonneur de Gibbs [5]. Cependant, selon le contexte, les contraintes peuvent prendre des formes très variées et il n’est pas toujours possible de représenter la loi a priori des paramètres inconnus par une distribution explicite ou facilement échantillonnable. Une alternative repose sur le calcul d’une estimation des paramètres inconnus par minimisation d’un critère pénalisé. Différentes formulations ont été présentées dans la littérature, fondées sur les multiplicateurs de Lagrange ou encore la pénalité quadratique [6]. Une limitation de ces approches est qu’elles fournissent

uniquement une estimation ponctuelle des paramètres d'intérêt, généralement au sens du maximum a posteriori, mais elles n'apportent aucune information sur la distribution a posteriori dans son ensemble (caractère multimodal, queue lourde etc.). En outre, elles peuvent converger vers des minima locaux dans le cas de contraintes non convexes. Dans cette communication, nous présentons une nouvelle approche qui présente l'intérêt d'être applicable quelles que soient les contraintes liant entre eux les paramètres inconnus et de fournir une approximation de la loi a posteriori dans son ensemble. Le principe est d'introduire une distribution a priori sur les paramètres d'intérêt, dépendant directement des contraintes et indexée par un paramètre dit de température. Ce dernier permet d'ajuster le respect des contraintes de la façon suivante : quand il tend vers zéro, le support de la loi a priori se réduit progressivement à l'espace des solutions admissibles. La distribution a posteriori des paramètres d'intérêt n'admettant pas une expression explicite dans le cas général, nous proposons de l'approcher en utilisant l'algorithme d'échantillonnage séquentiel de Monte Carlo introduit dans [7].

Cette communication est organisée comme suit : dans la partie 1, nous présentons la modélisation proposée pour résoudre un problème d'estimation sous contraintes égalité. Dans la partie 2, nous détaillons l'algorithme de Monte Carlo mis en œuvre pour approcher la loi a posteriori des paramètres inconnus. L'intérêt de la démarche présentée est illustré par des résultats de simulation dans la partie 4. Finalement, des conclusions et perspectives sont énoncées dans la partie 5.

2 Modélisation proposée

Nous notons \mathbf{x} le vecteur à estimer à partir d'observations concaténées dans un vecteur \mathbf{y} . La fonction de vraisemblance des paramètres inconnus s'écrit $p(\mathbf{y}|\mathbf{x})$. Dans notre contexte, le vecteur \mathbf{x} peut être décomposé en deux sous-parties : $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2]$, où \mathbf{x}_1 comprend les paramètres non sujets à des contraintes et \mathbf{x}_2 fait référence à l'ensemble des composantes qui sont liées entre elles par un jeu de m contraintes égalité $\{c_i(\mathbf{x}_2) = 0\}_{i=1, \dots, m}$. Le problème est de définir une loi a priori pour \mathbf{x}_2 assurant le respect de ces contraintes égalités. Nous proposons de considérer une distribution de la forme :

$$p_T(\mathbf{x}_2) = \frac{1}{Z_T} \exp\left(-\frac{\sum_{i=1}^m c_i^2(\mathbf{x}_2)}{T}\right), \quad (1)$$

où T est appelé paramètre de température et Z_T est une constante de normalisation. Cette loi a priori conduit à la loi a posteriori :

$$p_T(\mathbf{x}|\mathbf{y}) \propto \underbrace{p(\mathbf{y}|\mathbf{x}) p(\mathbf{x}_1|\mathbf{x}_2)}_{\gamma_T(\mathbf{x})} \exp\left(-\frac{\sum_{i=1}^m c_i^2(\mathbf{x}_2)}{T}\right), \quad (2)$$

avec \propto qui signifie proportionnel à. Ce choix de loi a priori présente de nombreux avantages. D'une part, dans le cas de contraintes linéaires, la distribution (1) est gaussienne et selon la fonction de vraisemblance, la loi a posteriori (2) admet une expression analytique. De façon plus générale, si les

contraintes sont multilinéaires, la distribution (1) n'est plus gaussienne mais la loi conditionnelle d'un paramètre sachant les autres reste gaussienne. Ainsi, il est facile d'explorer la loi a posteriori en utilisant un échantillonneur de Gibbs. D'autre part, une propriété intéressante est que le MAP s'obtient par minimisation du critère suivant :

$$J(\mathbf{x}) = -\ln p(\mathbf{y}|\mathbf{x}) - \ln p(\mathbf{x}_1|\mathbf{x}_2) + \frac{1}{T} \sum_{i=1}^m c_i^2(\mathbf{x}_2). \quad (3)$$

