



APPLICATION DU RECUIT SIMULE A L'INVERSION DE LA
TRANSFORMATION DE POISSON-MANDEL

F. SULTANI, H. LANTERI, C. AIME, G. RICORT, E. ARISTIDI

Département d'Astrophysique
U.R.A. 709 du CNRS
Université de Nice — Sophia Antipolis
Parc Valrose — 06108 Nice Cedex 2
FRANCE

RESUME

Dans cette communication, nous présentons une application de la méthode du recuit simulé à l'inversion de la transformation de Poisson-Mandel. Ce problème inverse étant mal posé, les techniques de régularisation par contraintes de douceur sont mises en œuvre.

ABSTRACT

We present in this paper the application of simulated annealing technique to the inversion of the Poisson Transform. This ill-posed problem is regularized by means of smoothness constraints.

INTRODUCTION

La transformation de Poisson-Mandel (TPM) intervient en optique pour les faibles flux lumineux où l'observation s'effectue en comptage de photons. La TPM est la transformation qui, appliquée à la densité de probabilité (DDP)

d'une intensité lumineuse aléatoire (fort flux), donne la répartition du nombre de photons détectés. Si l'on note $X(I)$ la DDP à fort flux, la répartition $Z(n)$ du nombre de photons détectés par pixel s'obtient par la relation :

$$Z(n) = \int_0^{\infty} e^{-I} \frac{I^n}{n!} X(I) dI \quad (1)$$

qui décrit la TPM [1]. Dans notre application, on cherche à obtenir $X(I)$ à partir d'une mesure bruitée $Y(n)$ de $Z(n)$.

$$Y(n) = Z(n) + b(n) \quad (2)$$

où $b(n)$ est une terme de bruit.

Dans le travail que nous présentons, nous nous sommes proposés de résoudre numériquement l'équation (1) pour un cas particulier où $X(I)$ est la DDP associée à l'addition de deux structures de speckles non corrélées. Il s'agit d'un cas classique [2].

La résolution de l'équation (1) est un problème inverse mal posé car il ne satisfait pas les conditions d'existence, d'unicité et de stabilité de la solution vis à vis des incertitudes de mesures dues au bruit [3]. La résolution numérique de cette intégrale passe par une étape de discrétisation qui conduit à :

$$Y = AX + b \quad (3)$$

où Y désigne le vecteur des données, A la matrice de transformation, X le vecteur des paramètres à estimer et b le vecteur contenant les bruits de mesures dus aux erreurs d'estimation (bruit statistique). L'équation (3) a une solution de norme minimale qui est unique $X^+ = A^+ Y$ (où A^+ désigne l'inverse généralisée) qui est obtenue par minimisation de :

$$J_Q(X) = \|Y - AX\|^2 \quad (4)$$

où $\|\cdot\|$ désigne la norme Euclidienne.

REGULARISATION

Le mauvais conditionnement de la matrice A ne permet pas d'obtenir une solution stable du problème. Un principe général pour s'affranchir des problèmes d'instabilités est celui de la régularisation. Régulariser un problème c'est admettre qu'on ne peut pas obtenir la vraie solution à partir des données entachées d'erreurs, et que le système d'équations initial ne permet que de définir un ensemble de solutions possibles parmi lesquelles une solution acceptable doit être choisie. Ainsi on est conduit à introduire, dans la formulation du problème, des informations sur les propriétés *a priori* de la solution. Ces informations se présentent sous forme d'une (ou plusieurs) contrainte(s). En ce qui concerne notre problème une contrainte de douceur est physiquement raisonnable.

