Astérisque

DIDIER DACUNHA-CASTELLE

Reconstruction des phases en cristallographie par maximum d'entropie

Astérisque, tome 121-122 (1985), Séminaire Bourbaki, exp. n° 628, p. 263-277

<http://www.numdam.org/item?id=SB_1983-1984__26__263_0>

Séminaire BOURBAKI
36e année, 1983-84, n° 628 Juin 1984

RECONSTRUCTION DES PHASES EN CRISTOLLOGRAPHIE PAR MAXIMUM D'ENTROPIE

[d'après G. Bricogne]

par Didier DACUNHA-CASTELLE

I. INTRODUCTION

Le problème de la reconstruction des phases en analyse par rayons X d'un cristal de grosses molécules d'origine biologique, à géométrie compliquée, est un problème dont la maturation a été lente, en particulier par non-utilisation des concepts probabilistes et statistiques adéquats. Si le modèle probabiliste a été introduit vers 1955 ([I-1], [I-2]), son exploitation est restée naïve du point de vue mathématique. Nous allons présenter (mathématiquement) certaines idées, notamment de G. Bricogne, indiquant le cheminement de l'outil sans doute transformable en bulldozer par des raffinements mathématiques encore à venir et par l'usage de gros moyens de calculs.

[La bibliographie est donnée par thèmes, relevant de chaque paragraphe.]

Cet exposé porte sur l'application d'idées (aussi fondamentales qu'élémentaires) de la statistique à la résolution d'un problème essentiel de la cristallographie X. La méthode du maximum d'entropie n'a évidemment que des fondements physiques. Ici, ils consistent à traiter le problème cristallographique comme un problème de mécanique statistique. On doit donc considérer que l'on traite mathématiquement une heuristique physique.

II. MODÈLE SIMPLIFIÉ ET PROBLÉMATIQUE

La protéine consiste en une répétition périodique d'un motif à identifier, compliqué, hautement "asymétrique", constitué de N_a atomes d'éléments différents. On trouvera en appendice des éléments mathématiques sur la cristallographie et les groupes finis associés. L'exposé serait de manière non essentielle compliqué par l'introduction de la notation algébrique convenable. Notons cependant que le travail numérique actuel comme le travail mathématique peuvent intégrer cet aspect.

Par diffraction rayons X, on mesure les *modules* des coefficients de Fourier de

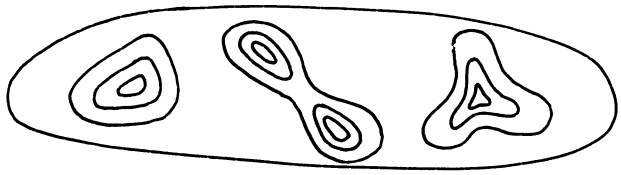
de Fourier de la densité électronique du motif, appelé molécule. On peut penser cette densité comme définissant la position des atomes dans une maille élémentaire, notée V donc comme une approximation d'une mesure décrivant la position des atomes $X_1 \ldots X_{N_a}$, soit

$$P_\infty = \frac{1}{N_a} \sum_{j=1}^{N_a} \delta_{X_j}$$

(le problème théorique qui vise à reconstruire P_∞ par la connaissance exacte des modules est classique).

En fait les mesures sont bruitées et donc les modules des coefficients sont connus avec une incertitude et de plus, sont en nombre limité, ce nombre est quantifié par le degré de résolution (mesuré en $\mathring{A}$) de l'expérience.

On cherche donc une densité électronique q_∞ approchant P_∞ par des estimations donnant des cartes de niveau de probabilité sur la position des atomes du type



avec des motifs caractéristiques, ces cartes étant déssinées à différents niveaux de résolution. Dans la suite, P,Q seront des probabilités sur le tore $\mathbb{T}^3 = V$.

Le modèle simplifié peut paraître stupéfiant : on suppose d'abord les N_a-atomes distribués au hasard suivant la loi q_∞ dans le volume indépendamment les uns des autres, c'est-à-dire que l'on *néglige* au *départ* tout ce que l'on sait des contraintes de la *stéréochimie*, en particulier les contraintes topologiques.

Il semble d'ailleurs que N_a ne doit pas être interprété strictement comme le nombre d'atomes, qu'il s'agit plutôt d'un paramètre mathématique lié au nombre d'atomes, de façon peu précise au départ, nous le noterons N.

III. PRÉLIMINAIRE : L'INFORMATION DE KULLBACK ET SA "GÉOMÉTRIE"

Soient P et Q deux probabilités, D une mesure dominante quelconque $dP = p d\nu$, $dQ = q d\nu$. L'information de Kullback de Q par rapport à P est

$$K(Q,P) = \int (\log \frac{q}{p})\, q d\nu .$$

On a $0 \leq K \leq \infty$, $K(Q,P)=0 \Leftrightarrow P = Q$. Si $dQ = q d\lambda_V$, λ_V loi uniforme sur V, alors $K(Q,\lambda_V) = -H(Q)$, H entropie de Q. Nous ne parlerons pas en général d'*entropie*, seulement d'*information*. Soit P une probabilité, résumant un modèle mathématique avant de faire une expérience. L'expérience impose une contrainte : rechercher Q dans un ensemble $\mathcal{C}$ de probabilités. L'information apportée par l'expérience est

$$K(\mathcal{C},P) = \inf_{Q \in \mathcal{C}} K(Q,P) .$$

Remarque.— K est une (la seule) information parce qu'elle a les propriétés suivantes : $K \geq 0$, K *diminue* par image, K est additive $K(P_1 \otimes P_2 , Q_1 \otimes Q_2) = K(P_1 \otimes Q_1) + K(P_2 \otimes Q_2)$, propriétés naïvement indispensables à toute information.

