1B4-OS-01a-4in1

時間分解分光計測のベイズ的最適設計

Bayesian-optimal design for time-resolved spectroscopic measurement

徳田悟 [*1] 永田賢二 [*1*2] 岡田真人 [*1]
Satoru Tokuda Kenji Nagata Masato Okada

[*1]東京大学 大学院新領域創成科学研究科 [*2]JST さきがけ研究員
Graduate School of Frontier Sciences, The University of Tokyo JST PRESTO

Microscopic dynamics have recently been captured by time-resolved spectroscopy. The peaks of measured data in each frame time exhibit the information of electronic state at the time. However improving frame rate causes quantum noise, which prevent us from identifying the peaks, for each frame. In this paper, we introduce a Bayesian framework, which provides a statistical peak separation and enables us to design an optimal frame rate.

1. 序論

ポンプ-プローブ分光法を初めとした時間分解分光法の発展により，ヘモグロビンの濃度変化や太陽電池の光触媒反応といった微小な対象のダイナミクスをフェムト秒オーダーで計測可能な時代が到来している．各フレーム時間で測定されたスペクトルのピークはその時間での電子状態を示す．しかし，微小時間での測定には量子雑音が顕著に現れ，各フレーム時間でのスペクトルからピークを判別することは困難となる．ベイズ推定に基づくピークの判別および分離は有効な解決策となる [永田 14, Mazet 15]．さらに，永田らは X 線光電子分光法を例にとり，事後分布の変化から，適切な推定結果を得るために必要なフレームあたりの測定時間の設計を報告している．本稿では彼らの枠組みにベイズ比熱 [徳田 15] を導入し，より精密に測定時間を設計する理論的枠組みを再構築する．人工データを用いたシミュレーションを行い，枠組みの有効性を示す．

2. 枠組み

単位時間あたりの測定で得られる理想的なスペクトル線形を次のように表す．

$$f(x;w) = F(x;w_F) + B(x;w_B) \tag{1}$$

ここで，$w = \{w_F, w_B\}$ である．$F(x;w_F)$ は興味のある物理過程からなるフォアグラウンド，$B(x;w_F)$ はそれ以外の物理過程からなるバックグラウンドをそれぞれ表す．本稿では固体表面の電子状態を対象とする X 線光電子分光法の場合を例とし，次のようにモデル化する．

$$F(x;w_F) = \sum_{k=1}^{K} a_k \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right) \tag{2}$$

$$B(x;w_B) = b\int_x^{\infty} dx' F(x';w_F) + c \tag{3}$$

ここで，$w_F = \{a_k, \mu_k, \sigma_k\}_{k=1}^{K}$，$w_B = \{b, c\}$ である．各バンドについて，強度 a_k は原子や分子の数密度，中心 μ_k はエネルギー準位，幅 σ_k は温度や寿命をそれぞれ表す．$B(x;w_B)$ はバルクにおける光電子の非弾性散乱を表す．光子の離散性から測定されるスペクトル y は次のようなポアソン分布に従う．

$$p(y \mid x; w, T) = \frac{(Tf(x;w))^y \exp(-Tf(x;w))}{y!} \tag{4}$$

ここで，T は測定時間である．信号雑音比は $\sqrt{Tf(x;w)}$ となる．[渡辺 12] に基づき，次のように事後分布を定義する．

$$p(w \mid D; \beta, T) = \frac{1}{Z_n(\beta)} \prod_{i=1}^{n} p(Y_i \mid X_i; w, T)^{\beta} \varphi(w) \tag{5}$$

ここで，$D = \{X_i, Y_i\}_{i=1}^{n}$ である．β は逆温度，$\varphi(w)$ は事前分布，$Z_n(\beta)$ は規格化定数を表す．$\beta = \infty$ のとき最尤推定，$\beta = 1$ のときベイズ推定に対応する．$Z_n(1) = p(D \mid K)$ を周辺尤度と呼び，この値を最大化するピーク数 K がベイズ最適な解とする．ベイズ比熱が次のように定義される [徳田 15]．

$$C_n(\beta) = \frac{\partial \langle nL_n(w)\rangle_\beta}{\partial \beta^{-1}} \tag{6}$$