On est donc amené à minimiser une fonctionnelle avec con-



trainte qui peut s'écrire de façon générale [4,5]:

$$J(X) = \Delta_1(X, X_0) + \lambda \Delta_2(X, X_\infty) \quad (5)$$

où $\Delta_1(X, X_0)$ est la distance quadratique entre X et X_0 , (solution consistante avec les mesures), et où $\Delta_2(X, X_\infty)$ est une distance (à déterminer) qui a pour rôle de régulariser la solution X en la contraignant à tendre vers la solution *a priori* (infiniment lisse) par le biais de λ . Le coefficient de régularisation λ traduit l'importance que l'on donne à la contrainte par rapport à la distance $\Delta_1(X, X_0)$ qui elle assure le lien entre les données et le paramètres à estimer. Deux types de régularisations sont utilisés :

A-La première et la plus classique fait partie de la classe des régularisation au sens de Tikhonov [6]. Dans ce cas l'information *a priori* de douceur est introduite sous forme d'un opérateur différentiel. Ainsi on cherche une solution telle que la norme de sa dérivée soit inférieure à une limite fixée. La fonctionnelle à minimiser s'écrit sous la forme suivante :

$$J(X) = \|Y - AX\|^2 + \lambda \|CX\|^2 \quad (6)$$

où C représente l'opérateur de différentiation.

B-Le principe de régularisation peut être interprété dans un cadre probabiliste. Dans ce contexte la connaissance *a priori* de l'objet est traduite par une loi de probabilité ; le principe de maximum d'entropie permet de déterminer cette loi [7].

Si cette connaissance *a priori* se limite à une contrainte globale sur l'entropie de l'objet, la loi de probabilité de l'objet s'écrit comme :

$$P(X) = \exp(\lambda S(X)) \quad (7)$$

avec $S(X)$ donné par la relation ci-dessous :

$$S(X) = - \sum_i X_i \log X_i \quad (8)$$

où les X_i sont les composantes de X . Si on fait l'hypothèse que $b(n)$ est un bruit blanc, gaussien de moyenne nulle et s'il y a indépendance statistique entre les $b(n)$, alors la solution X est obtenue en minimisant la fonctionnelle ci-dessous :

$$J(X) = \|Y - AX\|^2 + \lambda \sum_i X_i \log X_i \quad (9)$$

RECUIT SIMULE

Le recuit simulé [8] est une méthode de minimisation d'une fonctionnelle qui a la propriété de converger vers le

minimum global. Dans le cas d'une fonctionnelle quadratique, les méthodes algébriques classiques donnent de très bons résultats et l'utilisation du recuit simulé n'est pas indispensable. La méthode est plus intéressante pour sa simplicité de mise en œuvre dans le cas d'une régularisation par maximum d'entropie, et devient quasiment incontournable lorsque des fonctionnelles non convexes sont à minimiser.

A l'origine cette technique s'inspire de la physique statistique. En effet, en thermodynamique statistique à une température T et à l'équilibre thermique, la probabilité pour qu'une particule se trouve à un état d'énergie E est donnée par la distribution de Boltzmann :

$$P(E) = \frac{e^{-\frac{E}{kT}}}{\sum e^{-\frac{E}{kT}}} \quad (10)$$

Lorsque la température tend vers zéro, la seule probabilité non nulle pour que la particule se trouve à un état d'énergie E correspond à l'état d'énergie minimale du système. Si on considère un système de particules à haute température, il lui correspond un état de désordre total. Si on refroidit subitement celui-ci, le solide que l'on obtiendra ne sera pas dans son état d'ordre maximal. En effet, en refroidissant brutalement, on fige les particules dans leurs positions désordonnées et aléatoires (ceci est un état métastable). Par contre si on abaisse la température de façon lente et si pour chaque température on laisse s'établir l'équilibre thermique, lorsque l'on approche la température nulle le solide que l'on obtiendra sera dans son état d'énergie minimale, c'est à dire dans un état d'ordre maximal. Le recuit simulé est basé sur le même principe, la fonctionnelle à minimiser joue le rôle de l'énergie (coût).

A une température initiale assez élevée, pour qu'aucune solution (configuration) ne soit privilégiée, on part d'une configuration initiale (solution uniforme dans notre cas). On associe à cette configuration g_i une énergie E_i . A cette même température on choisit aléatoirement un pixel et on lui ajoute (ou retranche) de façon aléatoire une quantité Δ (le grain). L'énergie associée à cette configuration est E_{i+1} . Deux cas peuvent se présenter:

i - La différence $\Delta E = E_{i+1} - E_i$ est négative, cela correspond bien à une décroissance du coût. On accepte donc cette nouvelle configuration.

ii - La différence ΔE est positive ce qui correspond à une augmentation du coût. L'algorithme de recuit simulé permet d'accepter des configurations correspondant à une augmentation de coût si $\exp(-\Delta E/T)$ est plus grand que qu'un nombre aléatoire entre zéro et un (critère de Metropolis [9]).

