

APPROCHE BAYÉSIENNE POUR LA RÉOLUTION DES PROBLÈMES INVERSES EN IMAGERIE¹

Guy DEMOMENT et Jérôme IDIER
Laboratoire des signaux et systèmes (CNRS/SUPÉLEC/UPS)
Supélec, Plateau de Moulon, 91192 GIF-SUR-YVETTE Cedex, France

1 Introduction

Les problèmes inverses ont connu un intérêt croissant depuis une vingtaine d'années parce qu'on les rencontre dans de très nombreuses applications : optique, radar, calorimétrie, spectroscopie, géophysique, acoustique, radioastronomie, contrôle non-destructif, génie biomédical, instrumentation et imagerie en général. Il existe, schématiquement, deux grandes communautés scientifiques qui s'intéressent à ces problèmes :

— celle de la *physique mathématique*, que l'on peut rattacher aux travaux fondateurs de Tikhonov dans les années 1960, dont P.C. Sabatier fut un des pionniers en France (avec son ATP du même nom), et dont une revue représentative est *Inverse problems* ;

— celle du *traitement statistique des données*, que l'on peut rattacher aux travaux de Franklin à la fin des années 1960 [8], dont les frères Geman ont constitué les accélérateurs en traitement d'image il y a une dizaine d'années [9], et dont une revue représentative est *IEEE Trans. on Image Processing*.

On peut dire, très grossièrement, que les uns abordent le problème en dimension infinie, avec les questions d'existence, d'unicité et de stabilité qui deviennent très compliquées avec des problèmes directs non linéaires, et le résolvent numériquement en dimension finie ; alors que les autres partent d'un problème dont la discrétisation est déjà faite et non remise en cause, et profitent du caractère fini du problème pour introduire une information *a priori* élaborée au travers de modèles probabilistes.

Nous nous proposons d'indiquer quels sont, de notre point de vue, l'état de l'art et les questions ouvertes dans le domaine de la résolution des problèmes inverses par des approches probabilistes.

2 Discrétisation

Nous avons vu plus haut [cf. chap. 1] qu'une première description d'un problème direct fait en général intervenir des fonctions de variables réelles (temps, fréquence, variables d'espace, . . .), représentant les grandeurs physiques mises en jeu : grandeurs accessibles à la mesure et grandeurs d'intérêt inconnues (que nous appellerons par la suite *objet*). Nous avons aussi indiqué dans l'introduction que l'approche statistique de l'inversion est caractérisée par une discrétisation préalable du problème direct. Les données expérimentales disponibles sont presque toujours constituées de la mesure de grandeurs physiques accessibles en un nombre nécessairement fini de points du domaine de définition de leurs variables. Elles sont donc naturellement discrètes, et nous les regrouperons dans le vecteur \mathbf{y} . Dans les approches statistiques, l'objet inconnu est lui aussi discrétisé, d'emblée, par décomposition sur un nombre fini de fonctions. Si celles-ci sont des éléments d'une base de l'espace auquel l'objet appartient, la décomposition est nécessairement tronquée. En imagerie par exemple, on utilise comme fonctions de base dans la très grande majorité des cas les indicatrices des pixels, ou des sinus cardinaux, selon que l'objet est implicitement supposé à support limité ou à

1. Chapitre à paraître en février 1999 dans l'ouvrage collectif « *Problèmes inverses* » de la collection ARAGO (n° 21), publié par l'Observatoire français des techniques avancées (OFTA), sous la direction de M. Bonnet.

spectre limité. On commence aussi à utiliser des bases d'ondelettes [4]. Le point de départ est donc constitué d'un *modèle paramétré* par le vecteur \mathbf{x} des coefficients de la décomposition, c'est-à-dire d'un ensemble d'hypothèses exclusives dont chacune est indexée par la valeur des coefficients. Cet espace d'hypothèses est donc l'ensemble des valeurs possibles de ces paramètres inconnus, $\mathcal{H} = \{x_i\}$. Le choix de ces fonctions de base fait évidemment partie du problème d'inversion, même s'il est assez peu abordé.

3 Problème mal-posé

Après discrétisation, le problème direct prend la forme générale suivante $\mathbf{A}(\mathbf{x}, \mathbf{y}) = 0$, où \mathbf{A} est un opérateur reliant l'objet inconnu \mathbf{x} aux données expérimentales \mathbf{y} , appelé aussi *réponse instrumentale*. Il prend souvent la forme explicite $\mathbf{y} = \mathbf{A}(\mathbf{x})$, voire linéaire $\mathbf{y} = \mathbf{A} \mathbf{x}$. Connaissant \mathbf{A} et \mathbf{x} , le calcul de \mathbf{y} est le *problème direct*. Inversement, le calcul de \mathbf{x} connaissant \mathbf{A} et \mathbf{y} est le *problème inverse*. Mais ces problèmes inverses présentent la caractéristique désagréable d'être aussi très souvent des *problèmes mal-posés*, en deux sens.

— Premièrement, l'opérateur \mathbf{A} est habituellement singulier dans le sens où il existe une classe \mathcal{K} de solutions $\mathbf{x} \in \mathcal{K}$ telles que $\mathbf{A} \mathbf{x} = 0$ (le noyau $\text{Ker}(\mathbf{A}) = \mathcal{K}$ n'est donc pas vide). Tout élément de \mathcal{K} peut être ajouté à toute solution pour donner une autre solution, et nous ne pouvons donc inverser la relation directe pour déterminer uniquement \mathbf{x} à partir de \mathbf{y} . Ce manque d'unicité fait que le problème inverse discret est mal-posé au sens d'Hadamard [cf. chap.1]. Cette situation survient à chaque fois que la réponse instrumentale détruit une partie de l'information nécessaire pour reconstruire l'objet.

— Deuxièmement, aucun dispositif expérimental n'est complètement affranchi d'une incertitude, dont l'origine la plus simple est la précision finie des mesures. Il est donc plus réaliste de considérer que l'objet recherché et les mesures sont reliés par une équation de la forme

$$\mathbf{y} = \mathbf{A}(\mathbf{x}) \diamond \mathbf{b}. \quad (1)$$

Dans cette expression générale, \mathbf{A} est un opérateur décrivant la partie essentielle de l'expérience, et $\diamond \mathbf{b}$ prend en compte la dégradation de cette représentation idéale par différentes sources d'erreur (de discrétisation, de mesure) regroupées sous le terme de *bruit*. Quand le mécanisme d'observation peut être approché par une distorsion linéaire et l'ajout d'un bruit, alors l'équation (1) se réduit à

$$\mathbf{y} = \mathbf{A} \mathbf{x} + \mathbf{b}. \quad (2)$$

La présence de ce bruit a pour effet d'« élargir » l'ensemble \mathcal{K} puisque tout élément \mathbf{x} tel que $\mathbf{A} \mathbf{x} = \boldsymbol{\epsilon}$, où $\boldsymbol{\epsilon}$ est « petit » par rapport au niveau supposé de bruit, peut être ajouté à toute solution possible pour obtenir une autre solution acceptable. En pratique cela se traduit par l'instabilité des tentatives d'inversion directe de la relation (2), des petits changements dans les données entraînant de grandes variations de la solution calculée. Ce n'est plus, cette fois, nécessairement un problème mal-posé au sens d'Hadamard puisque la solution peut très bien — lorsque le noyau \mathcal{K} est vide — être unique et dépendre continûment des données. L'instabilité provient du *mauvais conditionnement* de \mathbf{A} [6].