Cette expression correspond exactement à la résolution d'un problème d'optimisation sous contraintes égalité en utilisant la méthode de pénalité quadratique, avec $1/T$ qui joue le rôle de paramètre de pénalité. Cependant, comme souligné en introduction, l'objectif de ce travail n'est pas seulement de calculer le maximum a posteriori mais d'estimer toute la loi a posteriori. Il convient en outre de noter qu'il est classique d'interpréter la minimisation d'un critère pénalisé comme la maximisation d'une loi a posteriori, et réciproquement. Néanmoins, l'objectif est généralement de favoriser certaines propriétés pour les paramètres inconnus telles que la parcimonie ou la régularité et non d'imposer le respect stricte d'une contrainte. Le paramètre dit de régularisation $\frac{1}{T}$ est alors soit réglé empiriquement, soit estimé au sens du maximum de vraisemblance ou du MAP. Notre démarche diffère dans le traitement de ce paramètre. Nous nous assurons que les contraintes soient exactement vérifiées en faisant tendre T vers 0. Le problème qui se pose est que la loi a posteriori (2) n'admet généralement pas une expression explicite et doit être approchée.

3 Algorithme d'estimation bayésienne

3.1 Démarche proposée

Une solution classique pour approcher une loi a posteriori complexe est d'appliquer une méthode MCMC, qui consiste à simuler une chaîne de Markov de distribution invariante la loi d'intérêt. Cependant, il est difficile de générer directement des échantillons selon $p_T(\mathbf{x}|\mathbf{y})$ avec T très petit. Nous proposons plutôt de simuler itérativement des échantillons selon une suite de distributions $\{p_{T_t}(\mathbf{x}|\mathbf{y})\}_{t>0}$ où le paramètre de température T_t diminue selon un schéma de refroidissement du type $T_t = \beta T_{t-1}$, avec $0 < \beta < 1$. Ainsi, les échantillons générés se répartissent tout d'abord dans les régions de l'espace des paramètres de forte vraisemblance, puis se concentrent graduellement sur l'espace admissible défini par les contraintes. À cet effet, une première idée serait de faire décroître le paramètre de température au fil des itérations d'un échantillonneur classique de type Métropolis-Hastings ou Gibbs. L'inconvénient de cette approche est que la chaîne ainsi simulée n'est pas homogène et l'échantillon \mathbf{x}_t généré à l'itération t n'est pas garanti d'être distribué selon la loi a posteriori $p_{T_t}(\mathbf{x}|\mathbf{y})$.

3.2 Echantillonneur séquentiel de Monte Carlo

Pour contourner cette difficulté, nous proposons plutôt d'appliquer à notre modèle l'algorithme générique développé ré-

vement par Del Moral et al. [7] pour générer des échantillons selon une suite de distributions définies sur un même espace de probabilité. Le principe est de mettre en œuvre plusieurs échantillonneurs en parallèle et de les faire interagir par une étape d'échantillonnage d'importance.

Si on note $\mathbf{x}_t^{(i)}$ l'échantillon issu de la i ème chaîne à l'itération t , aussi appelé particule, alors il est distribué selon la loi η_t définie récursivement par :

$$\eta_t(\mathbf{x}_t) = \int K_t(\mathbf{x}_{t-1}, \mathbf{x}_t) \eta_{t-1}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1}, \quad (4)$$

où $K_t(\cdot, \cdot)$ est le noyau de transition de l'échantillonneur qui dans notre cas varie avec le paramètre de température et dépend donc de l'itération considérée. La loi cible est alors approchée par une distribution discrète :

$$p_{T_t}(\mathbf{x}_t | \mathbf{y}) \simeq \sum_{i=1}^N w_t^{(i)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)}) \quad (5)$$

où δ est une mesure de Dirac et $w_t^{(i)}$ est le poids qui vérifie $w_t^{(i)} \propto p_{T_t}(\mathbf{x}_t^{(i)} | \mathbf{y}) / \eta_t(\mathbf{x}_t^{(i)})$, avec la contrainte que la somme des poids doit être égale à 1.