Dorénavant nous ne travaillons qu'avec des probabilités sur $\mathbb{R}^k$ telles que $\Phi(t) = \int_{\mathbb{R}^k} e^{t.x} \, dP(x) < \infty$ pour tout $t \in \mathbb{R}^k$. On supposera $P \neq \delta$ de sorte que $\log \Phi$ est strictement convexe et l'on pose pour tout a

$$h(a) = \sup[a.t - \log \Phi(t)]$$

(voir [III-1], [III-2], [III-3]). h est la transformée de Cramer de P (et caractérise P). On pose

$$\psi(t) = \operatorname{grad} \log \Phi(t)$$

et on note $G = \psi(\mathbb{R}^k)$ et G^o l'intérieur de G . Pour $a \in G^o$, on définit t_a par $\psi(t_a) = a$, alors

$$h(a) = at_a - \log \Phi(t_a) .$$

THÉORÈME.— *Soit* P *tel que* $\int x dP(x) = 0$, *et* $a \in G^o$, *il existe un point unique* P^a *tel que* $K(P^a,P) = \inf_{Q \in \mathcal{C}_a} K(Q,P)$, $\mathcal{C}_a = \{Q , \int x dQ(x) = a\}$, *et l'on a*

$$dP^a(x) = \frac{e^{t_a x}}{\Phi(t_a)} dP(x) , \quad K(P^a,P) = h(a) .$$

Démonstration.— $\inf_{Q \in \mathcal{C}_a} K(Q,P) = \inf_{Q \in \mathcal{C}_a} \int\left[\log \frac{q}{p^a} \frac{p^a}{p}\right] q d\nu$ où $p^a = \frac{e^{t_a x}}{\Phi(t_a)} p$.

$\inf_{Q \in \mathcal{C}_a} K(Q,P) = \inf_{Q \in \mathcal{C}_a} K(Q,P^a) + at_a - \log \Phi(t_a) = h(a)$, l'inf. étant obtenu pour $Q = P^a$.

Changement de probabilité sous contrainte

Soit P la probabilité avant l'expérience, le résultat de l'expérience impose une contrainte du type $\mathcal{C}_o$: $\int C(x) dQ(x) = c_o$ où $C : \mathbb{R}^k \longrightarrow \mathbb{R}^p$.

Nous appellerons C-arc exponentiel issu de P la famille de probabilité $\frac{e^{t.C(x)}}{\Phi_C(t)} dP(x) = dP_{t,C}$ avec $\Phi_C(t) = \int e^{t.C(x)} dP(x)$.

Supposons (pour simplifier), Φ_C défini sur $\mathbb{R}^p$, et $C_o \in \operatorname{grad} \log \Phi_C(\mathbb{R}^p) = \psi_C(\mathbb{R}^p)$, et soit t_o tel que $\int C(x) dP_{t_o}(x) = C_o$.

COROLLAIRE.— *La probabilité* $P_{t_o,C}$ *est la seule probabilité de la classe* $\mathcal{C}_o$ *telle que*

$$\inf_{\mathcal{C}_o} K(Q,P) = K(P_{t_o,C},P) = h_{CP}(C_o)$$

où CP *est l'image de* P *par* C , *et* $(\operatorname{grad} \log \Phi)(t_o) = C_o$.

Le problème de changement de probabilité après expérience est alors dit *bien réalisable*, la nouvelle probabilité est P_{t_o,C_o}. Elle est donc obtenue (par minimax) comme la probabilité respectant la contrainte expérimentale et apportant le moins

d'information par rapport à P .

Remarque.— L'idée de changement de probabilité est une clé essentielle des mathématiques expérimentales modernes. Si par exemple, on veut décrire une dynamique de très faible probabilité comme un changement de phase dans un modèle mathématique de la mécanique statistique, on sera amené à accélérer la simulation en changeant la probabilité de manière à amener une forte probabilité à la situation que l'on veut observer [III-4].

Comme nous l'avons vu, les contraintes sont en général plus floues que celles indiquées.

Nous appelons contraintes en tâches $\mathcal{C}_{c_0,\sigma^2}$ des contraintes du type

$$\int C(x)dQ(x) \in B(c_0,\sigma^2) \text{ , boule de rayon } \sigma$$

et contraintes type χ^2 des contraintes non sphériques du type

$$\frac{1}{N}\sum_{\alpha=1}^{N}\frac{[\int c_\alpha(x)dQ(x) - c_{\alpha o}]^2}{\sigma_\alpha} = 1 ,$$

où $c_{\alpha o}$, σ_α sont données, $\sigma_\alpha > 0$; enfin on peut définir un modèle de contrainte statistique

$$\int C(x)dQ(x) = C_o + \varepsilon ,$$

où ε est une loi gaussienne $\mathcal{N}(0,\Sigma)$ centrée, de covariance Σ donnée (nous n'utiliserons pas cette contrainte par souci de simplicité).

Remarquons par exemple pour la contrainte en tâche que le résultat précédent s'étend. Supposons $B(c_o,\sigma^2) \subset \psi_C(\mathbb{R}^p)$, alors

$$\inf_{Q\in\mathcal{C}_{c_0,\sigma^2}} K(Q,P) = \inf_{C_1\in B(c_o,\sigma^2)} \inf_{Q\in\mathcal{C}_{C_1}} K(Q,P) = \inf_{C_1\in B(c_o,\sigma^2)} h_{CP}(c_1) .$$

La fonction $h_{CP}(c_1)$ atteint son minimum sur la boule en un point $c_{o,\sigma}$ et on a encore une solution

$$\frac{e^{t_{o,\sigma}C(x)}}{\Phi(t_{o,\sigma})}\,dP(x) .$$

Remarque.— On indiquera en appendice les difficultés mathématiques dans le cas où $\psi_C(\mathbb{R}^p)$ n'est pas tout l'espace, (c'est la fermeture de l'enveloppe convexe du support de μ) et les propriétés de semi-continuité de K.

IV. LE PRINCIPE D'ENTROPIE MAXIMUM ET SON UTILISATION EN CRISTALLOGRAPHIE

Ce principe a une double origine : en théorie de l'information et en statistique (où il est plus utilisé sous des variantes comme le maximum de vraisemblance) et en physique (Jaynes).