On répète cette opération de perturbation des pixels de l'image jusqu'à ce que le coût atteigne une valeur stationnaire, ce qui correspond à l'équilibre thermique à tempéra-

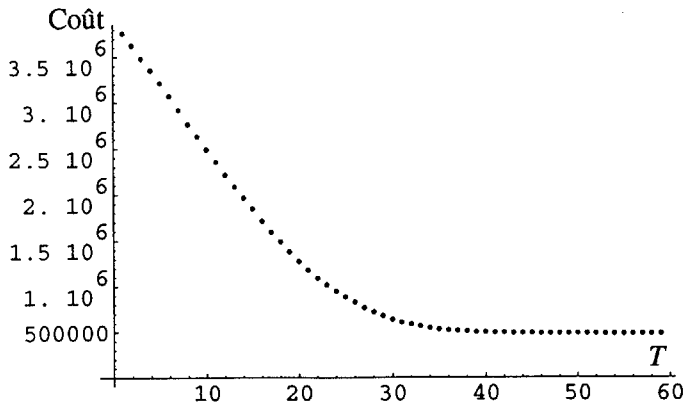


Fig 1: Variation du coût en fonction de la température. On constate que le coût converge de façon asymptotique au delà du 50 ème pas de température.

ture constante. Lorsque cette convergence est atteinte, on abaisse d'une petite quantité la température et on recommence le même procédé. On arrête l'algorithme dès que le coût atteint sa convergence asymptotique (système physique à température nulle). La mise en oeuvre de l'algorithme va comporter les étapes suivantes :

- a- Choix de la température initiale.
- b- Choix de la taille de la perturbation (grain).
- c- Estimation de l'équilibre approché à une T donnée.
- d- Loi de l'abaissement de la température.
- e- Critère d'arrêt.

a- La température dans l'algorithme de recuit simulé joue un rôle très important. C'est elle qui fixe la probabilité d'accepter ou de refuser les configurations défavorables à la minimisation. Cela conduit à accepter des configurations correspondant à une augmentation de coût et permet au système de sortir des minima locaux. Le choix de la température initiale se fait de façon empirique [10]. Il est basé sur le principe de taux d'acceptation des configurations :

Si ξ est le nombre de configurations augmentant le coût et acceptées et ϵ le nombre de configurations augmentant le coût, le rapport ξ/ϵ définit le taux d'acceptation. Nous avons opté pour une température initiale T_0 , déterminée par essais successifs, correspondant à un taux d'acceptation de 80%.

b-La taille de la perturbation (grain) Δ est un sujet traité assez évasivement dans la littérature concernant le recuit simulé. Compte tenu du fait que la solution a les dimensions d'une DDP et avec le choix du pas d'intégration utilisé, nous avons travaillé avec des grains de l'ordre de 10^{-3} ou 10^{-4} . Il est évident qu'un grain très petit augmente considérablement le temps de calcul, alors que

inversement une grande perturbation conduit à une solution non optimale [11].

c-La convergence du coût à température constante est théoriquement asymptotique. Pour atteindre cette convergence approchée il faut itérer un grand nombre de fois pour que l'équilibre s'établisse. En pratique ce n'est jamais réalisable. Nous nous sommes basés sur deux méthodes différentes:

i - Une méthode empirique est proposée par Kirkpatrick et Gelatt [10]. Elle consiste à fixer un nombre d'itérations permettant d'approcher l'équilibre, tout en gardant un compromis entre le temps de calcul et la qualité du résultat.

ii - Une méthode basée sur la convergence de la valeur moyenne relative du coût [11] en fonction du nombre d'itérations. Dans ce cas, une moyenne mobile des valeurs du coût est calculée à chaque itération. Lorsque l'écart relatif entre deux valeurs successives de cette moyenne est inférieure à un seuil fixé, on considère que la convergence à la température considérée est atteinte.

