

X線自由電子レーザーによる分子の電子密度推定

池田 思朗¹・河野 秀俊²

(受付 2012年6月29日; 改訂 10月9日; 採択 10月10日)

要 旨

米国, 欧州, 日本及び韓国では X 線自由電子レーザーと呼ばれる装置が開発され, 米国, 日本では運用段階にある. 現在, この装置によって初めて可能となった X 線領域の波長を持つ高強度レーザー光を用いた, 様々な実験が計画および実施されている. 本稿では, そうした実験のひとつ, X 線回折を用いた生体単粒子の電子密度推定について説明し, 2 次元回折画像から 2 次元電子密度を推定する位相回復問題に対する我々の研究成果について解説する.

キーワード: X 線回折, XFEL, 位相回復, MAP 推定, 疎性.

1. はじめに

近年開発された X 線自由電子レーザー (X-ray Free Electron Laser: XFEL)¹⁾によって, これまでにない高強度 X 線レーザー光が出力できるようになった. 現在, 日本では XFEL の運用が始まり, このレーザーを用いる様々な実験が開始されている. 本稿では, その実験のなかから生体単粒子の 3 次元電子密度推定について説明し, その中で必要となる位相回復問題に関する我々の研究成果について解説する.

生体単粒子, その中でも特にタンパク質分子などの生体高分子の 3 次元構造を知ることは, 医学を始め, 生化学の様々な分野で重要である. 生体高分子の詳細な立体構造は, 主に X 線結晶構造解析によって決定されている. X 線結晶構造解析では, 結晶中の分子の反復構造によって回折した X 線が作る膨大な数の回折点を取得する. この回折点を用いて分子の電子密度を推定し, 立体構造を決定する. この方法によって, すでに 8 万以上のタンパク質分子の立体構造が決定されている.

しかし, 全ての生体高分子の構造を X 線結晶構造解析によって決定するのは難しい. 細胞中の約半数のタンパク質分子は結晶にならないからだ. そうした結晶にならない生体高分子に対し, 単粒子の状態では X 線を当てて回折画像を計測し, 電子密度を解析する実験が開始されている. 単粒子であるがゆえに, 回折データを得るためには位相の揃った非常に強い X 線を照射しなければならない. しかし, このような非常に強い X 線にさらされると, 分子は数十フェムト秒で壊れてしまう. そのため, 実験に用いる X 線レーザーは, 高強度かつ短パルスでなければならない. XFEL の光は, 位相の揃った高強度かつ短パルスの X 線であるため, 分子が破壊される前に回折画像を撮影できると考えられている. 様々な方向から生体高分子の回折画像を撮り, 得られた複数の回折画像を処理することによって 3 次元電子密度を推定する方法が提案されている (Neutze et al., 2000; Huldt et al., 2003).

¹ 統計数理研究所: 〒190-8562 東京都立川市緑町 10-3

² 日本原子力研究開発機構: 〒619-0215 京都府木津川市梅美台 8-1-7

こうした方法を用い、実際に巨大ウィルスであるミミウィルスのコヒーレント X 線イメージングが行われている (Seibert et al., 2011). しかし、生体内のタンパク質分子などは一般にミミウィルスよりはるかに小さいため、XFEL であってもミミウィルスと同等のコントラストを持つ回折画像を撮るのは難しく、3次元電子密度を推定するには様々な困難があると想像される. 本稿ではそうした困難の中から2次元回折画像からの位相回復という問題を取りあげ、我々が提案する Sparse Phase Retrieval (SPR) 法 (Ikeda and Kono, 2012) について説明する. SPR 法は LASSO などと同様に情報源の疎性を仮定した推定法であり (Tibshirani, 1996; Zhao and Yu, 2006), ベイズ統計の枠組においては最大事後確率法 (Maximum a posteriori: MAP) 推定を行なうことに対応する. SPR 法は、従来の方法が適用できない回折画像面に到達する光子数が少ない状況下において、単粒子の電子密度推定を可能にする.

2. XFEL を用いた生体単粒子回折画像の計測

2.1 回折画像

まず、回折画像の基本的な性質について説明する. 図 1 に実験の概略を示す. 真空中に落下させた生体単粒子に X 線を当て、回折画像を得る. 図中に示したように I (photons/pulse/mm²) は X 線の強度、 L は分子の直径である.