En mécanique statistique, on peut décrire un "mode de fabrication" des lois fondamentales de la manière suivante (voir [IV-1] pour des détails) : soit P la probabilité uniforme sur l'ensemble fini des niveaux e d'énergie possibles des

différentes configurations d'un système régi par une contrainte d'énergie moyenne du type $\Sigma \Phi(e_k)p_k = c$. Alors la distribution donnant la loi des états sous cette contrainte est celle qui minimise $K(Q,P)$ sous contrainte, elle est donc de densité exponentielle par rapport à P comme nous l'avons vu.

Un *principe* de la physique pourrait être le suivant. Soit un phénomène décrit par une loi de probabilité P inconnue. La théorie (ou l'expérience) permettent de connaître une caractéristique C de P . Alors le principe de maximum d'entropie définira P de sorte que l'entropie de P soit maximum parmi les lois vérifiant C . Remarquons que cela nécessite le choix d'une mesure de référence dont le choix peut être soit une mesure uniforme considérée comme décrivant une "ignorance absolue", soit une mesure invariante pour un certain groupe lié au problème physique. La théorie des gaz parfaits est l'exemple classique.

Un exemple, qui est probablement l'utilisation première du principe, est celui du modèle de Gibbs-Ising sur $\mathbb{Z}^2$. Soit $V \subset \mathbb{Z}^2$, x la configuration de V ($x \in \{-1,1\}^V$ dans le cas du magnétisme), y celle de $\mathbb{Z}^2 - V$, $H_V(x|y)$ le hamiltonien donnant la distribution d'énergie de x lorsque y est fixée. Cherchons la densité ρ de probabilité sous la contrainte d'énergie moyenne fixée, soit $\int_V H_V(x|y)\rho(x)dx = C$. Par maximum d'entropie, on obtient $\rho(x) = \frac{1}{Z} \exp - \beta H_V(x|y)$ où Z est la fonction de partition normalisante et β lié à C . Des utilisations [IV-1] du principe aux phénomènes irréversibles sont importantes.

En cristallographie, l'argument de base est que dans un virus de 10^4 à 10^6 atomes, du point de vue de la diffraction X , les atomes H ne comptant pas, les atomes C , N , O ont à peu près le même poids atomique. Le modèle des cristallographes qui traite ces atomes comme des variables aléatoires indépendantes et équidistribuées à la manière de la mécanique statistique part de cette constatation. L'impasse est donc faite sur la dépendance due aux contraintes de la chimie. La méthode d'entropie a deux autres avantages : elle s'adapte très bien aux nécessités de rechercher des mesures invariantes pour certains groupes, et de plus, en cristallographie, des connaissances de phases à 30° près sont significatives. En premier certaines phases sont des signes, ensuite on connaît des "motifs" des cartes électroniques et leur nombre est relativement limité. Une image floue permet quand-même de classer le motif. La mise en oeuvre de la méthode d'entropie sera essentiellement de réaliser le découpage du mégaproblème en microproblème.

Dans le problème cristallographique nous utiliserons ce principe de deux manières différentes que nous allons décrire.

Soit A_M , $M = 1,\ldots K$ une famille croissante de parties finies de $\mathbb{Z}^3$ avec Card $A_M = \overline{M}$, $B_M = A_M \setminus A_{M-1}$. On désigne par $\mathcal{E}_M$ la famille de probabilités sur $V \subset \mathbb{R}^3$,

$$\mathcal{E}_M = \{Q\ ,\ \int e^{i\alpha x} dQ(x) = \hat{c}_\alpha \text{ pour } \alpha \in A_M\} \ .$$

On se donne Q_o et l'on définit par récurrence Q_M par

$$K(Q_M, Q_{M-1}) = \min_{Q \in \mathcal{C}_M} K(Q, Q_{M-1}) \ .$$

Nous verrons plus loin que la contrainte $\mathcal{C}_M$ est bien réalisable au sens du paragraphe II. Il résulte alors de la première partie que :

$$dQ_M(x) = \exp[t.C_M(x)]\Phi_{M-1}^{-1}(t_M)\, dQ_{M-1}(x) \ ,$$

où $t \in \mathbb{R}^{2\overline{M}}$, $C_M(x) = \{\cos \alpha x, \sin \alpha x, \alpha \in A_M\}$, $\Phi_{M-1}(t) = \int e^{t\, C_M(x)}\, dQ_{M-1}(x)$ avec $\text{grad}[\log \Phi_{M-1}](t_M) = \hat{c}_M$, $\hat{c}_M = \{\hat{c}_\alpha, \alpha \in A_M\}$.

Donc la connaissance de nouveaux coefficients de Fourier amènera à changer Q_{M-1} en Q_M suivant la voie indiquée. Indiquons que suivant la remarque déjà faite, il est possible de bruiter les mesures sans changer la méthode, ce qui donne une grande souplesse.

Il reste à décrire le principe de reconstruction des phases. A l'étape $M-1$ on suppose disposer d'une probabilité Q_{M-1}^E . Mais ici on ne mesure pas de nouveaux coefficients de Fourier, seulement leur module. On a donc pour passer de Q_{M-1} à Q_M à choisir l'argument φ_α de

$$\hat{c}_\alpha = \int e^{i\alpha x} dQ(x) \ , \qquad \alpha \in B_M$$

connaissant le module

$$\left[\int \cos \alpha x\, dQ(x)\right]^2 + \left[\int \sin \alpha x\, dQ(x)\right]^2 = \hat{d}_\alpha \ .$$

Soit alors $\mathcal{Q}_M$ l'ensemble des lois de probabilités Q telles que

$$\int e^{i\alpha x}\, dQ(x) = \int e^{i\alpha x} dQ_{M-1}(x) \ , \qquad \alpha \in A_{M-1}$$

$$\left|\int e^{i\alpha x} dQ(x)\right|^2 = \hat{d}_\alpha^2 \ , \qquad \alpha \in B_M \ .$$