On voit donc que, dans ces problèmes mal-posés, l'obtention d'une solution n'est pas du tout un problème de déduction mathématique, c'est un problème d'*inférence*, c'est-à-dire de traitement de l'information : comment tirer les meilleures conclusions possibles de l'information incomplète qui est disponible. Toute méthode d'inférence scientifique devrait, pour être acceptable, 1° prendre en compte *toute l'information pertinente disponible* et, 2° *éviter* soigneusement de supposer une information disponible alors qu'elle ne l'est pas. La modélisation probabiliste est un moyen commode et cohérent de décrire une situation d'information incomplète. Nous allons voir comment elle conduit à une approche statistique bayésienne.

4 Inférence statistique (1) : vraisemblance

Il faut préciser d'emblée que tout problème auquel on s'attaque par une approche bayésienne doit être *bien posé* dans le sens où une information suffisante doit être apportée pour permettre d'attribuer sans ambiguïté les distributions de probabilité nécessaires au calcul. Ceci signifie, au minimum, qu'un ensemble exhaustif de possibilités doit être spécifié au début de chaque problème. Nous l'appellerons *espace des données* (ou des *épreuves*) s'il s'agit des résultats possibles de l'expérience, ou *espace des hypothèses* s'il spécifie les hypothèses que nous voulons vérifier. Il est utile aussi de distinguer deux classes de problèmes, appelées *estimation* et *choix de modèle*. L'estimation étudie les conséquences du choix d'un modèle particulier, supposé « vrai », alors que le choix de modèle a pour but d'en retenir un par comparaison avec une ou plusieurs alternatives.

Dans un problème d'estimation, on suppose que le modèle est vrai pour *une* valeur (inconnue) de ses paramètres, et on explore les contraintes imposées aux paramètres par les données. L'espace des hypothèses est alors l'ensemble de toutes les valeurs possibles des paramètres, $\mathcal{H} = \{x_i\}$. Les données consistent en un ou plusieurs échantillons ; pour rendre le problème bien posé, l'espace de tout les échantillons possibles, $\mathcal{S} = \{y_i\}$, doit être aussi précisé. Les espaces \mathcal{H} et \mathcal{S} peuvent être tous deux discrets ou continus.

Désignant par x_0 la vraie valeur du paramètre, nous pouvons nous attaquer au problème d'estimation en calculant la probabilité que chacune des valeurs possibles du paramètre soit la vraie valeur. Désignons par D la proposition affirmant les valeurs des données expérimentales réellement observées, par H la proposition $x_0 = x$ affirmant que l'une des valeurs possibles du paramètre, x_0 , est la vraie valeur, et par I l'environnement logique du problème. C'est cet environnement logique qui définit notre problème en spécifiant l'espace des hypothèses, l'espace des données, comment les hypothèses (valeurs des paramètres) et les données sont reliées, et toute information supplémentaire que nous pourrions avoir sur les hypothèses ou les données. Nous pourrions définir I comme la proposition affirmant 1° que la vraie valeur du paramètre est dans \mathcal{H} , 2° que les données observées consistent en N échantillons de l'espace \mathcal{S}^N , 3° la manière dont les paramètres sont reliés aux données (le modèle direct \mathbf{A}), et 4° toute information supplémentaire. Bien entendu, la nature physique des paramètres et des données est spécifiée implicitement dans \mathcal{H} , \mathcal{S} et \mathbf{A} .

La première étape dans toute méthode d'inférence statistique destinée à résoudre un problème tel que (2) consiste à choisir une distribution de probabilité décrivant notre information — ou notre incertitude — sur les erreurs \mathbf{b} : $q(\mathbf{b} | I)$. C'est une étape essentielle puisqu'elle permet d'en déduire la distribution *directe*, ou d'*échantillonnage* :

$$p(\mathbf{y} | \mathbf{x}, I) = q(\mathbf{y} - \mathbf{A} \mathbf{x} | I). \quad (3)$$

Dans la très grande majorité des cas, on choisit une distribution gaussienne centrée pour les erreurs, indépendante de \mathbf{x} , ce qui donne :

$$p(\mathbf{y} | \mathbf{x}, I) = (2\pi |\mathbf{R}|)^{-N/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{A} \mathbf{x})^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{A} \mathbf{x}) \right\},$$

où \mathbf{R} désigne la matrice de covariance de la distribution $q(\mathbf{b} | I)$. Elle est souvent diagonale, voire même proportionnelle à l'identité. Une question se pose immédiatement : quel sens doit-on donner à un tel choix et dans quelles situations un tel modèle est-il approprié ?

Avec une interprétation *fréquentiste* d'une probabilité, la distribution pour le bruit devrait être la distribution de fréquence de ses valeurs dans un très grand nombre de répétitions des mesures. Elle est alors justifiée par référence au théorème central limite qui dit, sous des conditions assez larges, que si le bruit dans un échantillon des données résulte d'un grand nombre d'effets élémentaires cumulés, « aléatoires » et indépendants, la distribution gaussienne est une bonne approximation de sa véritable distribution de fréquence. Mais,

excepté les fluctuations d'origine électronique dans une chaîne de mesure, le bruit n'est pas en général le résultat d'un grand nombre d'effets indépendants (songeons par exemple aux erreurs de discrétisation qui dépendent de la solution \mathbf{x}_0). De plus, pour pouvoir effectuer une inférence avec cette interprétation, il serait nécessaire que nous disposions de nombreux résultats d'autres mesures pour pouvoir déterminer ces fréquences, ce qui est une situation expérimentale extrêmement rare.

Cette « hypothèse » gaussienne n'est donc pas une hypothèse sur le caractère « aléatoire » du bruit. Nous ne prétendons nullement que le phénomène donnant naissance au bruit soit véritablement aléatoire et suive une distribution gaussienne. Ce n'est même pas une hypothèse à proprement parler ; c'est plutôt le choix le moins compromettant — ou le plus conservateur — que nous puissions faire pour la distribution du bruit dans une situation d'incertitude. Nous supposons ici deux choses, 1° que le bruit peut prendre toute valeur réelle mais qu'il est de moyenne nulle, c'est-à-dire qu'il n'y a pas d'erreur de mesure systématique (ou que s'il y en a, nous avons été capable de la détecter et de la corriger), et 2° que nous nous attendons à ce qu'il y ait une « échelle typique » du bruit, c'est-à-dire que les grandes contributions du bruit ne sont pas aussi probables que les petites. Autrement dit, nous pensons que la distribution pour le bruit doit avoir une moyenne nulle et un écart-type fini, même si nous n'avons pas d'idée précise de la valeur de ce dernier. Par contre, nous n'avons aucune idée de l'existence ou non de cumulants d'ordre supérieur à deux. Dans ces conditions, le choix le moins compromettant vis-à-vis des caractéristiques que nous ignorons — que l'on peut justifier formellement par des principes informationnels [14] — est celui d'une distribution gaussienne. Et si l'on suspecte que les composantes du bruit affectant les N échantillons ont des échelles différentes et sont corrélées, la matrice de covariance de la distribution est là pour cela. Il n'est pas nécessaire d'en spécifier la valeur, mais si elle est inconnue, ses éléments, regroupés dans un vecteur d'*hyperparamètres* $\boldsymbol{\theta}$, viendront en général compliquer le problème. On les appelle paramètres de *nuisance* pour cette raison.