生体単粒子の電子密度を X 線と垂直な面に射影したものを $f_{xy} \geq 0$, $x, y = 1, \dots, M$ と表わす. x, y は 2次元座標のインデックスである. X 線レーザー光によって得られる回折画像はその f のパワースペクトラム, すなわち 2次元フーリエ変換したものの大きさに対応する量である. そこで, f_{xy} のフーリエ変換を F_{uv} とする. 定義は以下の通りである.

$$(2.1) \quad F_{uv} = \frac{1}{M} \sum_{x,y} f_{xy} \exp\left(\frac{2\pi i(ux + vy)}{M}\right).$$

通常、上式の右辺にある係数は $1/M$ ではなく $1/M^2$ だが, f_{xy} と F_{uv} の 2乗ノルムを等しくするため、本稿では上式の定義を用いる. 簡便のため、以下の記号を定義しておく.

$$\mathbf{f} = \{f_{xy}\} = \{f_{11}, \dots, f_{MM}\}, \quad \mathbf{F} = \{F_{uv}\} = \{F_{11}, \dots, F_{MM}\}.$$

また、フーリエ変換を $\mathbf{F} = \mathcal{F}(\mathbf{f})$ と表わすことにする.

2次元電子密度 \mathbf{f} と理想的な回折画像 S_{uv} の間には以下のような関係があることが知られている.

$$(2.2) \quad S_{uv} = \alpha |F_{uv}|^2 c_{uv} = Ir_c^2 \left(\frac{\lambda}{\sigma L}\right)^2 |F_{uv}|^2 c_{uv}.$$

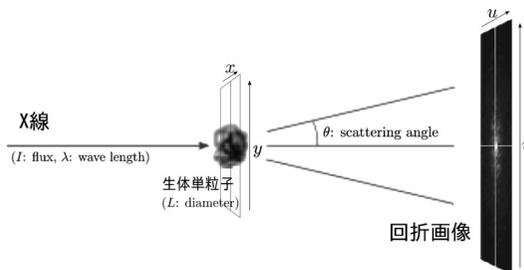


図 1. 生体単粒子を用いた XFEL による回折実験の概略図.

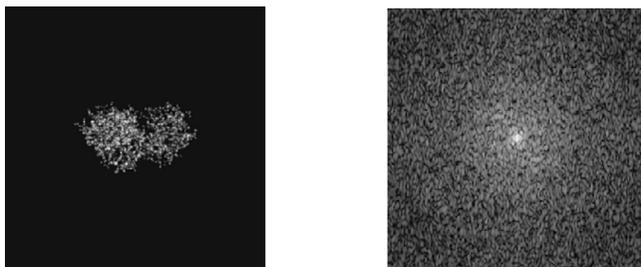


図 2. リゾチームの電子密度 (左) と理想的な回折画像 (右). 各画像は 308×308 の大きさの像である.

ここで r_c は電子の古典半径 (2.82×10^{-12} mm), λ は X 線の波長 (0.1 nm), σ は各次元のオーバーサンプリング比 (2), $c_{uv} = c(u, v) = \cos^3 \theta$ である (図 1 参照). 本来は回折画像は平面ではなく Ewald 球面上の画像として理解すべきだが, ここでは回折の角度が小さいとしてその影響は無視する. 本研究の方法を Ewald 球面上のデータに適用することは可能である.

図 2 に理想的な状態での電子密度と回折画像をシミュレーションで作成したものを示す. 結晶化した生体単粒子を用いた X 線回折画像も同様の関係式に従う. 結晶は非常に多くの粒子が規則的に並んだ構造を持つため, 電子密度のフーリエ変換である回折画像はブラッグの回折条件 (Bragg's Law) を満たした点のみで光子が観測される. 一方, 単分子の回折画像は図にあるような全体に連続的に広がった像となる.

2.2 XFEL による回折画像

図 3 に示すのは XFEL による X 線レーザーをリゾチームというタンパク質分子にあてて得られる回折画像を計算機シミュレーションによって作った例である. 図 2 に対して粗い画像であることがわかる. こうした画像となる理由は, (2.2) 式に示したように, 回折強度が古典半径の自乗に比例しているため小さい単粒子では回折強度が弱くなり, 高強度の XFEL であっても観測面に到達する光子の数が少なくなってしまうからである. この図では 4630 個の光子のみが観測面に到達している. 図 3 は約 90000 画素の図であるから, ほとんどの画素では光子が観測されていないことがわかる. また, 各画素で観測されるのは光子の数, つまり理想的な場合のような実数値ではなく, 非負の整数値である²⁾.

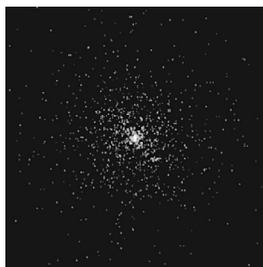


図 3. XFEL によって得られると予想される回折画像. 強度を表す I は 5.0×10^{21} photons/pulse/mm² であり, 観測面に到達した光子の数は 4630 個である.

