

混沌系工学特論 配布資料 #2

担当：井上 純一 (情報科学研究科棟 8-13)

URL : http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/

平成 17 年 10 月 24 日

目次

2	統計的手法に基づくデジタル画像の復元	17
2.1	原画像と劣化過程の「小磁石」による表現	18
2.2	事前情報の導入と事後確率の計算	20
2.3	最大事後確率推定と有限温度推定	21
2.4	確率モデルのパラメータ推定	22
2.4.1	周辺尤度関数最大化法	23
2.4.2	EM アルゴリズムによる周辺尤度の最大化	23
2.4.3	マルコフ連鎖モンテカルロ法の適用	24
2.4.4	デジタル画像復元の実行例	25
2.5	平均場アルゴリズムによる計算量の低減	26

2 統計的手法に基づくデジタル画像の復元

前回は簡単な最適化問題に対して、ノイズを利用したアルゴリズムを提示し、このアルゴリズムを同一のエネルギー関数 (同一の力学的性質) を持つ複数のシステムに対して動作させた場合、アンサンブルという考え方により、システムが各エネルギーレベルをとる確率を求める手続きを学んだ。このとき、ノイズの大きさをうまく制御することにより、十分な実行時間の後には確率 1 で最小エネルギー状態を求めることができることも見た。さらに、このとき得られる各状態の出現確率で物理量の期待値を計算することで、注目するシステムの物理量の長時間平均を間接的に計算できることも学んだ。ここで、得られる値は該当する物理量の実験室での観測値に相当する。これがこの講義で扱う統計力学の方法であった。

ここからは、データをノイズを含む通信路を介して転送する場合、もとのデータをいかに復元するか、という問題を考えてみよう。データとしては 2 次元デジタル画像を用い、劣化した画像から原画像を復元する状況を考える。この画像復元の問題においても、前回学んだ最適化手法である、[ノイズを用いたアルゴリズム] を基礎とするシミュレーテッド・アニーリングを用いることができる。このことは画像を復元する問題においても、やはり「エネルギー関数」が存在することを意味する。ここでは、まず、システムを「小磁石」で表すことからはじめ、ベイズ統計を介した最大事後確率推定という原画像データの推定法の下では、問題が最適化問題として定式化されることを示す。また、画像およびデータの劣化過程を確率モデルで表現することになるが、この分布を最尤推定で特定するための統計的手法である EM アルゴリズムについても触れる。また、画像の修復アルゴリズムを平均場近似という考え方をを用いて実時間で解く手続きについても詳しく見て行くことにする。これらの手法/アルゴリズムがどのように統計力学/熱力学と関係し、具体的にどのように用いられるのかを習得することがこの節の目標である。

2.1 原画像と劣化過程の「小磁石」による表現

本節で扱う 2 次元のデジタル画像は例えば次のような「白」と「黒」のピクセル (画素) からなるものであるとする。さて、これらのデジタル画像を Xemacs 等のテキストエディタを用いて開いてみると



図 10: ここで扱うデジタル画像の例. 原画像 (左) と劣化画像 (右).

```
P1
200 200
111111111111110000
111111111111111111
.....
.....
```

のようになっている。このファイルの先頭の P1 はこのファイルが白黒の 2 値画像であることを表すタグであり、次の 200 200 はこの画像のサイズ、つまり、図 10 が縦 200, 横 200 ピクセル (画素) の画像であることを示している。また、この次の行から始まる 11110000111111 の並びが画像の本体を表している。

このファイルに関するより詳しい取り扱いの説明に関しては、例えば

http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/LEC/ENS/inf_ens_2005.html

での情報工学演習 I(A) 「ファイル操作と gnuplot 及び GIMP の使い方に関する演習」のテキストを参照すると良いであろう。

もちろん、この画像は 2 値画像なので、この数の要素は 1 か 0 のみである。さて、この各要素の「番地」を指定するにはいくつかやり方があるが、ここでは各画素の番地の決め方を「右上から左したに順を追って付ける」という約束にし、先頭の 1 から数えて、 i 番目の画素の値を ξ_i で表現することにする。従って、この画素の値が 1 ならば

$$\xi_i = 1 \text{ (黒)} \tag{37}$$

であるし、逆に 0 ならば

$$\xi_i = 0 \text{ (白)} \tag{38}$$

というわけである。ところで、このデジタル画像を「小磁石」の塊としてみる場合、画素の値を 0, 1 であるとするのではなく、「小磁石」の N 極と S 極に対応した 1 か -1 とする方が何かと都合がよい。つまり、

$$\xi_i = \begin{cases} +1 \text{ (黒)} \\ -1 \text{ (白)} \end{cases} \tag{39}$$

とする。これは全ての画素で $\xi_i \rightarrow 2\xi_i - 1$ なる変換を行えば良いだけの話ではあるが、この講義では式 (39) の表現を用いることにしよう。