Cependant, il n'est généralement pas possible de calculer l'intégrale définie par (4) et donc d'évaluer les poids $w_t^{(i)}$. Cette difficulté est résolue dans [7] par une technique de variable auxiliaire. La loi cible est alors représentée comme étant la loi marginale d'une loi étendue :

$$\tilde{p}_{T_t}(\mathbf{x}_{0:t} | \mathbf{y}) = p_{T_t}(\mathbf{x}_t | \mathbf{y}) \prod_{k=0}^{t-1} L_k(\mathbf{x}_{k+1}, \mathbf{x}_k), \quad (6)$$

où $\mathbf{x}_{0:t} = (\mathbf{x}_0, \dots, \mathbf{x}_t)$ et $L_k(\mathbf{x}_{k+1}, \mathbf{x}_k)$ représente un noyau de transition choisi par l'utilisateur. Comme en filtrage particulaire standard, cette loi étendue \tilde{p}_{T_t} peut être approchée par échantillonnage d'importance :

$$\tilde{p}_{T_t}(\mathbf{x}_{0:t} | \mathbf{y}) \simeq \sum_{i=1}^N w_t^{(i)} \delta(\mathbf{x}_{0:t} - \mathbf{x}_{0:t}^{(i)}). \quad (7)$$

L'intérêt de considérer un espace étendu est qu'il n'est plus nécessaire d'intégrer par rapport aux paramètres des itérations passées pour calculer les poids. Ainsi, la distribution de la suite $\mathbf{x}_{0:t}^{(i)}$ admet cette fois une expression analytique :

$$\eta_t(\mathbf{x}_{0:t}) = \eta_t(\mathbf{x}_0) \prod_{k=1}^t K_k(\mathbf{x}_{k-1}, \mathbf{x}_k). \quad (8)$$

Il s'ensuit une expression récursive pour les poids :

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{\gamma_{T_t}(\mathbf{x}_t^{(i)}) L_{t-1}(\mathbf{x}_t^{(i)}, \mathbf{x}_{t-1}^{(i)})}{\gamma_{T_{t-1}}(\mathbf{x}_{t-1}^{(i)}) K_t(\mathbf{x}_{t-1}^{(i)}, \mathbf{x}_t^{(i)})}, \quad (9)$$

où γ_{T_t} est défini dans l'équation (2). Enfin, la loi marginale $p_{T_t}(\mathbf{x}_t | \mathbf{y})$ est simplement obtenue à partir de (7) par marginalisation.

Un défaut de cet algorithme, inhérent aux méthodes de Monte

Carlo séquentielles, est que le système de particules tend à dégénérer de sorte que tous les poids sont proches de 0 sauf 1. Pour prévenir ce problème, le choix du noyau de transition $L_{t-1}(\mathbf{x}_t, \mathbf{x}_{t-1})$ est crucial. Son expression optimale au sens où elle minimise la variance des poids est donnée dans [7] mais elle est difficile à mettre en œuvre en pratique. Il est montré que dans le cas où les lois cibles varient lentement d'une itération sur l'autre, ce qui dans notre cas correspond à une décroissance lente de la température, une bonne approximation de ce noyau optimal est :

$$L_{t-1}(\mathbf{x}_t, \mathbf{x}_{t-1}) = \frac{p_{T_t}(\mathbf{x}_{t-1} | \mathbf{y}) K_t(\mathbf{x}_{t-1}, \mathbf{x}_t)}{p_{T_t}(\mathbf{x}_t | \mathbf{y})}. \quad (10)$$

Les poids se calculent alors récursivement de la façon suivante :

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{\gamma_{T_t}(\mathbf{x}_{t-1}^{(i)})}{\gamma_{T_{t-1}}(\mathbf{x}_{t-1}^{(i)})}. \quad (11)$$