画素 (u, v) で得られる光子数を N_{uv} と置く. N_{uv} の期待値は (2.2) 式の S_{uv} だが, 実際に観測される光子数は全く同一の構造に同じ方向から X 線を照射したとしても同じになるとは限らず, 実験を行う度に確率的に振舞う. 全光子数 $N_{all} = \sum_{uv} N_{uv}$ の期待値は強度 I に比例し, 各画素で観測される光子数はポアソン分布に従うことが知られている. 各画素で観測される光子数は独立で, 各 N_{uv} が独立でポアソン分布に従うというのは光学で一般的に用いられている仮定である (Hau-Riege et al., 2005). このとき N_{uv} の分布は以下のように書ける.

$$p(N_{uv}|S_{uv}) = \frac{S_{uv}^{N_{uv}} \exp(-S_{uv})}{N_{uv}!}.$$

(2.2) 式より S_{uv} は α を正の定数として $S_{uv} = \alpha |F_{uv}|^2 c_{uv}$ とかける. フーリエ変換によって得られる F は f の関数であることから, $\mathbf{N} = \{N_{uv}\} = \{N_{11}, \dots, N_{MM}\}$ の分布は f の関数であり, 以下のように書ける.

$$(2.3) \quad p(\mathbf{N}|f) = \prod_{uv} \frac{(|F_{uv}|^2 c_{uv})^{N_{uv}} \exp(-|F_{uv}|^2 c_{uv})}{N_{uv}!}.$$

なお, ここでは α は 1 とした. データだけから α の大きさと $|F_{uv}|^2$ の大きさの両方を同時に決定できない. $|F_{uv}|^2$ の大きさは全エネルギーなど別の情報をもとに決定する必要がある.

2.3 位相回復

本稿での位相回復とは, 回折画像から電子密度を推定する問題である³⁾.

まずは, 光量は十分にであると仮定し, S_{uv} から f_{xy} を復元する際, なぜ位相回復問題が生じるかを説明する.

S_{uv} が観測されるならば, $|F_{uv}|$ が観測できるとして良い. F_{uv} は f_{xy} のフーリエ変換にあたることから, 逆フーリエ変換を用いて f_{xy} が復元できそうである. しかし観測されるのは, F_{uv} の大きさのみであって位相情報は得られないため, 逆フーリエ変換を適用できない. したがって f_{xy} の復元には困難が伴う.

別の見方をしよう. (2.1) 式の F_{uv} の定義, そして f_{xy} が実数をとることから,

$$(2.4) \quad \begin{aligned} F_{M-u, M-v} &= \frac{1}{M} \sum_{x,y} f_{xy} \exp\left(\frac{2\pi i((M-u)x + (M-v)y)}{M}\right) \\ &= \frac{1}{M} \sum_{x,y} f_{xy} \exp\left(\frac{-2\pi i(ux + vy)}{M}\right) = F_{uv}^*, \quad (* \text{ は複素共役を示す}). \end{aligned}$$

すなわち $|F_{M-u, M-v}| = |F_{uv}|$ となる. 回折画像 S_{uv} は M^2 画素の画像だが, そのうち半分は同じ値を持つことになり, S_{uv} からは $M^2/2$ 個の画素の値しか得られない. これらの値を用いて f_{xy} の M^2 の画素の値を復元しなければならない. したがって位相回復問題は不良設定問題である (図 4).

2.4 既存法

ここでは, 単粒子の 2 次元回折画像の解析における位相回復問題に対する既存の方法について説明する.

既存の方法では, 2 次元空間の電子密度 f が正となりうる領域, すなわち正となりうる成分が限定されていると仮定する. 言い換えると, ひとつの分子のみが真空状態に存在し, そのまわりには何も無い状態を仮定する. その領域を γ と書くことにする. そして, 次の関数の最適

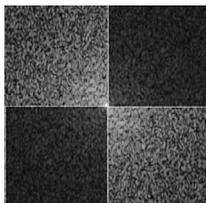


図 4. 回折画像の対称性. 同じ色の領域は同じ値をもっている.

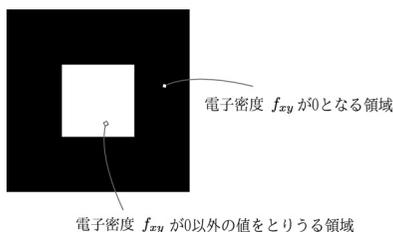


図 5. (2.5) 式における制約. この図では正方形の領域を仮定したが, 任意の形で構わない.