さて、この画像がノイズを含む通信路を通過し、各画素が独立に確率 p で反転する状況を考える。つまり、

$$\xi_i \rightarrow \begin{cases} -\xi_i & \text{確率: } p \\ +\xi_i & \text{確率: } 1-p \end{cases} \tag{40}$$

とする。この各ピクセルの劣化過程をグラフを用いて描くと図 11 のようなグラフとなる。これはまた、原

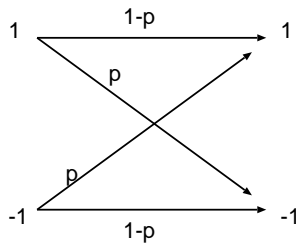


図 11: 2 元対称通信路.

画像 $\{\xi\} \equiv (\xi_1, \xi_2, \dots, \xi_N)$ に対する劣化画像を $\{\tau\} \equiv (\tau_1, \tau_2, \dots, \tau_N)$ とすれば、条件付き確率を用いて

$$P(\tau_i|\xi_i) = \frac{e^{\beta_\tau \tau_i \xi_i}}{2 \cosh(\beta_\tau)} \tag{41}$$

と表すことができることにも注意しよう。(40) 式の p , (41) 式の β_τ の間には

$$p = \frac{e^{-\beta_\tau}}{2 \cosh(\beta_\tau)}, \quad 1-p = \frac{e^{\beta_\tau}}{2 \cosh(\beta_\tau)} \tag{42}$$

つまり、

$$\beta_\tau = \frac{1}{2} \log \left(\frac{1-p}{p} \right) \tag{43}$$

が成り立つ。このような統計的性質を持つ通信路を情報理論の講義で既に学んだように 2 元対称通信路 (Binary Symmetric Channel : BSC) と呼ぶ。

各画素が独立にノイズからの影響を受けるとすれば、原画像全体 $\{\xi\}$ は劣化画像全体 $\{\tau\}$ へ

$$P(\{\tau\}|\{\xi\}) = P(\tau_1|\xi_1) \times P(\tau_2|\xi_2) \times \dots \times P(\tau_N|\xi_N) = \prod_{i=1}^N P(\tau_i|\xi_i) = \frac{e^{\beta_\tau \sum_{i=1}^N \tau_i \xi_i}}{[2 \cosh(\beta_\tau)]^N} \tag{44}$$

の確率で変化する。以上が原画像及び劣化過程の「小磁石」を用いた表現方法である。

2.2 事前情報の導入と事後確率の計算

前節でデジタル原画像と劣化過程を「小磁石」— スピン — の言葉で表した. この確率システム — スピン系 — に対し, ここからの問題は劣化画像 $\{\tau\}$ から原画像 $\{\xi\}$ を言い当てることである. もう少し明示的に書くと, 原画像 $\{\xi\}$ の推定値を $\{\sigma\} \equiv (\sigma_1, \sigma_2, \dots, \sigma_N)$ とすると, 我々はできるだけ $\{\sigma\}$ が $\{\xi\}$ に近くなるような戦略を選ぶべきであると言える. この目標下では当然 $\{\sigma\} = \{\xi\}$, つまり

$$\sigma_i = \xi_i \quad (\forall i) \quad (45)$$

のとき, デジタル原画像の完全な復元ができたことになる.

さて, ここで条件付き確率 $P(\{\sigma\}|\{\xi\})$ を求めることができれば, 画像を修復する上で有用であろう. この条件付き確率は, 劣化画像 $\{\tau\}$ を観測した条件下での原画像の推定値が $\{\sigma\}$ である確率である. 今の場合, この確率自体は直接手に入らないが, ベイズの公式 (Bayes formula) と呼ばれる関係式:

$$P(\{\sigma\}|\{\tau\}) = \frac{P(\{\tau\}|\{\sigma\})P(\{\sigma\})}{\sum_{\sigma} P(\{\tau\}|\{\sigma\})P(\{\sigma\})} \quad (46)$$

を用いることにより, 間接的に計算される. ここで, $\sum_{\sigma}(\dots) \equiv \sum_{\sigma_1=\pm 1} \sum_{\sigma_2=\pm 1} \dots \sum_{\sigma_N=\pm 1}(\dots)$ と定義してあり, 以降では断りの無い限り全てこの定義に従うものとする.