Remarque : lien avec le recuit-simulé. La méthode proposée présente des connections avec l'algorithme d'optimisation du recuit simulé dont nous rappelons brièvement le principe ci-après. Soit $\pi(\mathbf{x})$ une loi de probabilité dont on souhaite calculer le maximum. Le recuit simulé consiste à échantillonner selon une suite de distributions $(\pi(\mathbf{x}))^{\frac{1}{T_t}}$, avec T_t un paramètre de température qui suit un schéma de refroidissement. Quand T_t tend vers l'infini, la loi échantillonnée est uniforme de sorte que tout l'espace des paramètres est parcouru. Inversement, quand T_t tend vers 0, elle tend vers un Dirac localisé au niveau du maximum global de la loi cible $\pi(\mathbf{x})$. De cette façon, il est possible de converger progressivement vers le maximum global sans être piégé dans des extrema locaux. Notre approche se démarque du recuit-simulé à deux titres. 1/ Notre objectif n'est pas de résoudre un problème d'optimisation mais d'approcher toute la loi cible par MCMC. 2/ Le paramètre de température n'est pas appliqué à toute la loi cible mais intervient uniquement dans la loi a priori des paramètres inconnus de façon à contraindre progressivement les échantillons à respecter un jeu de contraintes égalité.

4 Résultats de simulation

L'approche développée a été testée sur données simulées pour différents types de vraisemblances et de contraintes et nous présentons l'un des scénarios à titre d'exemple. Nous nous intéressons à l'estimation d'un vecteur de paramètres inconnus $\mathbf{x} = [x_1, \dots, x_n]^T$ qui est relié linéairement à un vecteur d'observations \mathbf{y} de mêmes dimensions $n \times 1$: $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, où \mathbf{w} est un vecteur de bruits dont les composantes sont centrées, indépendantes, et de variance $\gamma = 0, 1$. En outre, les composantes du vecteur \mathbf{x} doivent satisfaire la contrainte $\sum_{i=1}^n |x_i| = 1$. La variance γ étant supposée inconnue, l'objectif est d'estimer conjointement cette dernière et les $\{x_i\}_{i=1, \dots, n}$.

L'échantillonneur utilisé est de type Gibbs avec des étapes de Métropolis-Hastings pour simuler les lois conditionnelles des composantes x_i qui sont non triviales. Pour la variance γ , nous choisissons une loi a priori inverse-gamma de paramètres

d'échelle et de forme égaux à 2 et 8, respectivement. Finalement, le paramètre de la loi de refroidissement de la température est $\beta = 0,999$ avec une température initiale $T_0 = 1$. Il convient de noter qu'une valeur trop faible de β poserait un problème pour calculer les poids de notre algorithme selon la formule (11).

Le nombre total d'itérations est fixé à 20000 et nous considérons 50 échantillonneurs en parallèle. Les paramètres inconnus sont estimés au sens du MAP. Notre approche est tout d'abord appliquée dans le cas où le nombre de composantes de \mathbf{x} est $n = 5$, puis sa robustesse à des problèmes de plus grande dimension est testée en considérant $n = 50$. Pour ce dernier cas, nous avons représenté sur la figure 1 la valeur de la contrainte en les estimés au fil des itérations de l'échantillonneur et nous pouvons observer qu'elle converge bien vers 0. Par ailleurs, sur la figure 2 sont tracées les probabilités a posteriori de deux des composantes du vecteur \mathbf{x} , estimées à partir des particules à la dernière itération. Ces lois apportent des informations complémentaires notamment sur la pertinence de l'estimé MAP. Finalement, dans le tableau 1, nous comparons notre algorithme à un échantillonneur MCMC classique ne prenant pas en compte les contraintes égalité. Les résultats reportés correspondent à des valeurs moyennes calculées à partir de 50 réalisations correspondant à des valeurs différentes du bruit de mesure. Notre méthode améliore sensiblement la racine de l'erreur quadratique moyenne (REQM) mais permet surtout d'assurer le respect strict des contraintes. Cependant, la contrepartie est un coût calculatoire plus important. Pour un Mac OS X 2,53 GHZ Intel Core 2 Duo, les 20000 itérations prennent 49s sous Matlab dans le cas où $n = 5$ et 12 minutes dans le cas où $n = 50$. Cependant, une grande partie des calculs est parallélisable et comme l'illustre la figure 1, il aurait été possible de diminuer significativement le nombre d'itérations sans dégrader fortement le respect des contraintes.