Ce choix est approprié à chaque fois que cette information est tout ce que nous savons du bruit. Comme c'est une situation fréquente, il est souvent fait. Si nous avons des informations supplémentaires sur le bruit, qui nous conduiraient à choisir une distribution non gaussienne, nous pouvons les inclure de la même manière, mais cela ne conduira à des résultats nettement améliorés que si la distribution diffère nettement d'une gaussienne. Il existe ainsi des situations — comme l'imagerie à faible taux de comptage de particules — où les données sont entières et de faible valeur. Le choix d'une distribution poissonnienne peut alors améliorer les résultats.

Munis de cette seule distribution directe $p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}, I)$, nous pourrions définir la solution du problème inverse comme étant celle du *maximum de vraisemblance*, la vraisemblance étant la distribution directe dans laquelle la variable \mathbf{y} prend sa valeur observée et le paramètre \mathbf{x} devient la variable :

$$\hat{\mathbf{x}}_{\text{MV}} = \arg \max_{\mathbf{x} \in \mathcal{H}} p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}, I).$$

La solution des *moindres carrés* est un cas particulier de celle du maximum de vraisemblance, lorsque la distribution directe est gaussienne :

$$\hat{\mathbf{x}}_{\text{MC}} = \arg \min_{\mathbf{x} \in \mathcal{H}} [\mathbf{y} - \mathbf{A} \mathbf{x}]^t \mathbf{R}^{-1} [\mathbf{y} - \mathbf{A} \mathbf{x}].$$

Introduite de cette manière, il s'agit toujours d'une méthode des moindres carrés *pondérés* (par la matrice \mathbf{R}^{-1}) qui possède la propriété indispensable d'invariance par changement d'unités dans \mathcal{H} et \mathcal{S} . Dans beaucoup de situations simples, cette méthode d'inférence fournit toute l'information recherchée, mais dans la plupart des problèmes inverses, la distribution directe ne contient pas toute l'information nécessaire pour rendre le problème bien posé et elle ne fournit pas tout l'appareil technique nécessaire au calcul :

(1) Dans le cas particulier d'un problème indéterminé $\mathbf{y} = \mathbf{A} \mathbf{x}$ où \mathbf{A} est singulier (problème dit d'*inversion généralisée*), il n'y a pas de « bruit », et donc pas de distribution directe, sauf

dans le sens rudimentaire où $p(\mathbf{y} | \mathbf{x}, I)$ est constant si \mathbf{x} est dans la classe \mathcal{C} des antécédents possibles de \mathbf{y} , et nul sinon. La vraisemblance étant constante dans la classe \mathcal{C} , sa maximisation ne nous est d'aucune aide pour choisir dans cette classe. L'essence du problème ne réside pas dans la présence d'un bruit « aléatoire » perturbant nos données, mais plutôt dans le fait que notre information soit incomplète, bien qu'essentiellement non bruitée.

(2) Dans le cas linéaire (2), la matrice \mathbf{A} du problème direct est souvent *mal conditionnée*. L'opérateur de résolution $\mathbf{A}^+ = (\mathbf{A}^t \mathbf{R}^{-1} \mathbf{A}) \mathbf{R}^{-1} \mathbf{A}^t$ est instable et la solution $\hat{\mathbf{x}}_{\text{MV}} = \mathbf{A}^+ \mathbf{y}$, est inacceptable : l'amplification du bruit est excessive.

(3) Le problème peut avoir des paramètres de nuisance sans intérêt pour nous, et en grand nombre. Ainsi, lorsque la matrice \mathbf{R} est pleine, il vient s'ajouter $N(N-1)/2$ hyperparamètres qui, lorsqu'ils sont inconnus, doivent en général être estimés par maximum de vraisemblance comme les paramètres d'intérêt \mathbf{x} , et le maximum global peut alors n'être plus un point mais toute une région.

(4) Nous pouvons avoir une information hautement pertinente sur la solution recherchée. Par exemple, nous pouvons savoir qu'elle doit être positive, ou qu'elle doit satisfaire certaines contraintes (comme en imagerie astronomique où l'intégrale de l'objet peut être connue par ailleurs), ou qu'elle est constituée de régions homogènes séparées par des contours nets. De telles informations ne sont pas contenues dans la distribution directe et il serait déraisonnable de ne pas en tenir compte.

(5) Dans beaucoup de problèmes, il est nécessaire de disposer non seulement d'une solution, mais aussi d'une mesure de la confiance à lui accorder. Munis de la seule distribution directe (3), les *intervalles de confiance* fournis par l'approche fréquentiste ne nous renseignent que sur le comportement à *long terme* de la solution, c'est-à-dire sur son comportement moyen dans un très grand nombre de répétitions de l'expérience. Or nous ne disposons que des résultats d'une seule expérience, d'ailleurs très souvent non reproductible.

(6) Enfin, l'estimation des paramètres d'un modèle supposé valable n'est souvent qu'une étape, et l'on peut avoir besoin de juger les mérites relatifs de différents modèles.

Il est donc nécessaire d'aller au-delà de l'inférence par maximum de vraisemblance. Toutes les extensions évoquées ci-dessus sont « automatiquement » fournies par l'approche bayésienne.

5 Inférence statistique (2) : règle de Bayes

L'inférence bayésienne est ainsi nommée parce qu'elle fait un large usage de la règle de Bayes, conséquence elle-même d'une règle fondamentale du calcul des probabilités, la *règle du produit* [5]. Soit H une hypothèse dont nous voulons évaluer la véracité, D un ensemble de données en rapport avec cette hypothèse, et I une proposition définissant l'environnement logique du problème. La règle du produit stipule que :

$$P(H D | I) = P(H | D, I) P(D | I) = P(D | H, I) P(H | I)$$

où, par exemple, $P(H | D, I)$ désigne de manière habituelle la probabilité que H soit vraie sachant D et I . On en tire la règle de Bayes :

$$P(H | D, I) = \frac{P(H | I) P(D | H, I)}{P(D | I)}$$

qui n'est rien d'autre qu'une règle d'*apprentissage* : elle nous dit comment nous devons ajuster la probabilité attribuée à la véracité d'une hypothèse lorsque notre état de connaissance change avec l'acquisition de données. La probabilité *a posteriori* pour H , $P(H | D, I)$, est obtenue en multipliant sa probabilité *a priori*, $P(H | I)$, par la probabilité d'avoir observé les

données D en supposant l'hypothèse vraie, $P(D | H, I)$, et en divisant le tout par la probabilité d'avoir observé les données indépendamment du fait que l'hypothèse soit vraie ou non, $P(D | I)$. Ce dernier terme, parfois appelé *vraisemblance globale*, joue le rôle d'une constante de normalisation.

Dans le cas d'un problème inverse tel que (1), et en supposant que les distributions de probabilité concernées sont à densité, la règle de Bayes devient :

$$p(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta}, I) = \frac{p(\mathbf{x} | \boldsymbol{\theta}, I) p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}, I)}{p(\mathbf{y} | \boldsymbol{\theta}, I)}. \quad (4)$$

Dans cette équation, $\boldsymbol{\theta}$ est un vecteur d'*hyperparamètres* constitué des paramètres des distributions *a priori* des erreurs et de l'objet. L'approche bayésienne fournit ainsi une réponse claire et sans ambiguïté au problème de l'inversion des données expérimentales : après avoir observé *les seules données* \mathbf{y} , l'incertitude sur l'objet \mathbf{x} est entièrement décrite par une distribution de probabilité dont on peut déduire aussi bien un estimateur ponctuel qu'une région d'incertitude.