d-La loi d'abaissement de la température joue un rôle important dans l'algorithme. Plusieurs méthodes sont proposées dans la littérature. Nous avons utilisé la loi dite géométrique :

$$T_n = \alpha^n T_0 \quad (11)$$

Avec $0.8 < \alpha < 1$; le choix de α est d'importance, il détermine la façon dont on refroidit le système. Comme nous l'avons déjà souligné précédemment, le rôle de la température est prépondérant en ce qui concerne le taux d'acceptation. Une décroissance lente de la température implique une baisse régulière du nombre des configurations augmentant le coût. Par contre un abaissement trop rapide de la température réduirait considérablement le rôle du critère de Metropolis et peut poser des problèmes de piègeages dans un minimum local.

e-Le critère d'arrêt de l'algorithme est basé sur la variation relative du coût en fonction de T lorsque l'équilibre est atteint. Soit m_1 la valeur moyenne du coût sur N paliers de températures, m_2 étant la moyenne sur N autres paliers suivants. On considère que la convergence est atteinte quand $(m_2 - m_1)/m_2$ est inférieur à un seuil fixé [11].

RESULTATS

Pour le problème qui nous intéresse, la DDP à fort flux correspondant à la somme de deux structures de speckles $X(I)$ est donnée par la relation suivante [2] :

$$X(I) = \frac{1}{\mu(1-\beta)} \left[\exp\left(-\frac{I}{\mu}\right) - \exp\left(-\frac{I}{\beta\mu}\right) \right] \quad (12)$$

L'équation (1) permet de calculer analytiquement la répartition du nombre de photons $Y(n)$:

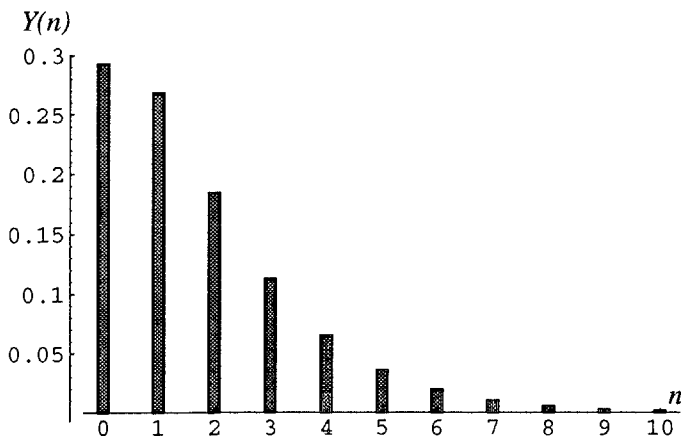


Fig 2: Distribution $Y(n)$ du nombre de photons détectés obtenue pour une simulation de 10 000 événements, pour un rapport d'intensité de 0.7 et une moyenne de 1.

$$Y(n) = \frac{\mu^n}{1-\beta} \left[\frac{1}{(1+\mu)^{n+1}} - \frac{\beta^{n+1}}{(1+\beta\mu)^{n+1}} \right] \quad (13)$$

où μ correspond au nombre moyen de photons par pixel et β désigne le rapport des intensités des deux structures. Les données $Y(n)$ sont simulées pour 10000 événements obéissant à la loi définie par la relation (13). Un exemple de ces données est montré sur la figure 2.

Les résultats de l'inversion sont donnés sur les figures 3 et 4 pour les deux types de régularisation envisagés. On peut constater le bon accord avec les résultats exacts fournis par l'équation 12. Les durées de calcul sont de l'ordre de deux minutes environ sur une station VAX 4000.

La simplicité de mise en œuvre quelle que soit la fonctionnelle à minimiser permet d'envisager l'application aux problèmes analogues concernant les DDP d'ordre 2 avec les contraintes supplémentaires liées à ce problème ; dans ce cas nous serons conduits à traiter des fonctionnelles présentant des minima locaux.