		MCMC	Méthode proposée
$n = 5$	REQM (\mathbf{x})	0,008	0,005
	REQM (γ)	0,074	0,048
	$c(\hat{\mathbf{x}})$	$0,002(\pm 10^{-2})$	$10^{-8}(\pm 10^{-5})$
$n = 50$	REQM (\mathbf{x})	0,012	0,009
	REQM (γ)	0,047	0,054
	$c(\hat{\mathbf{x}})$	$0,054(\pm 10^{-2})$	$5 \times 10^{-6}(\pm 10^{-5})$

TABLE 1 – Comparaison de l'approche proposée et d'un algorithme MCMC

5 Conclusion

Dans cette communication, nous présentons une méthode générique pour échantillonner une loi a posteriori dans le cas où certains paramètres inconnus sont liés entre eux par des contraintes égalités. L'approche proposée consiste à définir une loi a priori dépendant conjointement des contraintes et d'un paramètre de température qui permet d'ajuster leur respect. La loi a posteriori résultante est simulée par un échantillonneur séquentiel de Monte Carlo comportant un schéma de refroidissement.

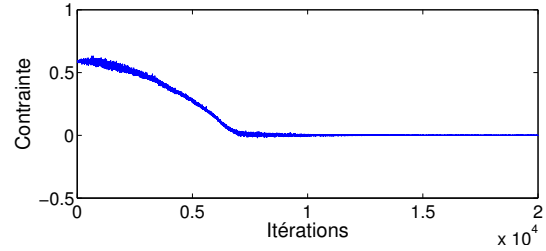


FIGURE 1 – Valeurs de la contrainte au cours des itérations pour $n = 50$.

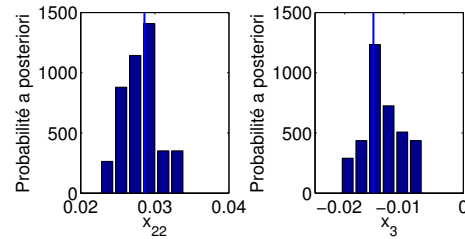


FIGURE 2 – Probabilités a posteriori estimées pour deux des composantes du vecteur \mathbf{x} quand $n = 50$. Les lignes verticales indiquent les valeurs réelles.

dissement. La méthode a été testée sur différents scénarios. Elle permet à la fois de calculer des estimés robustes aux problèmes de maxima locaux et de garantir un strict respect des contraintes. Une perspective intéressante est la généralisation au cas de contraintes inégalités. Une possibilité serait alors de définir des lois a priori fondées sur les fonctions barrière utilisées en optimisation.

Références

- [1] E. Chouzenoux, S. Moussaoui and J. Idier. *Algorithme primal-dual de points intérieurs pour l'estimation pénalisée des cartes d'abondances en imagerie hyperspectrale*. Actes du 23e colloque GRETSI, Bordeaux, France, 5-8 septembre 2011.
- [2] S. Narasimhan, C. Jordache. *Data reconciliation and gross error detection : an intelligent use of process data*. Gulf Publishing Company, Houston, Texas, 2000.
- [3] O. Eches, N. Dobigeon, J.-Y. Tourneret. *Estimating the number of endmembers in hyperspectral images using the normal compositional Model and a hierarchical Bayesian model*. IEEE J. Sel. Topics Signal Process., vol. 4, no. 3, pp. 1-10, Juin 2010.
- [4] M. Shuster, S. Oh. *Three-axis attitude determination from vector observations*. Journal of Guidance, Control and Dynamics, vol. 4, no. 1, pp. 70-77, 1981.
- [5] C. P. Robert. *Le Choix bayésien : principes et pratique*. Springer, 2005.
- [6] J. Nocedal, S.J. Wright. *Numerical Optimization : Second Edition*. Springer Series in Operations Research, 2006.
- [7] P. Del Moral, A. Doucet, A. Jasra. *Sequential Monte Carlo Samplers*. Journal of the Royal Statistical Society, Series B, vol. 68, no. 3, pp. 411-336, 2006.