6 Liens avec les méthodes déterministes de régularisation

Avec une approche bayésienne, le résultat (4) est la solution du problème inverse puisqu'il fournit, sous la forme d'une distribution de probabilité, toute l'information disponible sur l'objet \mathbf{x} . S'il faut ensuite retenir une solution particulière parmi toutes les solutions probables, un choix très répandu consiste à retenir celle qui maximise la distribution *a posteriori* :

$$\hat{\mathbf{x}}_{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta}, I).$$

Ce n'est qu'une possibilité parmi d'autres, qui correspond à minimiser un coût de décision en « *tout-ou-rien* », limite (lorsque $\varepsilon \rightarrow 0$) du coût moyen $P(\|\hat{\mathbf{x}} - \mathbf{x}_0\| > \varepsilon)$. Comme la distribution directe et la distribution *a priori* prennent souvent les formes respectives :

$$p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}, I) \propto \exp\{-\mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x})/2\sigma^2\} \quad \text{et} \quad p(\mathbf{x} | \boldsymbol{\theta}, I) \propto \exp\{-\alpha\mathcal{F}(\mathbf{x})/2\sigma^2\},$$

la recherche d'une solution $\hat{\mathbf{x}}_{\text{MAP}}$ équivaut à la minimisation du critère *composé* :

$$\mathcal{J}(\mathbf{x}) = \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \alpha\mathcal{F}(\mathbf{x}) \quad 0 < \alpha < +\infty. \quad (5)$$

On reconnaît là une forme désormais classique de *régularisation* telle qu'elle est apparue, dans les problèmes inverses en dimension finie, dans les travaux de Phillips et Twomey [16, 23], contemporains de ceux de Tikhonov [20]. L'idée de base est de renoncer à l'espoir d'accéder à la solution « exacte » à partir de données imparfaites, de considérer que l'équation (2) définit une classe de *solutions admissibles*, et de rechercher dans cette classe une solution qui peut être considérée comme étant physiquement raisonnable, c'est-à-dire compatible avec une certaine information *a priori*. Ceci s'obtient usuellement en recherchant une solution $\hat{\mathbf{x}}$ qui minimise un critère tel que (5) spécifiquement conçu pour : 1° être, jusqu'à un certain point, « fidèle » aux données (premier terme du critère, presque toujours de la forme $\mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_W^2$), et, 2° renforcer certaines propriétés « souhaitables » qui résument notre connaissance *a priori* sur la solution (second terme, quadratique lui aussi dans la théorie standard de la régularisation $\mathcal{F}(\mathbf{x}) = \|\mathbf{D}_k \mathbf{x}\|^2$) [22].

Le critère (5) résume une vision de la régularisation que l'on peut qualifier de *déterministe*. Elle a donné lieu à des développements théoriques importants, essentiellement en physique mathématique. Cependant les questions du choix de la *fonctionnelle régularisante* $\mathcal{F}(\cdot)$ et du *coefficient de régularisation* α sont encore largement ouvertes, et la minimisation proprement dite du critère est souvent un problème en soi.

Cadre énergétique		Cadre probabiliste
Énergie $\mathcal{F}(\mathbf{x})$	\leftrightarrow	Densité de probabilité $p(\mathbf{x} T) = \frac{\exp\{-\mathcal{F}(\mathbf{x})/T\}}{\int_{\mathbf{x}} \exp\{-\mathcal{F}(\mathbf{x})/T\} d\mathbf{x}}$
Énergie quadratique $\mathcal{F}(\mathbf{x}) = \ \mathbf{x} - \mathbf{m}\ _{\mathbf{W}}^2$	\leftrightarrow	Loi gaussienne $\mathcal{N}(\mathbf{m}, T\mathbf{W}^{-1})$
Énergie séparable	\leftrightarrow	Bruit blanc
Énergie gibbsienne	\leftrightarrow	Modèle markovien
$\mathcal{J}(\mathbf{x}) = \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \alpha\mathcal{F}(\mathbf{x})$	\leftrightarrow	$p(\mathbf{x}, \mathbf{y} \boldsymbol{\theta}, I) = p(\mathbf{y} \mathbf{x}, \boldsymbol{\theta}, I)p(\mathbf{x}, \boldsymbol{\theta}, I)$ (règle de Bayes)
Solution pénalisée $\arg \min_{\mathbf{x}} \mathcal{J}(\mathbf{x})$	\equiv	Maximum <i>a posteriori</i> $\hat{\mathbf{x}}_{\text{MAP}} = \arg \max_{\mathbf{x}} p(\mathbf{x} \mathbf{y}, \boldsymbol{\theta}, I)$
?	\leftrightarrow	Marginalisation (de paramètres de nuisance)
?	\leftrightarrow	Régression (prédiction, filtrage de Wiener, ...)
?	\leftrightarrow	Algorithmes pseudo-aléatoires (recuit simulé, <i>Monte-Carlo Markov Chain</i> ,...)

TAB. 1 – La minimisation du critère composé (5) possède une interprétation probabiliste bayésienne. Ce même mécanisme permet de définir des distributions de probabilité à partir de fonctions d'énergie, et inversement. Mais le cadre probabiliste offre également des outils qui lui sont propres, n'ayant pas nécessairement d'interprétation déterministe : marginalisation, régression, tirage pseudo-aléatoire.

Beaucoup de fonctions d'énergie locale utilisées en traitement d'image ont été introduites dans un cadre bayésien : elles définissent \mathbf{x} comme un champ de Markov [9, 1, 3]. Mais bien que le point de vue *énergétique* soit aussi représenté dans la communauté des traiteurs d'image, tous les critères de la forme (5) sont formellement réinterprétables dans un cadre bayésien. La question n'est pas de savoir si l'interprétation bayésienne représente une justification des autres approches, mais plutôt de voir ce qu'elle apporte en réponse aux problèmes évoqués ci-dessus. En plus de sa grande cohérence, elle fournit des outils originaux pour le choix des hyperparamètres (*marginalisation*) et pour la minimisation du critère composé (*algorithmes pseudo-aléatoires*), et la variété des coûts qu'elle permet d'utiliser conduit à des solutions qui n'ont pas d'équivalent dans le cadre énergétique (*régression*).

7 Modèle *a priori*

Il est souvent reproché à l'estimation bayésienne de dépendre de la connaissance d'un hypothétique « vrai modèle » aléatoire ayant engendré l'objet à reconstruire. Implicitement, pour formuler ce reproche, il faut admettre que la réalité peut être « enfermée » dans un modèle mathématique. Vaste débat... Dans le cas de l'approche probabiliste de l'inversion telle que nous la concevons, l'interprétation fréquentiste des probabilités entretient une confusion fâcheuse : il nous semble important de préciser que nos hypothèses probabilistes ne sont pas des hypothèses sur le caractère « aléatoire » de l'objet, ou du bruit affectant les données, mais des choix d'un mode de représentation d'une information *a priori* incomplète — ou d'une connaissance incertaine — compatible avec l'outil d'inférence choisi. Cette situation n'a d'ailleurs rien de singulier puisqu'il est rare que dans un problème réel l'information *a priori* disponible le soit sous une forme directement adaptée au cadre théorique choisi pour le traitement. Cela dit, la conversion de cette information *a priori* en une distribution de probabilité est un problème difficile, encore largement ouvert. Il existe de nombreuses méthodes dont voici trois grandes familles.