(46) 式の中に含まれる, $P(\{\tau\}|\{\sigma\})$ は尤度 (likelihood) と呼ばれる確率であり, 「通信路の確率モデル」を表している. 我々は通信路に関する統計的性質を事前に知ることはできないが, $0 \leq q \leq 1$ なるパラメータ q を用いて, 各画素 i に関して

$$P(\tau_i = -\sigma_i|\sigma_i) = q, \quad P(\tau_i = \sigma_i|\sigma_i) = 1 - q \quad (47)$$

と仮定することは自然であろう. 従って, (41)(43) 式で見たように

$$h = \frac{1}{2} \log \left(\frac{1-q}{q} \right) \quad (48)$$

を満たすような h を用いることにより, 通信路の確率モデルである尤度 $P(\{\tau\}|\{\sigma\})$ は

$$P(\{\tau\}|\{\sigma\}) = \frac{e^{h \sum_{i=1}^N \tau_i \sigma_i}}{[2 \cosh(h)]^N} \quad (49)$$

のように一般的に書くことができる.

一方, $P(\{\sigma\})$ は事前分布 (prior distribution) と呼ばれる「原画像の確率モデル」である. 原画像がどのような統計的性質を持つのか, 我々は知る由もないが, 例えば, 図 10 のような画像を見ると, 画像の局所的には同じ色の画素を取っていることが見て取れるであろう. これは図 10 にかかわらず, どの種類の画像であっても, 多かれ少なかれこのような一般的性質を持っていることが予想される. 従って, 次のように事前分布を選ぶことができる.

$$P(\{\sigma\}) = \frac{e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j}}{\sum_{\sigma} e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j}} \quad (50)$$

ここで, $J > 0$ であり, 上記確率分布に含まれる和 $\sum_{\langle ij \rangle}(\dots)$ は近接する画素についてのみ取られる. 従って, 上の確率分布からわかることは隣り合う画素対 σ_i, σ_j が同じ色を取るような画像 $\{\sigma\}$ が相対的に高確率で現れるということである.

もちろん, 実際に我々が修復する画像はこの性質を持っていないかもしれないし, 原画像としてより特殊な (あるいは特定の) 画像を想定して, (50) 式よりさらに複雑な形を持つ確率分布を事前分布とすべきかも

しれない。しかし、いずれにしても我々は原画像に関しては何も知りえないわけである。従って、特殊で複雑な事前分布を用いるよりも、どの画像についても最小限あてはまりそうな自由度を持つ (50) を選ぶことが得策であろうと、ここでは考えるわけである。

この事前確率 (50) からのスナップショットはパラメータ J の値に依存することに注意しよう。このパラメータ J の逆数をエネルギー関数が $E = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j$ で与えられる磁石 — 強磁性イジングモデル — が置かれた温度: $T = J^{-1}$ として定義すると、この温度を $T = 1.15, 2.15, 3, 15$ と変化させると、それぞれの温度での分布 (50) からのスナップショットは図 12 のようになる。 T は温度であったので、 T の値が大きな場合には、隣接する小磁石が同じ方向に揃おうとする影響より、熱によりバラバラな向きを取ろうとする影響の方が大きく、図 12 の右側の図を見てみると白黒のまだらな模様が現れることがわかる。逆に、温度 T が小さいと、今度は隣接する小磁石が同一の方向に揃う影響が強くなり、図 12 の左側の図のように同じ色を持つ大きな塊 (クラスター) を持つような画像が現れる。

ここで、我々は図 10 の「印鑑」のような 2 値デジタル画像を図 12 で与えられる複数枚のスナップショット画像の「張り合わせ」で表現することを試みる。このとき、どの温度 $T = J^{-1}$ でのスナップショットを用いれば望ましいような張り合わせを作ることができるか、は自明ではないので、 J の値自体を劣化画像から特定しなければならない。このための手続きについては後に詳しく見て行くことになる (EM アルゴリズムの節参照)。さて、(49)(50) 式を (46) に代入すれば事後確率 (posterior distribution) が得られる。