値を与える f を推定結果としている.

$$(2.5) \quad \begin{aligned} & \text{minimize} \sum_{uv} (|F_{uv}|c_{uv}^{1/2} - N_{uv}^{1/2})^2, \\ & \text{subject to } f_{xy} \geq 0, (x, y) \in \gamma \text{ and } f_{xy} = 0, (x, y) \notin \gamma. \end{aligned}$$

上式の目的関数は電子密度を f としたときの回折画像 $|F_{uv}|^2 c_{uv}$ と観測された回折画像の各画素の値 N_{uv} との近さを表している. γ 上のみで正となりうる f によってこの目的関数を小さくすることが目的である (図 5).

この問題を解くアルゴリズムとして広く用いられているのが hybrid input-output (HIO) 法 (Fienup, 1978; Fienup, 1982; Miao et al., 1998) である. XFEL による単粒子回折画像の解析でも HIO 法が用いられている (Neutze et al., 2000; Huidt et al., 2003). HIO 法が (2.5) 式の最適化問題を解いていること, また, HIO 法のアルゴリズムが既存の最適化法である Douglas-Rachford 法の適用例とみなせることは既に報告されている (Fienup, 1982; Bauschke et al., 2002).

単粒子の重さなど他の指標によって f の正となりうる領域 γ がわかっている, 理想的な回折画像 $|F_{uv}|^2 c_{uv}$ が正確に求まっている場合には, HIO 法によって電子密度が復元できる. しかし, 理想的な回折画像を実験的に得るためには, 十分な光量が必要である. ミミウィルスを対象とした XFEL による単粒子回折画像の解析に関する先行研究では HIO 法によって電子密度が復元されている (Seibert et al., 2011). これは粒子が大きいため十分な光量が得られ, 理想に近い回折画像が観測できたためだと考えられる.

3. 提案する方法

3.1 SPR 法

タンパク質分子などの単粒子に XFEL を用いて得られる回折画像は十分な光量が得られないため, 前節で示した既存の方法は必ずしもうまく働かない. そこで我々は SPR 法を提案し

た (Ikeda and Kono, 2012).

(2.3)式にあるように、観測された回折画像は確率モデルによって記述できることから、電子密度 f を復元することは推定の問題である。しかし、2.3 節で説明したように、位相回復問題は不良設定問題であり、最尤推定に関しても解は一意に求まらない。そこで、我々は電子密度に関する事前知識を事前分布として記述し、ベイズ統計の枠組における MAP 推定によってこの問題を扱った。位相回復に対してベイズ統計の方法を用いた方法は他にも提案されているが、我々とは問題設定、事前分布の選択が異なる (Irwan and Lane, 1998; Baskaran and Millane, 1999)。

事前分布は電子密度 f に関する知識を反映するものであることから、中心に高い確率で値を持ち、具体的にどの画素かは分からないが、多くの画素が 0 になると仮定するのが妥当だろう。こうした仮定を満たす事前分布として、LASSO (Tibshirani, 1996; Zhao and Yu, 2006) で用いられたものと同様に指数分布を用いた事前分布を仮定する。具体的には以下のような事前分布を仮定する。

$$p(\mathbf{f}) = \prod_{xy} p(f_{xy}), \quad p(f_{xy}) = \rho_{xy} \exp(-\rho_{xy} f_{xy}), \quad f_{xy} \geq 0, \quad \rho_{xy} \geq 0,$$

定義から分かるように、事前分布では各画素 f_{xy} は独立であるとしている。また、 ρ_{xy} はハイパーパラメータである。このような関数による事前分布を用いることによって、具体的に画素を指定することなく多くの画素が 0 となるが、 ρ_{xy} の値が大きいほど $f_{xy} = 0$ となり易い。我々は周辺のほうが 0 が多くなるように、 ρ_{xy} を x, y の関数として以下のような関数 $\rho(\mu)_{xy}$ と定めた。

$$(3.1) \quad \rho(\mu)_{xy} = \mu w_{xy}, \quad w_{xy} = a \left\{ \left(x - \frac{1+M}{2} \right)^2 + \left(y - \frac{1+M}{2} \right)^2 \right\} + b.$$

a と b は w_{xy} が中心で 0、端で 1 となるように調整するパラメータである (図 6 参照)。ここでは図 6 のように放物面の形状をした事前分布を仮定したが、電子密度に関してより詳細な事前知識があればそれを反映する関数を用いるべきである。なお、(3.1) 式にある μ は事前分布全体の大きさを調整するパラメータである。