Certaines méthodes reposent sur la théorie des *groupes de transformation* pour déterminer la mesure de référence « naturelle » pour le problème. Mais en pratique cette approche n'a guère permis que de justifier après coup l'emploi de la mesure de Lebesgue pour les paramètres de *localisation* (fournissant ainsi une extension au cas continu de la distribution uniforme résultant de l'application du « Principe d'indifférence » de Bernoulli dans le cas discret) et de la mesure de Jeffreys dans le cas des paramètres d'*échelle*.

D'autres méthodes reposent sur des principes informationnels. Il s'agit principalement des méthodes dites « à maximum d'entropie » dans lesquelles on recherche une distribution qui soit la plus proche d'une distribution de référence (au sens d'une distance de Kullback) tout en vérifiant une information incomplète connue *a priori*. Mais là encore, cette approche a surtout permis de justifier après coup certains choix. De plus, elle n'est véritablement praticable que lorsque cette information *a priori* est faite de contraintes linéaires sur la distribution recherchée (moments). On travaille alors dans la famille des *distributions exponentielles*.

Il existe enfin une dernière classe très importante, celle des constructions faites « à la main ». C'est dans cette catégorie qu'entrent les modèles markoviens qui ont connu un développement spectaculaire en imagerie depuis 1984, et qui permettent d'incorporer dans une distribution *a priori* des propriétés locales essentielles que doit posséder l'objet. La construction de ces modèles demande beaucoup de savoir-faire.

8 Choix du critère

L'approche bayésienne ramène l'inversion à la détermination d'une distribution *a posteriori*. Comme il n'est pas envisageable de calculer complètement de telles distributions, on se contente de rechercher un estimateur ponctuel qui est souvent celui du maximum *a posteriori*. D'autres alternatives existent (*maximum a posteriori marginal*, *moyenne a posteriori*, ...), mais il est important de bien évaluer les conséquences d'un tel choix et de proposer, si nécessaire, des alternatives.

La question peut être valablement soulevée de l'absolue nécessité d'imposer que la solution soit continue par rapport aux données et, par conséquence, de la convexité des critères de régularisation. Alors que les approches quadratiques et entropiques sont bien connues pour rendre les problèmes inverses bien posés, la minimisation d'une fonctionnelle non convexe ne peut assurer que la solution soit continue : une petite variation dans les données peut induire un « saut » d'une vallée à une autre, et donc une perte de continuité. Mais, dans beaucoup de problèmes, ces transitions sont non seulement désirables, mais nécessaires pour restaurer des discontinuités, des frontières, des interfaces, des points brillants, ... , sans limite en terme de résolution spatiale. Un éclairage différent peut être donné à ce problème en observant que certains critères non convexes introduits en imagerie possèdent une expression équivalente impliquant des *variables cachées*. Dans ce cas le problème quitte l'analyse convexe et incorpore une certaine dose d'analyse combinatoire, ou de test d'hypothèse, ce qui relève de la théorie de la décision plus que de l'estimation. L'analyse bayésienne reste pertinente dans ce contexte de détection-estimation combinées. Nombre de travaux récents ont été engagés dans cette voie, combinant plusieurs niveaux de variables, mélangeant descriptions à bas et haut niveau, ou des données recueillies par des modalités expérimentales différentes. C'est en ce sens que les concepts classiques de la régularisation, comme la continuité par rapport aux données, ne sont plus totalement appropriés et qu'un effort d'extension doit être fait.

9 Choix des hyperparamètres

Il existe assez peu de méthodes de détermination des hyperparamètres [21]. Dans le cas où ceux-ci se limitent au seul coefficient de régularisation, et où le régulariseur $\mathcal{F}(\mathbf{x})$ est quadratique, les méthodes de *validation croisée* fournissent des solutions acceptables [11]. Mais il s'agit de méthodes déterministes par essence, qui minimisent un critère de risque dépendant de l'objet inconnu \mathbf{x} , ce qui ne peut être effectué qu'asymptotiquement puisque cet objet est évidemment inconnu. C'est aussi dans cette catégorie qu'entre une méthode plus récente, dite *de la courbe en L* [12].

Les méthodes bayésiennes, qui attribuent une distribution de probabilité *a priori* à l'objet, ne présentent pas cette limitation. Les hyperparamètres θ constituent un second niveau de description du problème, indispensable pour « rigidifier » le premier niveau constitué par les paramètres eux-mêmes — c'est-à-dire l'objet \mathbf{x} . Dans un problème inverse, la valeur des hyperparamètres est importante pour obtenir une solution acceptable, mais ne présente pas d'intérêt en soi. Dans une approche bayésienne, on peut donc distinguer deux niveaux d'inférence. Le premier infère sur \mathbf{x} , pour une valeur donnée de θ , au travers de la distribution *a posteriori* de l'équation (4). Le second infère sur θ grâce à une relation analogue :

$$p(\theta | \mathbf{y}, \mathbf{A}) = \frac{p(\theta | \mathbf{A}) p(\mathbf{y} | \theta, \mathbf{A})}{p(\mathbf{y} | \mathbf{A})}.$$

On retrouve d'ailleurs là une caractéristique de l'utilisation de la règle de Bayes : la vraisemblance $p(\mathbf{y} | \theta, \mathbf{A})$ attachée aux données dans le second niveau est le coefficient de normalisation dans le premier.

Si, comme cela est souvent le cas, ce terme est suffisamment « piqué », l'influence de la

distribution *a priori* $p(\theta | \mathbf{A})$ est négligeable, et le second niveau d'inférence peut être résolu par maximisation de cette vraisemblance. Mais il faut pour cela résoudre un problème de marginalisation :

$$p(\mathbf{y} | \theta, \mathbf{A}) = \int p(\mathbf{x}, \mathbf{y} | \theta, \mathbf{A}) d\mathbf{x} = \int p(\mathbf{y} | \mathbf{x}, \theta, \mathbf{A}) p(\mathbf{x} | \theta) d\mathbf{x}. \quad (6)$$

Une telle intégrale conduit très rarement à un résultat explicite.

Pour contourner cette difficulté, on peut introduire des « variables cachées » \mathbf{z} qui viennent compléter les observations \mathbf{y} de manière que la nouvelle vraisemblance $p(\mathbf{y}, \mathbf{z} | \theta, \mathbf{A})$ soit plus simple à calculer. On est alors conduit à maximiser des espérances conditionnelles par des techniques itératives, déterministes ou stochastiques (algorithmes EM et SEM) [7].

On peut aussi remarquer que la *vraisemblance généralisée*

$$p(\mathbf{y}, \mathbf{x} | \theta, \mathbf{A}) = p(\mathbf{x} | \mathbf{y}, \theta, \mathbf{A}) p(\mathbf{y} | \theta, \mathbf{A}) = p(\mathbf{y} | \mathbf{x}, \theta, \mathbf{A}) p(\mathbf{x} | \theta)$$

résume toute l'information propre au premier niveau d'inférence, et vouloir en faire la maximisation conjointe par rapport à \mathbf{x} et θ . Le problème d'intégration soulevé par (6) est évidemment évacué. À θ fixé, le maximum de vraisemblance généralisée (MVG) coïncide avec le MAP. En revanche, il existe de nombreux cas de dégénérescence du MVG, ce qui rend cet outil plus délicat à utiliser que sa facilité de mise en œuvre ne le laisse paraître. Par contre, à \mathbf{x} fixé, la situation est beaucoup moins favorable : la vraisemblance généralisée n'est pas, en général, majorée sur son domaine de définition, et il n'existe même pas toujours un maximum local.