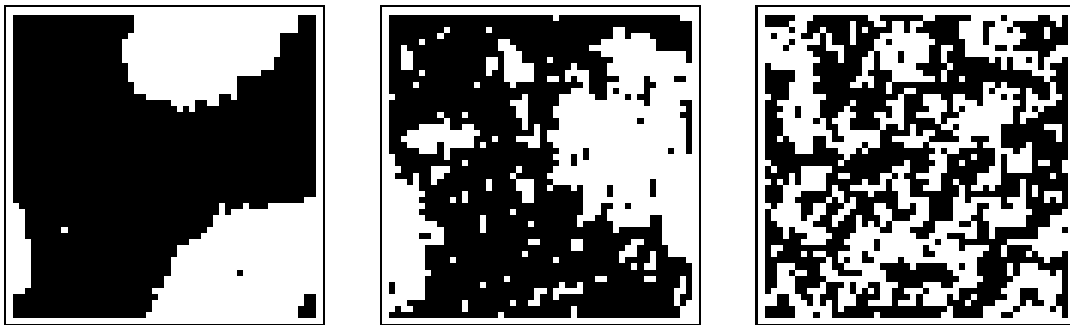


図 12: ここで用いる事前分布からのスナップショット。デジタル画像のサイズは 50×50 。左から $T = 1.15, 2.15, 3.15$ に選んである。

$$P(\{\sigma\}|\{\tau\}) = \frac{e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i}}{\sum_{\sigma} e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i}} \quad (51)$$

この分布の形を良くみると、エネルギー関数を

$$E(\{\sigma\}) = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - h \sum_i \tau_i \sigma_i \quad (52)$$

とした場合、前回学んだボルツマン分布の形をしている。従って、アンサンブルの考え方にに基づき、この分布での物理量の平均を計算することにより、例えば内部エネルギーや比熱などの値が得られる。しかし、ここでの我々の目標は、それらの物理量を求めることと言うよりもむしろ、この事後分布を用いてどのように原画像を推定するのか、ということにある。以下では、この事後確率の「使い方」を見て行くことにする。

2.3 最大事後確率推定と有限温度推定

前節で事後分布が得られたわけであるが、この事前分布を原画像の推定にどのように使うのか、が次の問題になるであろう。この目的の下に最も素朴には、事後確率 (51) を最大化するような $\{\sigma\} = (\sigma_1, \sigma_2, \dots, \sigma_N)$

を原画像の推定値に選ぶ戦略である。これを最大事後確率推定 (Maximum A Posteriori 推定: MAP 推定) と呼ぶ。

事後確率 (51) を最大化することは次に定義されるエネルギー関数:

$$E(\{\sigma\}) = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - h \sum_i \tau_i \sigma_i \quad (53)$$

を最小化する操作に等しい。従って、この組合せ最適化問題はシミュレーテッド・アニーリング (simulated annealing) 等の最適化ツールにより解くことができる。

さて、これとは別に、事後確率 (51) を着目する画素 σ_i の周りで「周辺化」し、次の周辺化事後確率:

$$P(\sigma_i | \{\tau\}) \equiv \sum_{\sigma \neq \sigma_i} P(\{\sigma\} | \{\tau\}) \quad (54)$$

を求め、確率 $P(\sigma_i = 1 | \{\tau\})$ と $P(\sigma_i = -1 | \{\tau\})$ とを比較し、 $P(\sigma_i = 1 | \{\tau\})$ の方が大きければ i 番目の画素の推定値 $\hat{\xi}_i$ を $\hat{\xi}_i = 1$ に、逆に $P(\sigma_i = -1 | \{\tau\})$ が大きければ、 $\hat{\xi}_i = -1$ とするような推定も考えられる。式で書けば

$$\begin{aligned} \hat{\xi}_i &= \text{sgn} [P(\sigma_i = 1 | \{\tau\}) - P(\sigma_i = -1 | \{\tau\})] \\ &= \text{sgn} \left[\sum_{\sigma_i} \sigma_i \sum_{\sigma \neq \sigma_i} P(\{\sigma\} | \{\tau\}) \right] \\ &= \text{sgn} \left[\frac{\sum_{\sigma_i} \sigma_i \sum_{\sigma \neq \sigma_i} e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i}}{\sum_{\sigma} e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i}} \right] \\ &= \text{sgn} \left[\frac{\sum_{\sigma} \sigma_i e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i}}{\sum_{\sigma} e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i}} \right] \\ &= \text{sgn} \left[\frac{\sum_{\sigma} \sigma_i e^{-E(\{\sigma\})}}{\sum_{\sigma} e^{-E(\{\sigma\})}} \right] \equiv \text{sgn} [\langle \sigma_i \rangle_{J,h}] \end{aligned} \quad (55)$$