REFERENCES

- [1] : Mandel L, *Proc.Phys. Soc.* **74**, 233 (1953).
 [2] : Goodman J.W "Statistical Optics", J.Wiley & Son (1985).
 [3] : Nashed M. Z. "Operator Theoretical and Computational Approaches to Ill Posed Problem with Applications to Antenna Theory", *IEEE Trans ap.* **29**, 220-231 (1981).
 [4] : Demoment G. "Image Reconstruction and Restoration: Overview of Common Estimation Structures and Problems" *IEEE Trans on Acoustics Speech and Signal processing*, **37** n° 12, (1989).
 [5] : Titterington D.M. "General Structure of the Regula-

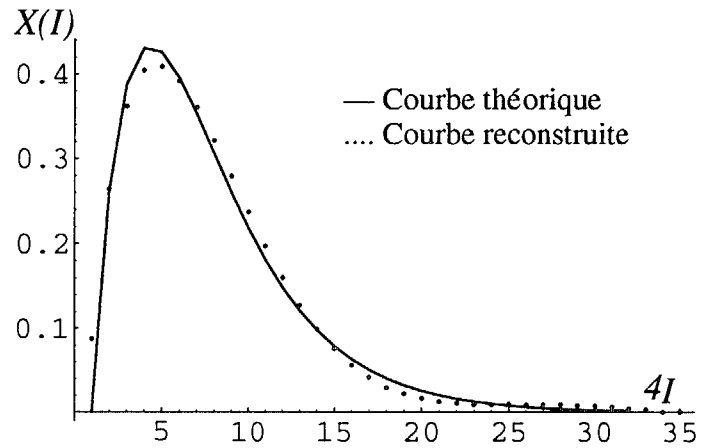


Fig 3 : En pointillé nous présentons la solution reconstruite $X(I)$ pour $\mu=1$ et $\beta=0.7$ régularisée par un opérateur différentiel (gradient) pour des données $Y(n)$ simulées sur 10000 événements. La solution théorique est illustrée en trait plein.

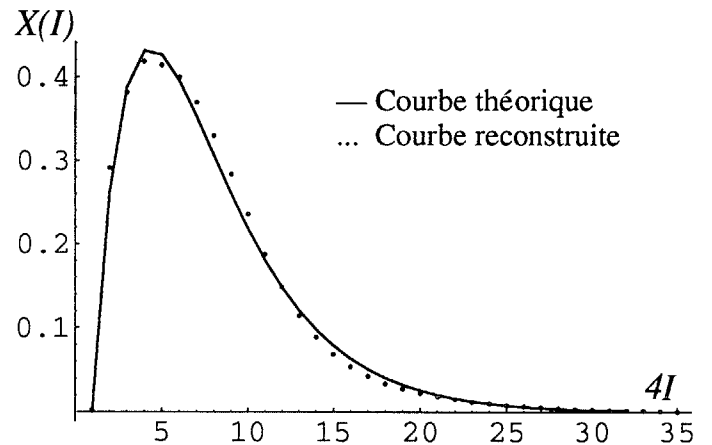


Fig 4 : En pointillé nous présentons la solution reconstruite $X(I)$ pour $\mu=1$ et $\beta=0.7$ régularisée par maximum d'entropie pour des données $Y(n)$ simulées sur 10000 événements. La solution théorique est illustrée en trait plein.

risation Procedure in Image Reconstruction", *Astron. Astroph.*, **144**, 381-387 (1985).

- [6] : Tikhonov A.V & Arsenin V.Y "Solutions of Ill Problems", V.H Winston and Son, Washington (1977).
 [7] : Mohammad-Djafari A.A. "Synthèse de Fourier Multivariables à Maximum d'Entropie", Thèse de Doctorat D'Etat., Univ Paris-sud (1987).
 [8] : Aarts A & Korst J "Simulated Annealing and Boltzmann Machine." Jones Wiley & Son (1990).
 [9] : Metropolis N., Rosenbluth H., Rosenbluth M., Teller A, Teller E, "Equation of state calculations by fast computing machines", *J. Chem. Phys.* **21**, 1087 (1953).
 [10] : Kirkpatrick .S & Gelatto.C.O "Optimisation by Simulated annealing", *Sciences*, **220**, n°4598 (1983).
 [11] : Sultani.F, "Inversion de la Transformation de Poisson par Recuit simulé", Rapport de DEA. Université de Nice (1992).