回折画像 N を観測したときの電子密度 f の事後分布は、事前分布と尤度に対して以下の関係を持つ。

$$p(\mathbf{f}|N) \propto p(N|\mathbf{f})p(\mathbf{f}).$$

MAP 推定を考えるために (2.3) 式の対数を取り $p(\mathbf{f}|N)$ から f に関連する項を取り出す。その

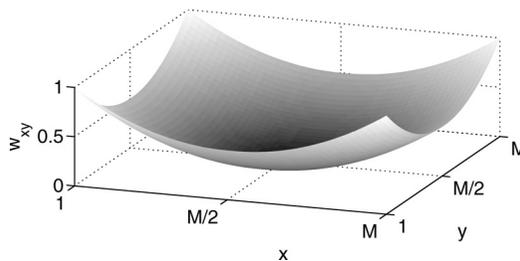


図 6. 事前分布の重み w_{xy} の 3 次元表示。周辺が 0 となり易い。

結果, 以下の関数が得られる.

$$(3.2) \quad \ell_\mu(\mathbf{f}|\mathbf{N}) = \sum_{uv} (N_{uv} \ln |F_{uv}|^2 - |F_{uv}|^2 c_{uv}) - \mu \sum_{xy} w_{xy} f_{xy}.$$

第一項はポアソン分布の尤度に関する項である (同様の考察は他の研究でも見られる (Irwan and Lane, 1998)). 第2項は指数分布による事前分布に関する項である. (3.1)式で導入した μ は2つの項のバランスを調整するパラメータとみなすことができる. SPR法では (3.2)式を最大にする \mathbf{f} を求める. すなわち, 以下の最適化問題を解く.

$$(3.3) \quad \text{maximize } \ell_\mu(\mathbf{f}|\mathbf{N}) \quad \text{subject to } f_{xy} \geq 0 \quad (1 \leq x, y \leq M).$$

一般に, 目的関数 $\ell_\mu(\mathbf{f}|\mathbf{N})$ は \mathbf{f} に関して上に凸な関数ではない. しかし, (3.3)式から, μ をひとつ決めればそれに応じた最適化問題が定まる. 従って, どのように最適値を与える \mathbf{f} を求めるかは別にして, μ をひとつ決めればそれに応じた \mathbf{f} がひとつ決まる. 我々は予め μ を決めることはせず, いくつかの μ に対してこの問題を解くことにする. その中からどの μ に対する電子密度を用いるかは後で議論する.

μ を非常に大きくとれば電子密度 \mathbf{f} のほとんどが0と推定され, 次節で示す勾配法が収束する \mathbf{f} は比較的簡単に求まる. 目的関数は凸でないが, この場合には最大値が求まる可能性が高いと考えられる. そこで最初は μ を大きく設定して問題を解くことで電子密度を推定し, 次に μ を少し小さくし, 前の推定結果を勾配法における初期値として用いて電子密度を推定する. これを繰り返し, 複数の μ に対して電子密度を推定する.

3.2 アルゴリズム

前節で示した (3.3)式を解く方法として勾配法に基づく方法を提案する. まず, 準備のために以下の関係式を示しておく.

$$\frac{\partial |F_{uv}|^2}{\partial f_{xy}} = \frac{2}{M} \text{Re} \left(F_{uv} \exp \left(-\frac{2\pi i (ux + vy)}{M} \right) \right).$$

逆フーリエ変換を以下のように定義する.

$$(\mathcal{F}^{-1}(\mathbf{F}))_{xy} = \frac{1}{M} \sum_{uv} F_{uv} \exp \left(-\frac{2\pi i (ux + vy)}{M} \right),$$

すると $\ell_\mu(\mathbf{f}|\mathbf{N})$ の \mathbf{f} による微分は以下のように書ける.

$$\frac{\partial \ell_\mu(\mathbf{f}|\mathbf{N})}{\partial f_{xy}} = 2 \text{Re}(\mathcal{F}^{-1}(\mathbf{g}(\mathbf{F}; \mathbf{N})))_{xy} - \mu w_{xy},$$

ただし $\mathbf{g}(\mathbf{F}; \mathbf{N}) = \{g_{uv}(F_{uv}; N_{uv})\}$ は以下に定義する.

$$g_{uv}(F_{uv}; N_{uv}) = \left(\frac{N_{uv}}{|F_{uv}|^2} - c_{uv} \right) F_{uv}.$$