La mise au point de méthodes exemptes de tout réglage empirique représente à la fois un enjeu pratique évident, mais aussi l'occasion d'asseoir la crédibilité de méthodes utilisant des modèles qui n'ont pas besoin d'être « vrais » pour être pertinents. Le support probabiliste est un outil formel très adapté car il permet l'écriture de vraisemblances marginales. Le principal obstacle reste d'ordre calculatoire. L'enjeu consiste à savoir contourner cet obstacle, par exemple par échantillonnage stochastique, dans des situations de plus en plus complexes.

10 Problèmes algorithmiques

L'approche bayésienne offre un cadre remarquablement cohérent pour traiter de problèmes d'inférence dans une situation incertaine où l'on dispose de plusieurs sources d'information : données expérimentales, modèle direct, *a priori*. Les limitations en sont bien connues, elle sont essentiellement d'ordre calculatoire. Il est clair que l'on ne peut pas espérer obtenir par cette approche de résultats satisfaisants dans des problèmes complexes, avec des modèles simplistes et des méthodes d'optimisation approximatives. C'est ce qui explique sans doute en grande partie l'échec des tentatives faites il y a une vingtaine d'années en restauration d'image et en filtrage de bruit, avec des modèles gaussiens. Mais on a assisté, dans le domaine de la vision par ordinateur, et depuis l'article fondateur des frères Geman en 1984, à une renaissance spectaculaire des méthodes de cette famille, associant champs markoviens, échantillonnage de Gibbs, techniques de recuit, parallélisation des calculs, dans des problèmes de segmentation, de détection de contours, d'extraction de textures, etc... Ces travaux ont fourni une base rationnelle très solide à ce domaine. Les critères étant choisis, se pose encore la question de leur extrémalisation, d'autant qu'ils sont souvent multimodaux. Lorsque la résolution du problème conduit à une *optimisation non convexe*, le *recuit simulé* est la méthode la plus souvent mise en avant, compte tenu de ses propriétés théoriques. Mais cette approche stochastique se révèle encore très souvent trop coûteuse numériquement pour être réaliste, et d'autres approches, bien connues pour ne pas posséder le même potentiel théorique, demeurent tout-à-fait concurrentielles. Citons par exemple les procédés d'optimisation par *non convexité graduelle* qui sont des méthodes sous-optimales de relaxation

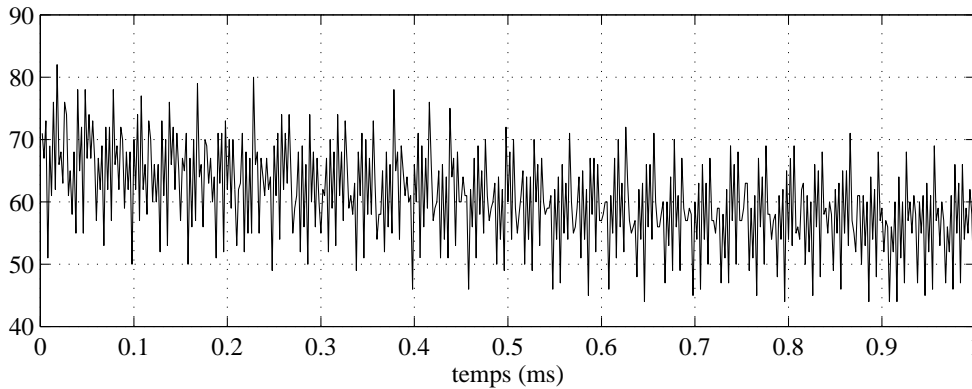


FIG. 1 – *Interférogramme obtenu pour une lampe à vapeur de cadmium (500 échantillons temporels prélevés à 500 kHz).*

déterministe initialement introduites en segmentation d'image. Le domaine de la restauration 1-D, dont les développements passés ont souvent préfiguré ceux de la restauration 2-D, l'illustre parfaitement : on connaît de nombreux exemples récents qui montrent qu'une mise en œuvre bien faite — c'est-à-dire qui exploite la structure intime du problème — d'une méthode théoriquement imparfaite conduit à des résultats de qualité comparable à celle qu'offre une technique de recuit, à un coût calculatoire bien moindre. Il faut bien voir qu'à ce niveau il n'existe pas de réponse générale à un problème d'optimisation. Pour être crédible, il faut accepter de quitter le domaine des mathématiques appliquées pour celui des sciences pour l'ingénieur.

11 Un exemple : la spectrométrie

La spectrométrie par transformation de Fourier est apparue il y a une cinquantaine d'années, et le premier appareil ayant été employé est l'interféromètre de Michelson. Mais cet instrument nécessite des réglages très précis et des miroirs de grande qualité optique, ce qui rend son utilisation trop coûteuse dans des domaines tels que l'analyse des gaz polluants. D'autres principes, reposant sur les propriétés des matériaux optiques anisotropes, ont permis de développer des interféromètres biréfringents plus compacts et ne nécessitant qu'un minimum de réglages, mais au prix d'une dégradation du pouvoir de résolution [17]². La Fig. 1 donne un exemple d'interférogramme obtenu avec ce genre d'appareil observant une lampe à vapeur de cadmium dont le spectre d'émission est connu, et la Fig. 2 donne le « spectre de Fourier » correspondant — qui est en fait le périodogramme du signal de la Fig. 1 — calculé dans la bande de longueurs d'ondes accessible par l'instrument. Le manque de résolution se traduit par l'impossibilité de séparer deux des quatre raies du cadmium, celles dont les longueurs d'onde ne diffèrent que de 12,2 nm.

Le problème à résoudre est donc d'améliorer la résolution des spectres obtenus par analyse de Fourier classique avec ces interféromètres. Il existe bien sûr plusieurs manières d'aborder un tel problème. Le manque de résolution venant essentiellement de la longueur limitée de l'interférogramme, nous avons choisi, pour illustrer ce chapitre, de modéliser les données recueillies de la manière suivante :

$$\mathbf{y} = \mathbf{F} \mathbf{X} + \mathbf{b}. \quad (7)$$

2. Les auteurs tiennent à remercier C. Durieu et S. Prunet, du LESIR de l'ENS de Cachan, qui leur ont aimablement fourni les données expérimentales utilisées dans ce chapitre.

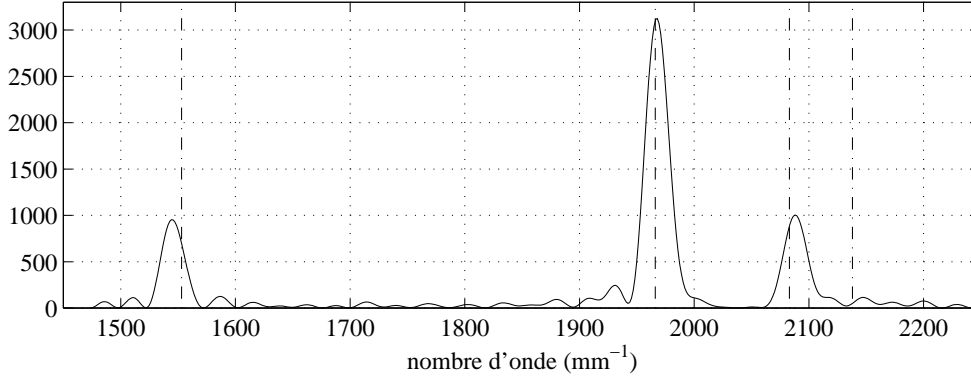


FIG. 2 – Périodogramme déduit des données de la Figure 1. Un « zoom spectral » a été effectué sur la zone fréquentielle d'intérêt, c'est-à-dire celle qui contient les quatre raies du cadmium, repérées par des droites verticales tiretées. Le manque de résolution ne permet pas de discriminer la quatrième raie.