である。この推定法を有限温度推定 (finite temperature estimation) と呼ぶことにしよう。この名称の由来は、(55) の形がエネルギー関数 (53) を持つ磁性体の温度 $T = 1$ での局所磁化に相当することに由来する。この考え方下では、前出の MAP 推定は $T = 0$ つまり、絶対ゼロ度での推定ということになる。これをもう少し良く見るためには、(55) 式の 3 行目を

$$\hat{\xi}_i = \text{sgn} \left[\frac{\sum_{\sigma} \sigma_i e^{-E(\{\sigma\})/T}}{\sum_{\sigma} e^{-E(\{\sigma\})/T}} \right] \quad (56)$$

と書き直せば 上式右辺で $T = 1$ と置いたものが有限温度推定値を、 $T = 0$ と置いたものが $E(\{\sigma\})$ を最小にするような σ_i を与えることから MAP 推定値に対応することが容易に見て取れるであろう。

劣化の度合いが大きい場合には MAP 推定と比べて有限温度推定が有効である (ノイズを「揺らぎ」とみなしたとき、(55) 式中の温度による「熱揺らぎ」を巧くこの「ノイズ揺らぎ」と調節することにより、1 画素当たりの誤り率が最小化される.)。

2.4 確率モデルのパラメータ推定

前節までで、デジタル画像及び、劣化の過程を確率モデルで表現し、我々の画像復元の問題を定式化した。既に説明したように、各確率モデルはパラメータ J, h により特定され、このパラメータをどのような値に選ぶかにより、MAP 推定、及び有限温度推定の精度が変化する。このとき、例えば、原画像と修復画像との間

のハミング距離を用い、この距離を最小にするようなパラメータ J, h を選ぶ方策をとることはできない。なぜならば、我々は原画像に関する情報を用いることができないからである。それでは、どのような基準のもとにこれらのパラメータを選んだらよいのであろうか。以下でその手続きを説明する。

2.4.1 周辺尤度関数最大化法

確率モデルのパラメータ J, h を決定するには劣化画像 $\{\tau\}$ に対し、次に定義される周辺尤度をコスト関数として導入し、パラメータに関する最適化問題を解けばよい (周辺尤度最大化法)。

$$\begin{aligned} \mathcal{L}(J, h : \{\tau\}) &= \log \text{tr}_{\{\sigma\}} P_h(\{\tau\} | \{\sigma\}) P_J(\{\sigma\}) \\ &= \log \text{tr}_{\{\sigma\}} e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j + h \sum_i \tau_i \sigma_i} - \log \text{tr}_{\{\sigma\}} e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j} - \log 2 \cosh h \end{aligned} \quad (57)$$

ここに、上式右辺の第 2,3 項はそれぞれ、事前分布、尤度の規格化因子 Z_m, Z_l の対数をとったものであることに注意しよう。劣化画像に関する平均操作 $[\dots]_{\{\tau\}}$ を施した場合の周辺尤度が確率モデルのパラメータの真値 (J_*, h_*) で最大値をとること、つまり、不等式: $[\mathcal{L}(J_*, h_* : \{\tau\})]_{\{\tau\}} \geq [\mathcal{L}(J, h : \{\tau\})]_{\{\tau\}}$ が成り立つことは確率分布 $P_{h_*}(\{\tau\} | \{\sigma\}) P_{J_*}(\{\sigma\})$ と $P_h(\{\tau\} | \{\sigma\}) P_J(\{\sigma\})$ 間のカルバック距離の非負性を用いることにより、によって簡単に示すことができる。この問題において我々は原画像に関する情報を何も持たず、その部分のデータが欠落しているの、推定値 $\{\sigma\}$ の自由度に関して和 $\text{tr}_{\{\sigma\}}(\dots) = \sum_{\sigma_1=\pm 1} \dots \sum_{\sigma_N=\pm 1}(\dots)$ をとり、観測データ $\{\tau\}$ を残して確率分布 $P_h(\{\tau\} | \{\sigma\}) P_J(\{\sigma\})$ を周辺化したものの対数をもってパラメータ決定のコスト関数としているわけである。従って、パラメータの最尤推定値を求めるには周辺尤度 \mathcal{L} をパラメータに関して最大化すればよい。