Il y a plusieurs points de vue pour faire la reconstruction qu'il est bon de clarifier :

1) La méthode strictement analytique. On choisit les $Q_M \in \mathcal{Q}_M$ telles que

$$K(Q_M, Q_{M-1}^E) = \inf_{\mathcal{Q}_M} K(Q, Q_{M-1}^E) \ .$$

2) Le modèle probabiliste des cristallographes. Il considère en fait que les $\hat{c}_\alpha$ sont *estimés* de la manière suivante. Soit $X_1 \ldots X_N$ les positions de N atomes identiques, positions tirées au sort suivant la loi Q_{M-1}^E . Soit C la fonction $V \to \mathbb{R}^{2k}$ définie par $x \longrightarrow \{\sin \alpha x, \cos \alpha x, \alpha \in B_M\}$ et soit

$$\hat{C}_\alpha^N = \frac{1}{N} \sum_{j=1}^{N} \exp i\alpha X_j \qquad \alpha \in B_M$$

et

$$\hat{C}^N = \{\sqrt{N}\, \text{Re}\, \hat{C}_\alpha^N, \sqrt{N}\, \text{Im}\, \hat{C}_\alpha^N, \alpha \in B_M\} \ .$$

La loi de $\hat{C}^N$ est notée $C^N Q$ si X a la loi Q . Un point de vue intéressant est de chercher Q minimisant $K(C^N Q, C^N Q_{M-1}^E)$ sous la contrainte $|\hat{C}_\alpha^N|$ donnée par $\alpha \in B_M$, et $\hat{C}_\alpha^N$ donnée par $\alpha \in A_{M-1}$.

Si l'on interprète cette méthode, elle consiste en fait à minimiser sur les phases la distance d'information des lois conditionnelles de $C^N Q$ et $C^N Q^E$ lorsque $|\hat{C}^N_\alpha|$ est *donnée*.

Le modèle est physiquement raisonnable. N est considéré comme le nombre de *degrés de liberté* du système et sera finalement estimé (ce n'est pas nécessairement le nombre d'atomes). En pratique la précision sur $|\hat{C}^N_\alpha|$ dépend de N .

La transformation de Fourier calculée sur des éléments de $\mathbb{Z}^3$ fait perdre de l'information par image. La convolution renormée en regagne par addition de v.a. indépendantes. Nous verrons dans le dernier paragraphe en quoi, pour le problème de recherche de maxima la méthode analytique est l'asymptotique pour $N \to \infty$ du modèle probabiliste, absolument indispensable à traiter, pour justifier l'asymptotique et aussi pour suivre le traitement numérique. Il y a de gros obstacles à la mise en pratique :

a) Analytique d'abord car il n'y a pas nécessairement unicité de Q_M , le problème n'étant pas convexe. Rien n'indique que lorsque, *faisant des calculs de taille raisonnable*, on travaille de manière séquentielle en ajoutant les contraintes de B_M , un choix de minimum sur B_M ait vraiment de bonnes propriétés globales pour l'algorithme entier (i.e. lorsque M varie) ;

b) les problèmes sont de toute manière de telle taille que des approximations seront indispensables.

Le reste de l'exposé sera donc consacré à la description de l'algorithme et aux fondements des approximations faites. Ceci nous permettra d'éclaircir plusieurs points. Les approximations quadratiques des deux méthodes amènent au même problème d'optimisation qui est un problème de maximum de vraisemblance sur des lois gaussiennes. Les approximations non quadratiques sont utiles : les méthodes ne sont pas équivalentes et des problèmes nouveaux se posent.

V. APPROXIMATIONS NÉCESSAIRES À LA MISE EN OEUVRE

Nous aurons besoin de trois types d'approximations :

a) L'approximation de $\frac{dC^N Q}{d\lambda}$, pour N grand où $d\lambda$ est la mesure de Lebesgue sur $\mathbb{R}^{2k}$. Le nombre d'atomes joue donc le rôle de coefficient de l'asymptotique du modèle probabiliste.

b) L'approximation de $K(Q, Q^E_M)$ lorsque Q est paramétrée sous la forme

$$dQ = q(\vartheta_1 \ldots \vartheta_{k(M)})\, d\lambda ,$$

avec $q^E_M = q(0,\ldots,0)$, $\|\vartheta\|_{\mathbb{R}^{k(M)}}$ petit donne l'asymptotique.

c) Pour M grand, soit $T_M q$ la matrice de Toeplitz de $Q = q d\lambda$,

$$T_M q = \left\{ \int e^{i(j-k)x} dQ(x) \, , \, j,k \in A_M \right\} ,$$

avec Card A_M comme infiniment grand.

Nous aurons besoin d'approximations du type Szegö du genre

$$\Phi[T_M q] \# T_M(\Phi q) ,$$

où Φ est une fonction analytique, les approximations ayant lieu en norme nucléaire par exemple.

V.1. Le développement d'Edgeworth indirect. Le cadre du cas direct est le suivant :

Soient $U_1 \ldots U_N$, $U \in \mathbb{R}^k$, N v.a. indépendantes de même loi, de densité f telle que $\Phi(t) = E e^{tU}$ existe pour $t \in \mathbb{R}^k$. Commençons par le cas $k = 1$. Soient K_ℓ les cumulants de f , qui sont les coefficients du développement en série entière de $\log \Phi$. Dans le cas gaussien $K_\ell = 0$, $\ell > 2$, K_1 est la moyenne, K_2 la variance. On pose

$$\rho_\ell = \frac{K_\ell}{\sqrt{K_2}} \quad , \quad S_n = \frac{U_1 + \ldots + U_n - nK_1}{\rho_2 \sqrt{n}} .$$

On note g_n la densité de S_n et H_k les polynômes d'Hermite. Un argument simple d'inversion de transformée de Laplace donne

$$g_n(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\left[1 + \frac{\rho_3}{6\sqrt{n}} H_3(x) + \frac{1}{n}\left(\frac{\rho_4 H_4(x)}{24} + \frac{\rho_3^2 H_6(x)}{72}\right)\right]\right\} + \mathcal{O}_x\left(\frac{1}{n^{3/2}}\right) ,$$

développement que l'on peut poursuivre à un ordre quelconque.