\mathbf{y} est un vecteur réel constitué des N échantillons de l'interférogramme, \mathbf{F} est une matrice de Fourier rectangulaire de dimensions $(N \times M)$, \mathbf{X} est la solution recherchée, vecteur complexe de dimension $M \gg N$, et dont le carré du module constituera le spectre recherché, et \mathbf{b} désigne les erreurs expérimentales, ou « bruit ». Le problème est donc vu comme une *synthèse de Fourier*, avec une discrétisation dans le domaine spectral beaucoup plus fine que celle qui est naturellement imposée par l'échantillonnage des données. On pourrait aussi considérer que le problème est vu comme une *extrapolation* de signal puisque, une solution $\widehat{\mathbf{X}}$ ayant été trouvée, il lui correspond, dans le domaine spatial de l'interférogramme, une solution $\widehat{\mathbf{x}} = \text{TFD}^{-1}(\widehat{\mathbf{X}})$, de même longueur $M \gg N$.

Posé en ces termes, la solution à l'équation (7) est évidemment indéterminé. Pour lever cette indétermination, on pourrait songer à une solution *inverse généralisée*, mais il est facile de vérifier qu'elle nous conduit à nouveau au spectre du périodogramme, l'extrapolation étant obtenue par *bourrage de zéros* [13]. Il faut donc aussi, après avoir modélisé les données, choisir une information *a priori* permettant d'aboutir à une solution mieux résolue. Le choix pour le bruit d'une distribution gaussienne centrée et non corrélée permet de définir une vraisemblance :

$$p(\mathbf{y} | \mathbf{X}, \sigma, I) \propto \exp \left\{ - \|\mathbf{y} - \mathbf{F} \mathbf{X}\|^2 / (2 \sigma^2) \right\},$$

dont on remarque que la maximisation par rapport à \mathbf{X} n'aboutit à rien puisque \mathbf{F} est rectangulaire, ainsi que nous l'avions déjà signalé dans le paragraphe 4. Il nous faut donc, pour que le problème inverse soit bien posé, choisir une distribution de probabilité *a priori* pour \mathbf{X} , ou, ce qui est équivalent avec une interprétation énergétique, la fonctionnelle de régularisation $\mathcal{F}(\mathbf{X})$ d'un critère régularisé de la forme :

$$\mathcal{J}(\mathbf{X}) = \|\mathbf{y} - \mathbf{F} \mathbf{X}\|^2 + \alpha \mathcal{F}(\mathbf{X}).$$

Le choix le plus simple et le plus commode à mettre en œuvre est la distribution gaussienne, circulaire, et non corrélée, qui conduit à $\mathcal{F}(\mathbf{X}) = \|\mathbf{X}\|^2$, mais dont il est facile de montrer qu'elle conduit encore, à un coefficient multiplicatif près, au périodogramme usuel [13]. La fonctionnelle $\|\mathbf{X}\|^2$ pénalise trop les composantes spectrales de forte amplitude, il faut en choisir une qui croisse à l'infini moins vite qu'une norme ℓ_2 .

Parmi les distributions à queue plus lourde que la gaussienne, nous retenons la distribution de Laplace séparable: $p(\mathbf{X} | \lambda, I) \propto \exp(-\lambda \sum_{m=1}^M \|X_m\|)$, ce qui conduit à une

distribution *a posteriori* de la forme : $p(\mathbf{X} | \mathbf{y}, \sigma, \lambda, I) \propto \exp[-\mathcal{J}(\mathbf{X})/\sigma^2]$, avec :

$$\mathcal{J}(\mathbf{X}) = \|\mathbf{y} - \mathbf{F} \mathbf{X}\|^2 + \lambda \sigma^2 \sum_{m=1}^M \|X_m\|. \quad (8)$$

Le critère régularisé (8) n'est pas une fonction strictement convexe de \mathbf{X} et, de plus, il n'est pas différentiable à l'origine, ce qui limite l'emploi d'algorithmes classiques de descente, tels que le gradient, dans la recherche du mode de la distribution. Un moyen de contourner cette difficulté consiste à introduire des variables auxiliaires réelles d_m , $m = 1, \dots, M$, et à considérer le critère *augmenté* :

$$\mathcal{K}(\mathbf{X}, \mathbf{d}) = \sigma^{-2} \|\mathbf{y} - \mathbf{F} \mathbf{X}\|^2 + \lambda \left\{ \sum_{m=1}^M [\|X_m\|^2 / (4d_m) + d_m] \right\}. \quad (9)$$

Celui-ci est cette fois convexe et différentiable en \mathbf{X} et \mathbf{d} . On montre de plus que le minimum du critère $\mathcal{K}(\mathbf{X}, \mathbf{d})$ est égal à celui de $\mathcal{J}(\mathbf{X})$, et qu'il est atteint par le même $\widehat{\mathbf{X}}$ [10]. Le calcul de la solution peut ainsi se faire de manière efficace avec une méthode de relaxation par bloc.

La solution ainsi obtenue est conditionnée par les valeurs des hyperparamètres σ et λ qu'il faut soit déterminer, soit intégrer hors du problème. C'est à ce stade que l'approche bayésienne se révèle particulièrement utile en offrant des outils sans équivalent dans l'approche énergétique. La technique que nous avons retenue consiste : 1° à remarquer que (9) définit une loi de probabilité $p(\mathbf{X}, \mathbf{d}, \sigma^2, \lambda | \mathbf{y})$; 2° à intégrer (\mathbf{X}, \mathbf{d}) hors du problème par échantillonnage pseudo-aléatoire afin d'en déduire des estimateurs de σ^2 et λ maximisant la vraisemblance $p(\lambda, \sigma^2 | \mathbf{y})$; 3° à calculer la solution $\widehat{\mathbf{X}}_{\text{MAP}}$ pour ces valeurs des hyperparamètres. L'étape d'échantillonnage concerne les 4 quantités inconnues $(\mathbf{X}, \mathbf{d}, \sigma^2, \lambda)$. Alternativement, trois de ces quantités sont fixées pour rééchantillonner la quatrième :

- $p(\mathbf{X} | \mathbf{y}, \mathbf{d}, \lambda, \sigma^2)$ est une loi normale, donc facile à échantillonner.
- $p(\mathbf{d} | \mathbf{y}, \mathbf{X}, \lambda, \sigma^2)$ est une loi non classique, ne dépendant que de \mathbf{X} et λ . On ne sait pas l'échantillonner directement, mais on peut procéder par réjection [18], à l'aide d'une classe de lois dominantes du type gamma.
- $p(\lambda | \mathbf{y}, \mathbf{X}, \mathbf{d}, \sigma^2)$ et $p(\sigma^2 | \mathbf{y}, \mathbf{X}, \mathbf{d}, \lambda)$ sont respectivement des lois gamma et gamma inverse, dont l'échantillonnage est classique.

La suite des quadruplets ainsi constituée est une chaîne de Markov dont la loi tend vers $p(\mathbf{X}, \mathbf{d}, \lambda, \sigma^2 | \mathbf{y})$ [18]. La Fig. 3 montre le résultat de l'échantillonnage pseudo-aléatoire de la loi $p(\mathbf{X}, \mathbf{d}, \lambda, \sigma^2 | \mathbf{y})$, permettant d'estimer λ et σ^2 . Le spectre estimé ensuite par $|\widehat{\mathbf{X}}_{\text{MAP}}|^2$ est indiqué à la Fig. 4. Les 4 raies du cadmium sont maintenant bien séparées. Il est vraisemblable que les erreurs de localisation sont imputables à une erreur de calibration.