$\ell_\mu(\mathbf{f}|\mathbf{N})$ の \mathbf{f} による微分を計算するためには \mathbf{f} のフーリエ変換が必要になり, さらに $\mathbf{g}(\mathbf{F}; \mathbf{N})$ に対して逆フーリエ変換を行なう必要がある. SPR法は勾配法を用いるため, 繰り返し演算が必要になるが, その1ステップの計算量は, 同じく繰り返し演算を必要とするHIO法の1ステップの計算量と同程度である (Fienup, 1982). 電子密度 f_{xy} , $x, y = 1, \dots, M$ の勾配法に基づく更新則は以下のように書ける.

$$(3.4) \quad f_{xy}^{t+1} = \max \left(0, f_{xy}^t + \eta_t \frac{\partial \ell_\mu(\mathbf{f}^t|\mathbf{N})}{\partial f_{xy}^t} \right),$$

ここで t は繰り返しの回数を示し, \max 操作によって f_{xy} は非負となる. 正の定数 η_t がステップ幅を示すが, このステップ幅を直線探索によって変化させることによって収束速度が飛躍的に向上する.

一般に勾配法では初期値が重要である. 最初 μ を非常に大きな値にとれば, 多くの f_{xy} が 0 となり, 初期値によらず解が比較的簡単に求まる. そうして求めた解を μ を少し小さくした問題の初期値とする. このように, 複数の μ に対して解を求める.

3.3 既存法のベイズ的解釈

既存法では (2.5) 式の最適化問題を解いている. この最適化問題は形式的には SPR 法と同様に MAP 推定と考えることができる.

以下のように尤度と事前分布を定義する.

$$p(\mathbf{N}|\mathbf{f}) \propto \exp\left(-A \sum_{uv} (|F_{uv}|c_{uv}^{1/2} - N_{uv}^{1/2})^2\right), \quad p(f_{xy}) = \begin{cases} 1/C, & (x,y) \in \gamma \\ 0, & (x,y) \notin \gamma, \end{cases}$$

上式で A, C は正の定数である. 尤度は指数の肩に 2 乗誤差が乗った正規分布の形をしているが $|F_{uv}|$ が非負であるために規格化定数は正規分布のものとは異なる. また, 事前分布は一様分布としているが, 各 f_{xy} は $0 \leq f_{xy} \leq C$ に制限されると仮定している. ただし C は十分大きい正の実数とする.

こうして定義した尤度と事前分布を考えると, 事後確率の対数をとったものは (2.5) 式の目的関数と一致する. したがって, 既存法も形式的には MAP 推定としてとらえることができる.

既存法では γ を設定しなければならないが, これをあらかじめ知ることは一般に難しい. SPR 法では γ を設定する必要はない. ただし, μ という新たなパラメータを調整する必要がある.

3.4 数値実験

本節では図 3 に示した回折画像に対して SPR 法を適用した結果を示す. 図 7 は μ を変化させて復元したタンパク質分子の電子密度を示したものである. 大きい μ に対しては多くの画素で電子密度が 0 と推定されていることがわかる. ある程度小さい μ に対しては図 2 左にある電子密度を十分に復元できる.

ここで問題となるのは μ の選択である. μ を小さくすると, 一般に尤度は大きくなる. 適切な μ を求めるためには, 尤度ではない何らかの規準によって最適値を与える μ を選ぶ必要がある. もし, 同じ条件で測定したデータが複数あるのならば, cross validation などの方法で μ を選ぶことができるが, 位相復元の問題では同じ条件で測定したデータはひとつしかない

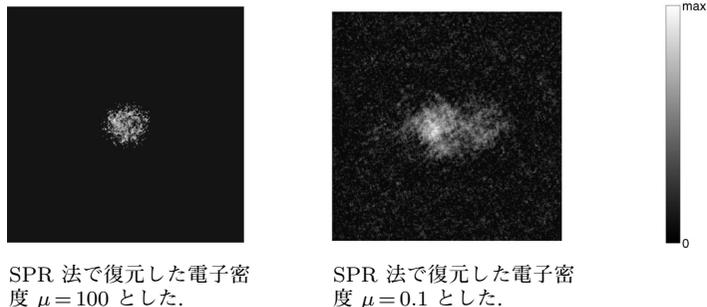


図 7. SPR 法によって復元された 2 次元電子密度.