2.4.2 EM アルゴリズムによる周辺尤度の最大化

周辺尤度 \mathcal{L} を最大化する手法として EM アルゴリズム (Expectation-Maximization) が知られている。この方法では周辺尤度を直接最大化せず、尤度関数のステップ依存した事後確率での平均: $\text{tr}_{\{\sigma\}} P_{J_t, h_t}(\{\sigma\} | \{\tau\}) \log P_h(\{\tau\} | \{\sigma\}) P_J(\{\sigma\})$ で定義される Q 関数を最大化する手続きにより周辺尤度を間接的に最大化する。

$$\begin{aligned} Q(J, h | J_t, h_t) &= -NJ \frac{\text{tr}_{\{\sigma\}} \varepsilon_B e^{J_t \sum_{\langle ij \rangle} \sigma_i \sigma_j + h_t \sum_i \tau_i \sigma_i}}{\text{tr}_{\{\sigma\}} e^{J_t \sum_{\langle ij \rangle} \sigma_i \sigma_j + h_t \sum_i \tau_i \sigma_i}} - Nh \frac{\text{tr}_{\{\sigma\}} \varepsilon_C e^{J_t \sum_{\langle ij \rangle} \sigma_i \sigma_j + h_t \sum_i \tau_i \sigma_i}}{\text{tr}_{\{\sigma\}} e^{J_t \sum_{\langle ij \rangle} \sigma_i \sigma_j + h_t \sum_i \tau_i \sigma_i}} \\ &\quad - \log \text{tr}_{\{\sigma\}} e^{J \sum_{\langle ij \rangle} \sigma_i \sigma_j} - N \log 2 \cosh h \end{aligned} \quad (58)$$

ここに $\varepsilon_B = -(1/2N) \sum_{\langle ij \rangle} \sigma_i \sigma_j, \varepsilon_C = -(1/N) \sum_i \tau_i \sigma_i$ である。EM アルゴリズムは Q 関数の計算 (Expectation) とその最大化 (Maximization) :

$$J_{t+1} = \arg \max_J Q(J, h | J_t, h_t), \quad h_{t+1} = \arg \max_h Q(J, h | J_t, h_t) \quad (59)$$

の繰り返しで構成される。ここではまず劣化過程の反転確率 p が既知であるとし、 $h = h_* = (1/2) \log((1-p)/p)$ とおいて J の更新のみを考えよう。このとき更新式の『方程式』は具体的に

$$u_m(J_{t+1}) \equiv \frac{\text{tr}_{\{\sigma\}} \varepsilon_B e^{J_{t+1} \sum_{\langle ij \rangle} \sigma_i \sigma_j}}{\text{tr}_{\{\sigma\}} e^{J_{t+1} \sum_{\langle ij \rangle} \sigma_i \sigma_j}} = \frac{\text{tr}_{\{\sigma\}} \varepsilon_B e^{J_t \sum_{\langle ij \rangle} \sigma_i \sigma_j + h_* \sum_i \tau_i \sigma_i}}{\text{tr}_{\{\sigma\}} e^{J_t \sum_{\langle ij \rangle} \sigma_i \sigma_j + h_* \sum_i \tau_i \sigma_i}} \equiv u_p(J_t, h_*) \quad (60)$$

と書ける。幸いなことに上式左辺 u_m はデータ $\{\tau\}$ を含まない確率変数 $\{\sigma\}$ のみからなる系における統計量 ε_B の期待値であり、十分に多数個 ($N \rightarrow \infty$) の画素が 2 次元正方格子に配置されている場合には以下

に示す厳密解 $u_m(J)$ が知られており (Onsager (1944)), 更新式が

$$u_m(J_{t+1}) = u_p(J_t, h_*), \quad u_m(J) = -\coth 2J \left\{ 1 + \frac{2}{\pi} (2 \tanh^2 2J - 1) K(k) \right\} \quad (61)$$

と書けることになる. ただし, $K(k)$ はその母数が $k = 2 \tanh 2J / \cosh 2J$ で与えられる完全楕円関数:

$$K(k) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2 \sin^2 \phi}} \quad (62)$$