Ce développement a plusieurs graves défauts. Sa bonne qualité en 0 se perd rapidement quand x croît, le terme en $1/\sqrt{n}$ est prépondérant, et l'échelle cristallographique pourra exiger de travailler avec des x de l'ordre de $\sqrt{n}$ où l'approximation gaussienne classique n'a plus de sens.

On passe alors au développement d'Edgeworth indirect obtenu à partir du changement de probabilité déjà introduit.

Soient

$$f_t(y) = \frac{e^{ty}}{\Phi(t)} f(y)$$

et $\hat{t}$ tel que

$$\int y f_{\hat{t}}(y)\, dy = \frac{x}{\sqrt{n}} \quad , \quad \hat{t} = t(x,n)$$

et soit $g_{n,\hat{t}}(x)$ la densité de $S_n = \frac{U_1 + \ldots + U_n}{\sqrt{n}}$ pour U de densité $f_{\hat{t}}$. Sous $g_{n,\hat{t}}$, S_n est centrée en x , on peut lui appliquer la formule précédente, calculée pour $x = 0$, d'où

$$g_{n,\hat{t}}(x) = \frac{1}{\sqrt{2\pi}}\left(\frac{1}{n}\frac{\hat{\rho}_4}{8} + \frac{5}{24}(\hat{\rho}_3)^2\right) + \mathcal{O}\left(\frac{1}{n^{3/2}}\right) ,$$

où $\hat{\rho}_j = \frac{\hat{K}_j}{\hat{K}_2}$, $\hat{K}_j$ cumulant de $f_{\hat{t}}$. On en déduit

$$g_n(x) = e^{n \log \Phi(\hat{t})} e^{-\hat{t}\, x/\sqrt{n}} g_{n,\hat{t}}(x)$$
$$= e^{-nh(x/\sqrt{n})} g_{n,\hat{t}}(x) ,$$

développement à l'ordre $\frac{1}{n}$ pour x fixé, mais qui vaut pour x de l'ordre de $\sqrt{n}$.

On a à travailler sur le cas multidimensionnel où les approximations sont du même

type [V-1]. Nous ne détaillerons pas l'utilisation critiquable faite par les cristallographes des approximations gaussiennes, mais indiquons un point essentiel : la nature même des contraintes imposées dans le problème de reconstruction des phases par minimum d'entropie implique de chercher des approximations type Edgeworth. Si p et q sont deux densités, le problème peut être éclairé par une analyse fine de la différence

$$K(p,q) - K(\sqrt{n}p^{*n}(\sqrt{n}\cdot),\sqrt{n}q^{*n}(\sqrt{n}\cdot))$$

entre l'information et ses approximations pour des lois gaussiennes.

V.2. Développement de l'information de Kullback

Soit une famille de probabilités $dP_\vartheta = p_\vartheta d\mu$, indexées par $\vartheta \in \mathbb{R}^k$, telles que l'application $\vartheta \longrightarrow \dot{p}_\vartheta = \frac{dp_\vartheta}{dx}$ soit définie presque partout pour une version convenable de p_ϑ et $\frac{\dot{p}_{\vartheta_o}}{p_{\vartheta_o}} \in L^2(p_{\vartheta_o})$. On appelle alors information de Fisher $I(\mathcal{O})$ matrice

$$\int \left(\frac{\dot{p}_\vartheta}{p_\vartheta} \otimes \frac{\dot{p}_\vartheta}{p_\vartheta}\right) p_\vartheta d\lambda .$$

On a, sous de faibles conditions de régularité [IV-3]

$$K(P_\vartheta, p_{\vartheta_o}) = {}^*\vartheta\, I(\mathcal{O}_o)\vartheta + \mathcal{O}(\theta-\theta_o)^2,$$

$I(\mathcal{O}_o)$ définit la métrique locale d'information que nous utiliserons plus loin.

V.3. L'homomorphisme asymptotique de Toeplitz

Un énoncé faible mais pratique est le suivant [V-3] :

Soit A l'algèbre des fonctions $\mathbb{T} \longrightarrow \mathbb{R}$ dont les coefficients de Fourier satisfont

$$\sum_k (1+|k|)\,|\hat{g}_k| = \alpha(g) < \infty .$$

Sur les matrices (n,n) soit b la norme $b(A) = \sum_{i,j} |a_{ij}|$. Soit T_n l'opérateur de Toeplitz

$$g \longrightarrow (\hat{g}_{i-j})_{1\le i,j\le n} = T_n(g) .$$

Alors pour $g_1 \ldots g_n \in A$,

$$b\left([T_n(g_1)\ldots T_n(g_k)] - T_n(g_1\ldots g_k)\right) \le (k-1)\alpha(g_1)\ldots\alpha(g_k) .$$

On en déduit aisément des approximations pour

$$\|T_n^{-1}(g) - T_n(\tfrac{1}{g})\|_{nucl}$$

et pour

$$|\log\det T_n(g) - n\int \log g| .$$

VI. INVARIANTS ASSOCIÉS AU GROUPE SPATIAL CRISTALLOGRAPHIQUE

Nous ne détaillerons pas cette partie (bien mathématisée !).