め、他の方法を考える必要がある。

この点は LASSO を用いたデータ解析でも問題となる。Zou らは LASSO では、0 でないパラメータの個数が確率モデルの自由度 (degree of freedom) の推定量となることを示している (Zou et al., 2007)。そして μ を選択するときに自由度を罰則項に含むモデル選択基準、Akaike information criterion (AIC) や Bayesian information criterion (BIC) などを使うことを提案している。SPR 法においても、モデルの自由度を推定し、モデル選択基準によって μ を求めることが考えられる。しかし、ポアソン分布を尤度としている SPR 法では、0 でないパラメータの個数は自由度の推定値とはならない。自由度の推定法については今後の課題である。

光学の分野では伝統的に規格化された 2 乗誤差を規準にしている。SPR 法でも 2 乗誤差を基準として μ を選ぶことにする。具体的には、ある μ に対して SPR 法によって $\ell_\mu(f|N)$ を最大化して求めた電子密度を $\hat{f}(\mu)$ とし、以下の関数によって誤差を定義する。

$$(3.5) \quad \text{Error}_F = \frac{\sum_{uv} (|\hat{F}(\mu)_{uv}|c_{uv}^{1/2} - N_{uv}^{1/2})^2}{\sum_{u'v'} N_{u'v'}}$$

ここで $\hat{F}(\mu)_{uv}$ は $\hat{F}(\mu)_{uv} = \mathcal{F}(\hat{f}(\mu))_{uv}$ とした。 $\text{Error}_F = 0$ となるのなら、 $|\hat{F}(\mu)_{uv}|c_{uv}^{1/2} = N_{uv}^{1/2}$ であり、尤度は最大になる。しかし、 Error_F は SPR 法の尤度とは異なるため、 μ を小さくしたとしても Error_F は必ずしも減少しない。厳密な議論ではないが、 μ を非常に大きな値から小さくしていくと、 Error_F は徐々に小さくなり、どこかで増加に転じる。そこで Error_F を最も小さくする μ を選ぶことにする。図 3 の回折画像に対して μ を変化させ、それぞれに対する Error_F を表示したものが図 8 である。 μ は 1×10^4 から 1×10^{-2} まで変化させた。 Error_F を最も小さくする μ は $\mu = 0.1$ であった。なお、このときの復元された電子密度画像は図 7 の右、 Error_F の値は 0.209 であった。同じデータに対して HIO 法を用いて電子密度を復元したとき (ただし、 f_{xy} が正となりうる領域 γ は 159×159 画素の正方形の領域とした。) の Error_F は 0.220 であった。

最後に計算速度について述べる。SPR 法は Intel Xeon X5570 (2.93 GHz 4 cores) を搭載したデスクトップ機に C 言語で実装した。上記の実験では、全ての μ に対して 10 分以内で解が求まった。

4. まとめ

本稿では、我々が提案した位相回復の解法を解説した。提案する SPR 法はベイズ統計の MAP 推定に基づく方法であり、今後 XFEL によって得られる生体単粒子などの回折画像からの位

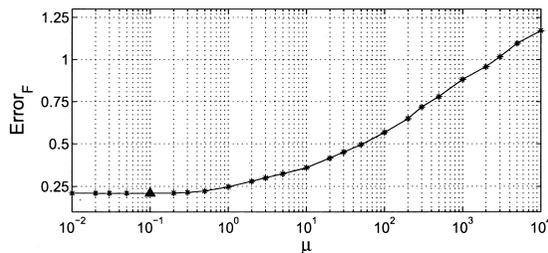


図 8. 図 3 の回折画像に対して μ を変化させて SPR 法を行った結果得られた Error_F を示したもの。黒い三角形は最小の Error_F に示した。横軸 μ は左から右に増加するが、実験を行なう際には大きい μ から始めて小さくしていった。

相回復に対して有効な方法であると考えられる。

ベイズ統計の立場から考えると、SPR 法も既存法も MAP 推定とみなせる。ただし、二つの方法では尤度と事前分布、それぞれの分布が異なる。光子の数が十分に多い場合、SPR 法と既存法による電子密度の復元結果に大きな差は見られなかった (Ikeda and Kono, 2012) が、光子の数が少ない場合、復元結果には差が見られた。以下ではその差について簡単に考察する。

まず、尤度について考える。単粒子の回折画像では中心およびその周辺に大部分の光子が観測される。回折画像の中心部は電子密度の空間周波数を考えると低い周波に対応し、単粒子のおおまかな形を表現している。SPR 法では光子数 N_{uv} をポアソン分布として記述しているため、光子数の多い画素の方が尤度に大きく寄与し、回折画像の中心部分を重視することになる。したがって電子密度の推定値は、単粒子のおおまかな形を反映したものになると考えられる。一方、既存法は回折画像のどの画素も等価に扱っており、光子の観測されなかった周辺の画素の影響も中心と同じ重みとなる。このため、電子密度の空間周波数を考えると高周波のノイズの影響を受け易いと考えられる。