である. 従って, 具体的なアルゴリズムの処理過程は, まず初期値 J_0 を適当に設定し, $u_p(J_0, h_*)$ を計算する. ついで, その値 u_p と $u_m(J)$ の値が等しくなるような J を J_1 とし, その J_1 に対して再度 $u_p(J_1, h_*)$ を計算し, その値 u_p と $u_m(J)$ の値が等しくなるような J を J_2 とする … というように逐次的に進んで行くことになる. そしてパラメータ J は $J_0 \rightarrow J_1 \rightarrow J_2 \rightarrow \dots$ のように更新される. このダイナミックスの固定点が EM アルゴリズムの与える解である. このように書くだけならば容易だが, しかし問題はそう簡単ではない. $u_p(J, h_*)$ は観測データ $\{\tau\}$ を含む非一様な系での期待値であり, u_m のような厳密解は無い. 従って, これを正確に評価するには 2^N の和 $\text{tr}_{\{\sigma\}}(\dots)$ を計算する必要があるのだが, N が大きな場合には, この計算が現実的ではなくなるのである.

2.4.3 マルコフ連鎖モンテカルロ法の適用

ここでは, ミクロ変数である各画素 $\sigma_i (i = 1, \dots, N)$ に次の確率過程を課し, 十分な時間が経過した後に系 $\{\sigma\}$ が平衡分布 (事後分布): $e^{-H(\{\sigma\}|\{\tau\})} / \text{tr}_{\{\sigma\}} e^{-H(\{\sigma\}|\{\tau\})}$ に収束することに着目する. このとき事後分布でのアンサンブル平均が上述確率過程からのサンプリング平均 (時間平均) で置き換えられることを利用し, 式 (60) で与えられる ε_B の重み付き平均を $(1/N_{\text{MCS}}) \sum_{k=1}^{N_{\text{MCS}}} \varepsilon_B(k)$ で近似計算することを考えよう. この方法をマルコフ連鎖モンテカルロ法 (MCMC: Markov Chain Monte Carlo 法) と呼ぶ.

ミクロ変数の従う確率過程 (シングルスピントリップ・メトロポリス法)

- (i) $i \in \{1, 2, \dots, N\}$ をランダムに選び, その位置の画素を反転させる. $\sigma_i \rightarrow -\sigma_i$.
- (ii) $\Delta E = H(F_i\{\sigma\}|\{\tau\}) - H(\{\sigma\}|\{\tau\})$ を計算し (F_i は位置 i の画素を反転させる演算子), $\Delta E \leq 0$ であれば, (i) の反転を受け入れ, $\Delta E > 0$ であっても, 確率: $e^{-\Delta E}$ で反転を受け入れる.
- (iii) (i)(ii) を $N \times N_{\text{MCS}}$ 回繰り返す.

これは前回学習した 2 変数関数であるエネルギー関数に対するノイズを利用したアルゴリズムの多変数関数であるエネルギー関数への拡張版であることに注意する. なお, この MCMC 法は u_m の厳密解を知らない場合, $H = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j$ と置くことで u_m の計算にも適用できる. つまり, EM アルゴリズムを適用する前に予め MCMC 法で J と u_m の対応関係をテーブルにしておく. そして J の更新毎に $u_p = u_m$ を与える J をそのテーブルから拾い出し, 該当値を次ステップの J 値とするように更新を進めればよい. 図 13 (左) に比較のため u_m の厳密解と MCMC 法による結果 ($N = 50 \times 50, N_{\text{MCS}} = 2 \times 10^5$) を載せる. ところで, MCMC 法では N_{MCS} の値をいくらに設定するか, 言い換えれば, パラメータを更新する際, 直前のパラメータ値で特徴づけられる系がどのくらい平衡状態から遠いのか, によって結果が異なるので注意が必要である. 事実, 図 13 (左) からわかるように, u_m の厳密解と MCMC 法から求めた u_p の交点が EM アルゴリズムの決定する J 値を与えるのだが, この交点の値は $N_{\text{MCS}} = 100$ と $N_{\text{MCS}} = 2 \times 10^5$ とでは有意な差があり, N_{MCS} を大きくとり, 系を十分に緩和させた場合の方が真値 $J_* = 0.465$ に近いことが見てとれる. 前に述べた確率伝播法は基本的に平均場近似の考え方に基づいており, ミクロ変数 (の期待値) の更新は EM