Un point important est le suivant : dans la construction par maximum d'entropie de la densité électronique à partir de la loi uniforme comme loi a priori, on obtient des relations algébriques sur les coefficients de Fourier, résultant des propriétés d'invariance de la densité électronique pour le groupe spatial. Le formalisme adéquat est d'utiliser l'expression de la fonction de partition à l'aide de multiplicateurs de Lagrange (formalisme qui occulte en partie le problème comme nous l'avons présenté). La fonction de partition Φ est donc exprimée à l'aide des multiplicateurs κ_α. On a

$$\frac{\partial}{\partial \kappa_\alpha} \log \Phi = |c_\alpha| \cos \vartheta_\alpha ,$$

$$\frac{1}{\kappa_\alpha} \frac{\partial}{\partial \vartheta_\alpha} \log \Phi = -|c_\alpha| \sin \vartheta_\alpha$$

Soit maintenant $h = \{h_\alpha , \alpha = 1,\ldots M\}$ un ensemble fixé de $\mathbb{Z}^3$. Soit u_α un élément du groupe cristallographique (spatial) G, $u_\alpha = (R_\alpha, t_\alpha)$ où R est une transformation unitaire et t_α une translation.

On considère les éléments de $\mathbb{Z}G$ notés $\sum_{g \in G} m(g).g$. On pose

$$\sum_{u \in G} m(\alpha,u)\, u = \sigma_\alpha$$

élément de $\mathbb{Z}G$, $\sigma = (\sigma_1 \ldots \sigma_M)$,

$$K = \{\sigma \in (\mathbb{Z}G)^M , \sum_\alpha R^*(\sigma_\alpha) h_\alpha = 0\} .$$

Soient $\mathcal{J}_\alpha$ les fonctions de Bessel, $\exp z \cos t = \sum_{-\infty}^{\infty} I_\alpha(z) \exp i\alpha t$,

$$\eta_\alpha = \sum_{u \in G} m(\alpha,u) ,$$

$$\mathcal{J}_\alpha\left(\frac{\kappa_\alpha}{|G|}\right) = \prod_{u \in G} I_{m(\alpha,u)}\left(\frac{\kappa_\alpha}{|G|}\right) \exp[-2\pi i\, h_\alpha t(\sigma_\alpha)] ,$$

$$\mathcal{J}_\sigma\left(\frac{\kappa}{|G|}\right) = \prod_\alpha \mathcal{J}_\alpha\left(\frac{\kappa_\alpha}{|G|}\right) ,$$

$$\psi_\alpha = \varphi_{h_\alpha} + \vartheta_\alpha ,$$

$$\eta\psi = \sum_\alpha \eta_\alpha \psi_\alpha ,$$

alors

$$\Phi(\kappa,\vartheta) = \sum_{\sigma \in \kappa} \mathcal{J}_\sigma\left(\frac{\kappa}{|G|}\right) \exp i\eta\psi ,$$

formule contenant les propriétés d'invariance par G de q et de ses coefficients de Fourier. Ces formules seront plus utiles pour *tester* de la qualité des solutions du problème de phases que pour la recherche elle-même, et aussi pour interpréter les méthodes utilisées jusqu'à ce jour, comme des approximations.

VII. L'ÉTAT ACTUEL DES ALGORITHMES

On est loin d'être capable d'exploiter numériquement l'approche précédente dans tout ce qu'elle doit apporter, et il reste beaucoup à faire. Le principal problème est de chercher, à chaque pas, dans des espaces de taille raisonnable et de contrôler la croissance de la taille de l'arbre des solutions possibles, croissance due à la non-convexité.

Dans l'état actuel des choses, la solution utilisée est un algorithme quadratique utile en restauration d'image et modifié pour faire face à la non-convexité, peu traitée semble-t-il, dans ce type de situation par les professionnels de l'optimisation.

Le travail se fait pour sortir du quadratique en utilisant des apports plus fins de géométrie différentielle.

VII.1. L'algorithme de minimum d'information en restauration d'image

On mesure l'intensité (de gris) d'une image en ses divers points notés i (par exemple un carré $2^6 \times 2^6$).

Un modèle élémentaire, intéressant en restauration est

$$d_i = \sum_j f_j k_{i-j} + e_i \quad (*)$$

où f est la vraie densité, k un filtre numérique d'étalement (donné par le procédé d'imagerie employé) et e un bruit blanc par exemple gaussien, de variance σ_i^2 en i (σ_i pouvant croître avec f_i).

Partant d'une distribution gaussienne de moyenne constante, on va chercher par minimum de distance d'information (minimum de "dissemblance") la moyenne f d'une distribution inconnue.

Soit A la valeur choisie pour la distribution uniforme. On supposera $\Sigma f_i = A$ ("conservation de l'énergie") et il faut ajouter une contrainte (de type statistique) pour ne pas chercher f "trop loin" de l'observation d .

Si N est le nombre de points de l'image et $\chi^2(N)$ la distribution du χ^2 à N degrés de liberté, χ^2_{95} le quantile à 95% de ce χ^2 , la proximité est assurée par la contrainte d'*observation*

$$\sum_i (d_i - \sum_i f_j k_{i-j})^2 \frac{1}{\sigma_i^2} < \chi^2_{95} .$$

On peut, et c'est un problème intéressant, utiliser d'autres distances associées à la distribution des résidus e_i sous l'hypothèse que (*) est vérifié.

L'algorithme d'image consiste donc à minimiser la dissemblance entre la gaussienne de moyenne uniforme A et celle de moyenne $f * g$, ce qui revient à minimiser la forme quadratique

$$\mathcal{V}(f) = \sum_i [(f*k)_i - A]^2$$

sous les contraintes $\Sigma f_i = A$ et $\sum_i [d_i - (f*k)_i]^2 \frac{1}{\sigma_i^2} \leq \chi^2_{95}$.

En fait le problème majeur est de limiter l'espace de recherche de f . En effet, le problème étant convexe, il n'y a aucune difficulté majeure à appliquer la méthode des multiplicateurs de Lagrange agrémentée de la méthode du gradient.

Le choix technique est différent de celui du problème de l'espace de recherche cristallographique.

Remarquons qu'en général la solution est unique et astreinte au bord des contraintes.