ここで，$L_n(w) = -(\sum_{i=1}^{n} \log p(Y_i \mid X_i; w, T))/n$ である．ベイズ比熱のスケーリング関数を次のように定義する．

$$\tilde{C}(nT\beta) = (nT\beta)^2 \left(\langle \tilde{L}_{n,T}(w)^2\rangle_\beta - \langle \tilde{L}_{n,T}(w)\rangle_\beta^2\right) \tag{7}$$

ここで，$\tilde{L}_{n,T}(w) = L_n(w)/T$ である．$T \gg 1$ かつ $n \gg 1$ のとき，[徳田 15] の場合に帰着し，$C_n(\beta) = \tilde{C}(nT\beta)$ となる．

3. シミュレーション

測定時間が異なるスペクトルの人工データをそれぞれ作成した．真のピーク数が $K_0 = 2$，サンプル数は $n = 101$，測定時間は $T = 10^m$，m は -2 から 6 までの 9 通りの整数とした．単位時間あたりの真のスペクトルの線形は図 1(B) の通り．交換モンテカルロ法 [Hukushima 96] を用いて次の同時分布から $\{w_l\}_{l=1}^{L}$ をサンプリングした．

$$p(w \mid D; \{\beta_l\}_{l=1}^{L}, T) = \prod_{l=1}^{L} p(w \mid D; \beta_l, T) \tag{8}$$

ここで，$L = 400$ である．$\beta_1 = 0$ とし，β_l $(l > 2)$ は区間 $[10^{-10}, 1]$ で対数的に等間隔に設定した．$K = 0, 1, 2, 3$ の各モ

連絡先: 岡田真人，okada@k.u-tokyo.ac.jp

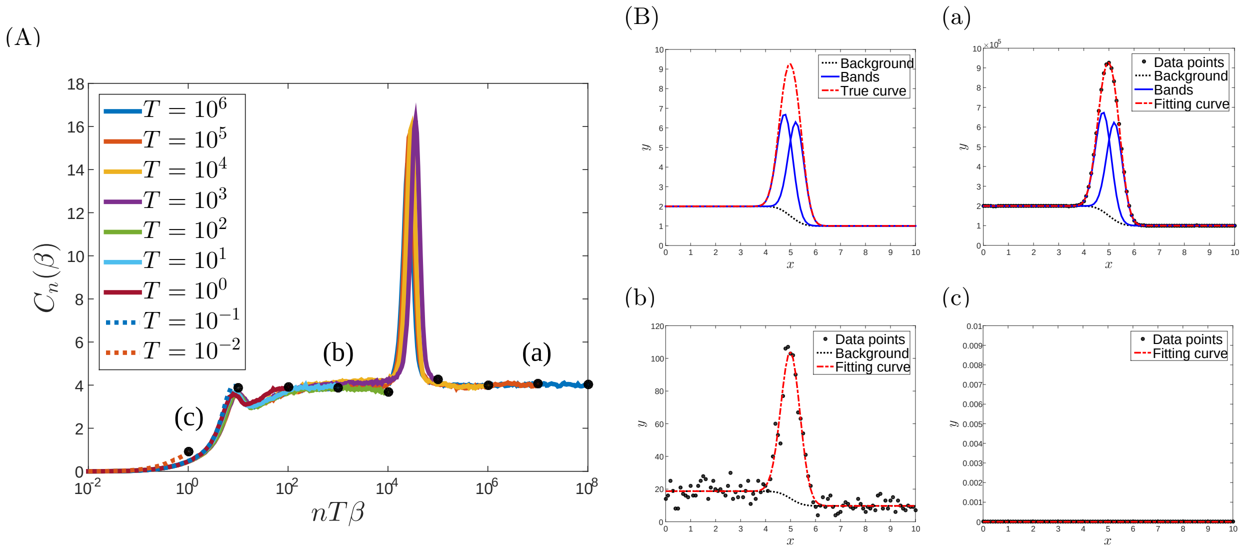

図 1: (A)$K = 2$ のベイズ比熱．横軸は $nT\beta$ を対数表示している．縦軸は $n = 101, T = 10^m$ でのベイズ比熱の各値を表している．m の値と線種の対応は凡例の通り．黒点は各 m における $\beta = 1$ の点をプロットしている．(B) 単位時間あたりの真のスペクトルの線形．黒の点線はバックグラウンド，青の実線はフォアグラウンドの各バンド，赤の一点鎖線は回帰曲線をそれぞれ表す．(a-c) ピーク分離の結果．(a)$T = 10^5$，(b)$T = 10^1$，(c)$T = 10^{-2}$ においてそれぞれ $K = 2$，$K = 1$，$K = 0$ のモデルがベイズ最適となっている．各モデルの最大事後確率推定による回帰をそれぞれ表している．黒点は人工データを表し，その他は (B) に準ずる．