12 Conclusion

Ce chapitre a permis de présenter brièvement les caractéristiques essentielles de l'approche bayésienne de l'inversion. Nous nous sommes limités à l'estimation paramétrique et n'avons qu'évoqué le choix d'un modèle. Nous insistons à nouveau sur le fait que la plupart des méthodes d'inversion classiques peuvent être chacune établie, ou réinterprétée, dans plusieurs cadres théoriques, et qu'il n'y a pas de lien exclusif entre les méthodes (plus précisément les algorithmes de traitement des données) et leurs interprétations théoriques. Ces dernières, par contre, peuvent être rangées en fonction de leur degré de généralité, c'est-à-dire leur capacité à affronter tous les problèmes soulevés lors de la résolution effective d'un problème inverse. De ce point de vue, l'approche bayésienne est particulièrement cohérente et complète puisque, outre le choix d'un critère régularisé, elle permet aussi le développement

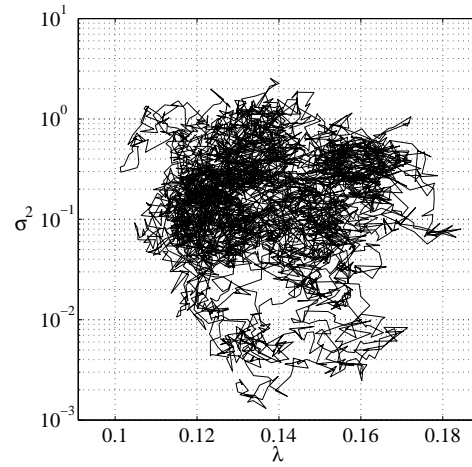


FIG. 3 – Trace dans le plan des hyperparamètres (λ, σ^2) de la trajectoire de l'échantillonneur utilisé. Environ 5 000 itérations ont été effectuées ; en pratique, beaucoup moins suffisent pour évaluer que les valeurs les plus vraisemblables de (λ, σ^2) se répartissent autour de $(0,14; 0,1)$.

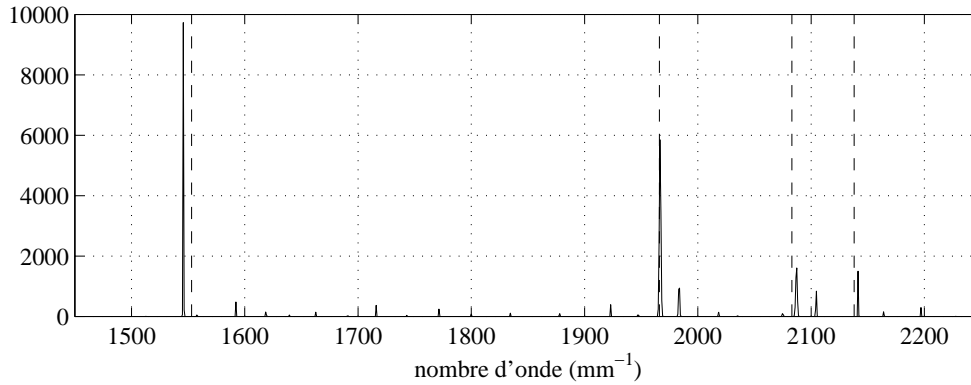


FIG. 4 – Spectre estimé par $|\widehat{\mathbf{X}}_{MAP}|^2$, où $\widehat{\mathbf{X}}_{MAP}$ est le minimum du critère (8), pour les valeurs d'hyperparamètres $(\lambda = 0,14; \sigma^2 = 0,1)$. Le vecteur $\widehat{\mathbf{X}}_{MAP}$ estimé contient 1024 points dans la zone fréquentielle d'intérêt, et aucun en dehors.

de techniques de calcul automatique des hyperparamètres et d'algorithmes d'optimisation pseudo-aléatoires, sans équivalents dans un cadre non probabiliste.

RÉFÉRENCES

- [1] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. ser. B*, vol. 48, pp. 259-302, 1986.
- [2] A. Blake and A. Zisserman, *Visual Reconstruction*, Cambridge, MA: MIT Press, 1987.
- [3] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Processing*, vol. 2, pp. 296-310, Jul. 1993.
- [4] F. Murtagh, J.-L. Starck and A. Bijaoui, "Image restoration with noise suppression using a multiresolution support," *Astron. Astrophys. Suppl. Ser.*, vol. 112, pp. 179-189, 1995.
- [5] R.T. Cox, *The Algebra of Probable Inference*, Johns Hopkins University Press, Baltimore, 1961.
- [6] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, pp. 2024-2073, Dec. 1989.
- [7] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.*, ser. B, vol. 39, pp. 1-38, 1977.
- [8] J.N. Franklin, "Well-Posed Stochastic Extensions of Ill-Posed Linear Problems," *J. Math. Anal. Applic.*, vol. 31, pp. 682-716, 1970.
- [9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721-741, Nov. 1984.
- [10] D. Geman and G. Reynolds, "Constrained Restoration and Recovery of Discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-14, pp. 367-383, Mar. 1992.
- [11] G. Golub, M. Heath and G. Wahba, "Generalized Cross-Validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, pp. 215-223, 1979.
- [12] P.C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Review*, vol. 34, pp. 561-580, 1992.
- [13] J. Idier, J.-F. Giovannelli and Ph. Ciuciu, "Interprétation régularisée des périodogrammes et extensions non quadratiques," in *16^e Colloque GRETSI*, Grenoble, sept. 1997, pp. 695-698.
- [14] E.T. Jaynes, "On the Rationale of Maximum-Entropy Methods," *Proc. IEEE*, vol. 70, pp. 839-952, 1982.
- [15] J. Marroquin, S. Mitter and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Stat. Assoc.*, vol. 82, pp. 76-89, 1987.
- [16] D.L. Phillips, "A Technique for the Numerical Solution of Certain Integral Equations of the First Kind," *J. Assoc. Comput. Mach.*, vol. 9, pp. 84-97, 1962.
- [17] S. Prunet, C. Durieu and B. Journet, "Improvement of the resolution of Fourier spectrometers via modern parametric spectral methods," in *IMACS/IEEE Multiconference, CESA '98*, Nabeul-Hammamet, Tunisia, Apr. 1998, pp. 376-381.
- [18] Chr. Robert, *Simulations par la méthode MCMC*, Economica, Paris, 1997.
- [19] A. Tarantola, *Inverse problem theory: Methods for data fitting and model parameter estimation*, Amsterdam: Elsevier Science Publishers, 1987.
- [20] A. Tikhonov et V. Arsénine, *Méthodes de résolution de problèmes mal-posés*, Éditions Mir, Moscou, 1976.
- [21] A.M. Thomson, J.C. Brown, J.W. Kay and D.M. Titterington, "A study of methods of choosing the smoothing parameter in image restoration by regularization," *Trans. IEEE Pattern Anal. Mach. Intell.*, vol. PAMI-13, pp. 326-339, 1991.

- [22] D.M. Titterington, "Common structures of smoothing techniques in statistics," *Int. Statist. Rev.*, vol. 53, pp. 141-170, 1985.
- [23] S. Twomey, "On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature," *J. Assoc. Comp. Mach.*, vol. 10, pp. 97-101, 1962.