VII.2. L'algorithme précédent modifié en cristallographie

Soit Q_M à une étape M donnée la probabilité a priori. $\mathcal{J}_M(P)$ sera une approximation de $K(C^{M+1}P, C^{M+1}Q^M)$ avec les notations du chapitre II. Soit c_α , $\alpha \in A_M$ les coefficients de Fourier sur lesquels on a une information sur $d_\alpha = |c_\alpha|$, et pour lesquels on recherche la phase φ_α , $c_\alpha(P) = d_\alpha e^{i\varphi_\alpha(P)}$.

A partir de l'algorithme précédent, on globalise la contrainte en prenant

$$\frac{1}{\text{Card } A_M} \sum_\alpha |d_\alpha(P) - d_\alpha(\text{obs})|^2 \leq \sigma^2 \quad \text{(donné)}$$

où $d_\alpha(\text{obs})$ est le module observé.

Comme précédemment la recherche de P minimisant $\mathcal{J}_M(P)$ sous la contrainte ci-dessus, se fera par des méthodes de gradient (améliorées), à condition de limiter l'espace dans lequel on cherche P . L'algorithme le plus simple consiste à paramétriser le problème à $\text{Card } A_M$ dimensions, en cherchant simplement

$$dP(x) = dQ_M(x) + \Sigma(c_\alpha - c_\alpha(Q_M))e^{i\alpha x} .$$

Donc, si

$$\vartheta = \{c_\alpha - c_\alpha(Q_M) \ , \alpha \in A_M\} \ ,$$

on cherche ϑ tel que $\mathcal{J}(P_\vartheta)$ soit minimal.

Pour calculer une approximation de $K(C^{M+1}P, C^{M+1}Q_M)$ où $C^{M+1}P$ a toujours pour densité $\sqrt{n}p^{*n}(\sqrt{n}\cdot)$, on peut utiliser l'approximation d'Edgeworth des densités de $C^{M+1}P$ et $C^{M+1}Q_M$ en ne retenant que le terme gaussien. Ce calcul amène simplement à la minimisation de l'approximation au 2e ordre de $K(P, Q_M)$ soit

$$\int \log \frac{p_\vartheta}{p_o} \, p_\vartheta dx$$

qui vaut $\frac{1}{2} \vartheta I(\mathcal{P}) \vartheta + \mathcal{O}(\vartheta^2)$, où $I(\mathcal{P})$ est la matrice d'information de Fisher du problème paramétrisé

$$I(\mathcal{P}) = \int \frac{(dp/d\vartheta)^{\otimes 2}_{\vartheta=0}}{p_o} dx \ .$$

Ici $I(\mathcal{P}) = T_M(\frac{1}{\sigma^M})$ où T_M est la M-matrice de Toeplitz d'où un calcul économique. Sous cette approximation on remarque que la quantité à recherche n'est autre que le maximum de vraisemblance $\hat{\vartheta}$ du modèle P_ϑ à calculer sous la contrainte non convexe

$$\frac{1}{\text{Card } A_M} \sum_\alpha |d_\alpha(P_\vartheta) - d_\alpha(\text{obs})|^2 \leq \sigma^2 \ .$$

Ces estimations font douter de l'utilité du modèle probabiliste un peu sophistiqué introduit par les cristallographes et le signification de N. En fait, le processus d'optimisation se fait donc sur le modèle *macroscopique* obtenu par passage à la limite (pour $N = \infty$). Ce passage à la limite est justifié par le résultat (déjà esquissé) suivant, soit Q la probabilité initiale $Q+\delta$ la probabilité perturbée de façon à réajuster les modules de certains coefficients, δ étant petite en norme L^2, alors si $P^{\circ n}$ est la loi de $\frac{Y_1+\ldots+Y_n}{\sqrt{n}}$ lorsque P est la loi de X_i, $Y_i = \{e^{i\alpha X_i},\ \alpha \in A\}$ on peut montrer à partir des approximations indiquées que la perturbation δ étant centrée on a, lorsque le nombre de coefficients de Fourier perturbés permet d'appliquer convenablement l'homomorphisme asymptotique de Toeplitz, $K((Q+P)^{\circ N}, Q^{\circ N}) = I_o(Q)\delta + o\|\delta\|^2 = K(Q+\delta,Q) + o\|\delta\|^2 = K(G_\delta,G)$, où G_δ, G sont les lois normales limites. On sait que si la perturbation est finie, $I_o(Q)$ est la matrice d'information de Fisher et $I_o(Q) = T_{\bar{A}}(\frac{1}{f})$ où $\bar{A}$ est le nombre de coefficients de Fourier perturbés. De plus on peut borner en fonction de N, M l'erreur faite.

L'application $\{X \to e^{i\alpha X},\ \alpha \in A \subset \mathbb{Z}^3\}$ amène, notamment quand A est petit, du fait que les α sont des éléments de $\mathbb{Z}^3$, une perte d'information par manque de "bijectivité" dans l'image. Asymptotiquement pour N et M grands l'information perdue est évidemment très petite, et contrôlable. Ceci explique donc que le processus d'optimisation puisse être mené sans tenir compte du nombre d'atomes N. Par contre, la situation devient plus complexe lorsque l'on voudra sortir du quadratique. Le facteur N intervient devant les vraisemblances des différentes solutions qui vont intervenir dans le problème de reconstruction de phases, ensemble dû à la non-convexité des contraintes.

En effet dans la procédure séquentielle, à chaque étape il faudra prendre en compte les minima locaux. Pour cela, *on recale* à chaque étape Q_{M+1} sur la loi uniforme, on obtient pour chaque minimum de $K(Q_{M+1},Q_o)$ une loi de $\{\frac{e^{i\alpha X_1}+\ldots+e^{i\alpha X_N}}{\sqrt{N}},\ \alpha \in A_{M+1}\}$ sous Q_{M+1}, ou plus exactement une approximation. On obtient donc la loi conjointe du module et de la phase et finalement la loi marginale du module. *Le contrôle de la procédure se fait en calculant la vraisemblance au point observé du module*, ce niveau de vraisemblance dépend de N, il servira de test pour supprimer des branches de l'arbre des solutions éventuelles, N est donc essentiellement estimé à cette étape par maximum de vraisemblance.