デルを用い，それぞれ以上の手続きを行った．ただし，$K = 0$ のとき $f(x; w) = c$ とした．各人工データについてベイズ比熱を計算し，ベイズ最適な解を以てピーク分離を行った．

図 1(A) に $K = 2$ のベイズ比熱を示している．測定時間 T に依らず，横軸を $nT\beta$ として各曲線の概形が一致している．これは式 (7) の有限サイズスケーリングが成立していることを裏付ける．$nT\beta = 10^1, 10^4$ 付近の臨界点を境に，ベイズ比熱は異なる値のプラトー領域をそれぞれ示している．[徳田 15] で議論したように，統計物理との対応から，臨界点を跨ぐ変化を相転移，各プラトー領域を異なる相として解釈できる．異なる相である (a)$T = 10^5$，(b)$T = 10^1$，(c)$T = 10^{-2}$ のピーク分離結果をそれぞれ図 1(a-c) に示している．表 4. に示すように，(a) は $\hat{K} = 2$，(b) は $\hat{K} = 1$，(c) は $\hat{K} = 0$ がそれぞれベイズ最適解となっている．測定時間に応じて相転移が生じていることがベイズ最適なモデルの変化として反映されている．測定時間が不十分であると，$\hat{K} = K_0 = 2$ と適切に推定されていないとも言える．$T = 10^{-1}$ について $\hat{K} = 2$ と推定されているが，これは臨界点付近では事後分布が複雑な構造を取り，推定結果が不安定になるためと考えられる．今回は，安定して適切な推定結果を得るために必要な測定時間が比熱曲線のプラトー領域に差し掛かる $T = 10^5$ 程度であることが示唆された．

4. 結論と議論

本稿において，我々はスペクトルの適切なピーク分離に必要な測定時間をベイズ比熱を指標に設計する枠組みを構築した．例として X 線光電子分光法の場合を想定し，人工データを用いたシミュレーションを行うことで枠組みの有効性を示した．

本枠組みを運用する上で，予め十分に長い時間測定したデータが手元にないと，ベイズ比熱の示す臨界点や相を把握できない，すなわち測定時間の設計を行えないことに注意なければならない．例えば，我々は図 1(A) における $T = 10^6$ の比熱曲線から必要な測定時間を割り出したのであり，$T = 10^{-2}$ の比熱曲線からはそれを知ることはできない．さらには，スペクトルの真の線形を知らない立場では，$T = 10^6$ の結果ですらも，十分に比熱曲線の概形を表しているかはわからない．これに対し，取るべき方策は二つある．一つは第一原理計算のような演繹計算によりスペクトルの真の線形を予測した上で，本稿のシミュレーションを行うことである．ある種の「答え」を知っている立場に立つのである．もう一つは，同じ物理系について，予備実験として十分に長い時間測定することである．実験コストの短縮ではなく，あくまで時間分解分光を用いて動的な情報を最大限に引き出す手だてとしては有用であり，多くの分光計測が再現性のある観測であることと親和性が高い．

表 1: 測定時間に対するベイズ最適解の変化

T	10^{-2}	10^{-1}	10^0	10^1	10^2	10^3	10^4	10^5	10^6
$\hat{K}$	0	2	1	1	1	2	2	2	2

参考文献

[永田 14] 永田賢二，村岡怜，佐々木岳彦，岡田真人，信学技報，Vol.113, No.500, 115-120 (2014).

[Mazet 15] Mazet, V., Faisan, S., Awali, S., Gaveau, M-A., Poisson, L., *Signal Processing*, Vol.109, 193-205 (2015).

[徳田 15] 徳田悟，永田賢二，岡田真人，*JSAI 2015*, 2F1-5in, 1-2 (2015).

[渡辺 12] 渡辺澄夫:ベイズ統計の理論と方法，コロナ社 (2012).

[Hukushima 96] Hukushima, K. and Nemoto K., *J. Phys. Soc. Jpn.*, Vol.65, No.6, 1604-1608 (1996).