次に事前分布について考える。光子数の少ない場合、観測された回折画像は理想とは程遠いものとなる。既存法では γ 領域内での f で回折画像を表現しなければならないが、SPR 法では γ を設定しないため、正の値をもつ f_{xy} は領域の全体に広がる。 μ を適切に設定すれば、観測された回折画像をうまく表現できる可能性がある。また、HIO 法を用いる場合、 γ 上で定義される f によって回折画像が説明できなければ収束性が保証されない。一方、SPR 法は局所最適解しか得られない可能性はあるが、収束はする。計算量も現実的な範囲であり、既存の HIO 法よりも小さな誤差を持つ解を与えることが確認できた。

XFEL による生体単粒子の回折画像からの 3 次元構造解析には、この位相回復だけではなく、いくつかの問題を克服する必要がある。回折画像はある方向からの観測だけである。3 次元構造を知るためには数多くの画像を取り、それらを正しく組み合わせて 3 次元の回折像を構築して位相回復を行うか、2 次元回折像から 2 次元の実像を求めてそれらを 3 次元再構成する必要がある。輝度が低い画像を組み合わせる方法は未だ十分には検討されておらず、今後の研究が必要である。

注.

- 1) 独立法人理化学研究所播磨研究所 X 線自由電子レーザー, <http://xfel.riken.jp> を参照.
- 2) 分子のサイズが大きい場合にはより鮮明な画像が得られるが、生体単粒子の回折画像はこの程度の光子数だと推定される.
- 3) 「計測と制御」2011 年 5 月号にミニ特集「回折イメージング～位相回復の新展開～」がある.

謝 辞

本研究の一部は文部科学省の X 線自由電子レーザー利用推進研究課題の事業として行なわれました。

参 考 文 献

- Baskaran, S. and Millane, R. P. (1999). Bayesian image reconstruction from partial image and aliased spectral intensity data, *IEEE Transaction on Image Processing*, **8**, 1420-1434.

- Bauschke, H. H., Combettes, P. L. and Luke, D. R. (2002). Phase retrieval, error reduction algorithm and Fienup variants: A view from convex optimization, *Journal of the Optical Society of America A*, **19**, 1334–1345.
- Fienup, J. (1978). Reconstruction of an object from the modulus of its Fourier transform, *Optics Letters*, **3**, 27–29.
- Fienup, J. (1982). Phase retrieval algorithms: A comparison, *Applied Optics*, **21**, 2758–2769.
- Hau-Riege, S. P., London, R. A., Huldt, G. and Chapman, H. N. (2005). Pulse requirements for x-ray diffraction imaging of single biological molecules, *Physical Review E*, **71**, 061919.
- Huldt, G., Szőke, A. and Hajdu, J. (2003). Diffraction imaging of single particles and biomolecules, *Journal of Structural Biology*, **144**, 219–227.
- Ikeda, S. and Kono, H. (2012). Phase retrieval from single biomolecule diffraction pattern, *Optics Express*, **20**, 3375–3387.
- Irwan, R. and Lane, R. G. (1998). Phase retrieval with prior information, *Journal of the Optical Society of America A*, **15**, 2302–2311.
- Miao, J., Sayre, D. and Chapman, H. (1998). Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects, *Journal of the Optical Society of America A*, **15**, 1662–1669.
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. and Hajdu, J. (2000). Potential for biomolecular imaging with femtosecond X-ray pulses, *Nature*, **406**, 752–757.
- Seibert, M. M. et al. (2011). Single mimivirus particles intercepted and imaged with an X-ray laser, *Nature*, **470**, 78–81.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.
- Zhao, P. and Yu, B. (2006). On model selection consistency of lasso, *Journal of Machine Learning Research*, **7**, 2541–2563.
- Zou, H., Hastie, T. and Tibshirani, R. (2007). On the “degrees of freedom” of the lasso, *Annals of Statistics*, **35**, 2173–2192.

Molecular Electron Density Estimation with X-ray Free Electron Laser

Shiro Ikeda¹ and Hidetoshi Kono²

¹The Institute of Statistical Mathematics

²Japan Atomic Energy Agency

New laser equipment x-ray free electron laser facilities are built in the United States, Europe, Japan, and Korea. A variety of new experiments in physics, chemistry, and biology are being carried out with this new X-ray laser, and new data will be collected. We explain one such experiment, electron density estimation from a single biological particle, and show our recent result on phase retrieval, which recovers the two-dimensional electron density from a two-dimensional X-ray diffraction pattern.