Le passage au non linéaire est difficile. La contrainte est pour chaque coefficient de Fourier du type "quartique en cul de bouteille", la contrainte globale pourrait être traitée par déployement des singularités, l'approximation quadratique globale étant probablement insuffisante. L'approximation gaussienne est également insuffisante vu l'ordre de grandeur des observations. Le développement Edgeworth sous la forme utilisant la transformée de Cramer permet d'analyser précisément

des développements d'information à l'ordre 3 par rapport à la norme de la perturbation.

VII.3. Branchement et non convexité des contraintes (résumé)

La contrainte n'étant pas convexe, il est nécessaire de tenir compte à chaque étape de la procédure séquentielle de tous les minima locaux. On développe donc un arbre dont les noeuds à la hauteur M sont des minima locaux. La taille de l'arbre doit être contrôlée. On utilise pour ce faire

1) l'arrêt lorsque le gain d'information est inférieur à un seuil donné ;

2) le test sur les relations algébriques ;

3) des considérations externes au problème (utilisant par exemple des informations stéréo-chimiques) impossibles à introduire dans l'état actuel des choses dans le modèle d'optimisation.

CONCLUSION

Les idées statistiques nécessitent encore des années de travail technique pour être toutes utilisables. Il faudrait introduire dès à présent dans le processus, notamment au niveau branchement, des idées topologiques qui rendraient impossibles certaines solutions, incompatibles avec la stéréo-chimie, et aideraient à la lecture des cartes. D'autres problèmes mathématiques importants se posent dans le domaine esquissé ici.

APPENDICE BIBLIOGRAPHIQUE

I. Cet exposé part d'un article (à paraître) de Gérard BRICOGNE, *Maximumentropy and the Foundations of direct methods.*

Les fondements du modèle probabiliste remontent à E.F. BERTAUT, Acta Cryst. 8 (1955), 537-543, et à A. KLUG, Acta Cryst. 11(1958), 519-543.

Un point de vue différent se trouve dans V. LUZZATI and D. TAUPIN, *Accuracy and resolution in Protein Crystallography : a probabilistic approach*, J. App. Cryst. (1984), 17-30.

Le livre de C. GIACOVAZZO, *Direct methods in Crystallography*, Londres Acad. Press, 1980, pourra être consulté pour le formalisme algébrique.

Le modèle cristallographique général est exposé dans tous les livres de physique. Nous n'étudions ici que la *diffraction* dans un cristal formée de répétitions périodiques d'un motif (très compliqué) qui est d'origine biologique et contient de 10^3 à 10^4 atomes pour les applications en vue. Le faisceaux est monochromatique.

Si (a,b,c) est une base du réseau périodique, (a*,b*,c*) une base du réseau réciproque, les directions de diffractions sont de la forme $s = ha^* + kb^* + \ell c^*$, h, k, ℓ entiers, ce qui s'exprime classiquement par la loi de Bragg.

III. Un exposé assez complet sur les méthodes de grandes déviations se trouve dans [III-1] Séminaire d'Orsay sur les grandes déviations, Astérisque n° 68 (1979).
[III-2] le livre de S. KULLBACK, *Information theory and statistics*, Wiley New York, 1959, est déjà ancien et on peut signaler pour les difficultés d'ordre mathématique :
[III-3] I. CSISZAR, *I-divergence geometry of probability distributions and minimization problems*, Annals of Prob. vol. 3 n° 1(1971), 146-158.

L'application du changement de probabilité à la simulation peut être vue par exemple dans [III-4] M. COTTRELL et J.C. FORT, *Evénements rares pour certains algorithmes stochastiques*, Orsay (1980) (Version anglaise dans IEEE).

IV. Les articles de Jaynes (et ses livres) sont nombreux. Une synthèse avec exemples : [IV-1] E.T. JAYNES, *The maximum entropy formalism*, ed. Raphael D. Levine, Myron Tribus, MIT Press 1979.

Une reprise en cristallographie est S.W. WILKINS, J.N. VARGHESE and M.S. LEHMANN, Acta Cryst. A39(1983), 47-60.
[IV-2] donne les propriétés élémentaires citées : D. DACUNHA-CASTELLE et M. DUFLO, *Probabilités et statistiques*, Tome 1 (chap. 2 et 7), Masson, 1982.

V. Le principe du développement d'Edgeworth indirect est plus connu sous le titre de "approximation du point col" et est dû à S. DANIEL (saddle point) et fondé sur un raisonnement de fonctions analytiques.

Une présentation des aspects probabilistes est : [V-1] O. BARNDORFF NIELSEN and D.R. COX, *Edgeworth and saddle-point approximations with statistical applications*, J. Roy. Stat. Soc. B n° 3(1979), 279-312.

L'information de Fisher est introduite dans [IV-2].

L'approximation de Szegö se trouve dans [V-2] H. WIDOM, *Asymptotic inversion of convolution operators*, Annales I.H.E.S. (1979), 191-240. Son usage en statistique est développé dans [V-3] R. AZENCOTT et D. DACUNHA-CASTELLE, *Séries d'observations irrégulières*, Masson, 1984.

Des majorations fines sont à paraître : A. SEGHIER (Orsay prépublications, 1984).

VI. Pour l'aspect algébrique, renvoyons au livre cité en I de GIACOVAZZO.

VII. Les algorithmes de restauration d'images peuvent être consultés dans : S.F. BURCH, S.F. GULL and J. SKILLING, *Computer Vision*, Graphics and Image Processing, n° 23(1983), 113-128.

Didier DACUNHA-CASTELLE

Université de Paris-Sud
Département de Mathématiques
Bâtiment 425
F-91405 ORSAY