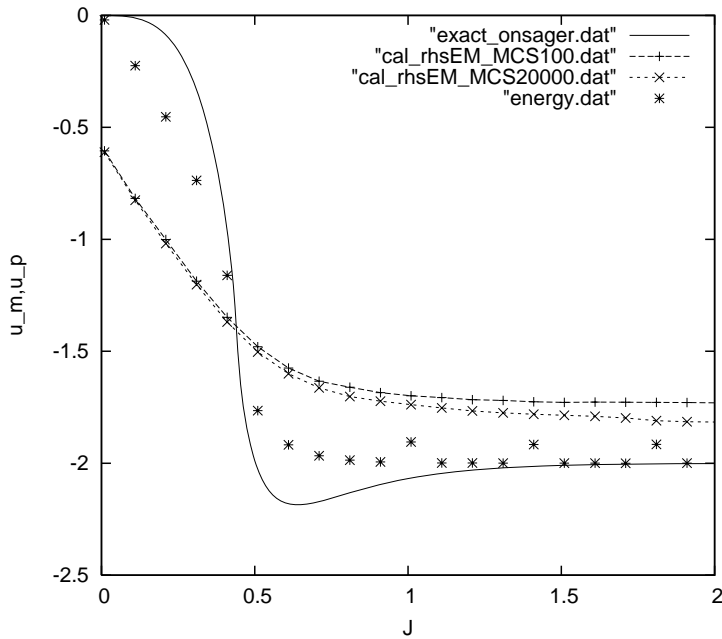


図 13: $u_m(J)$ の厳密解と $N_{\text{MCS}} = 2 \times 10^5$ の MCMC 法による近似解, 及び, $u_p(J, h_*)$ の $N_{\text{MCS}} = 100, 2 \times 10^5$ の MCMC 法による計算結果. u_m, u_p の交点が J の推定値を与える (真値は $J_* = 0.465$.)

アルゴリズムにおけるマクロ変数の更新式と同様に確定的な式で与えられ, 逐次的に画素間の相関を取り込むことにより精度が向上する. 一方, MCMC 法ではミクロ変数に確率過程を課し, そこから生成される系の時系列の長時間平均で期待値を近似計算するため, 系の平衡状態への緩和時間がミクロ変数だけでなく, パラメータの推定精度にとっても決定的である. これは MCMC 法を用いた EM アルゴリズムではミクロ変数の確率過程 (ダイナミクス) がマクロ変数 J, h に共役な統計量 $\varepsilon_B, \varepsilon_C$ の期待値を決定し, それが Maximization ステップを介して J, h に影響し, さらに, そうして更新された J, h がミクロ変数の確率過程にフィードバックされる... といった具合にミクロ変数とマクロ変数が相互に絡み合いながら情報処理が進んで行くことになるからである. 特に確率的情報処理の課題では事後確率が必ず何らかの観測データを含み, 空間的に非一様なものとなる. そこで統計物理の知見を借りるならば, 空間的に非一様な系の平衡状態への緩和は一般的に言って非常に遅い, という事実が知られている. そうであるならば Expectation ステップ, つまり u_p の計算では系の緩和に関して慎重にならざるを得ず, どのタイミングで処理を Maximiation ステップへと切り替えてマクロ変数を更新したならば, どの程度の精度が得られるのか, を調べることはアルゴリズムの性能を議論する上ではとても重要になる. それはここで示した簡単な数値実験からも明らかであろう.

2.4.4 デジタル画像復元の実行例

さて, 最後に劣化度 h をも未知とし, これと J を EM アルゴリズムで同時推定しよう. h の更新式は $\hat{u}_p(J, h)$ を ε_C の事後確率での平均値として, $h_{t+1} = \tanh^{-1}(\hat{u}_p(J_t, h_t))$ と書けるので, これと h も未知であるとした場合の J の更新式: $u_m(J_{t+1}) = u_p(J_t, h_t)$ とを組んで反復計算を行う. 結果を図 13 (右) に示す. この図より, 原画像として $J_* = 0.465$ と選んだ事前確率 (50) からのスナップショットに選び, 劣化率を $p = 0.1$ ($h_* = 1.1$) に定め, $N_{\text{MCS}} = 100$, 及び, $N_{\text{MCS}} = 2 \times 10^5$ としてアルゴリズムを動作させた場合, それらの収束値には小さいが有意な差があることが見て取れる. なお, 当然のことながら図 16 (b1) のよう

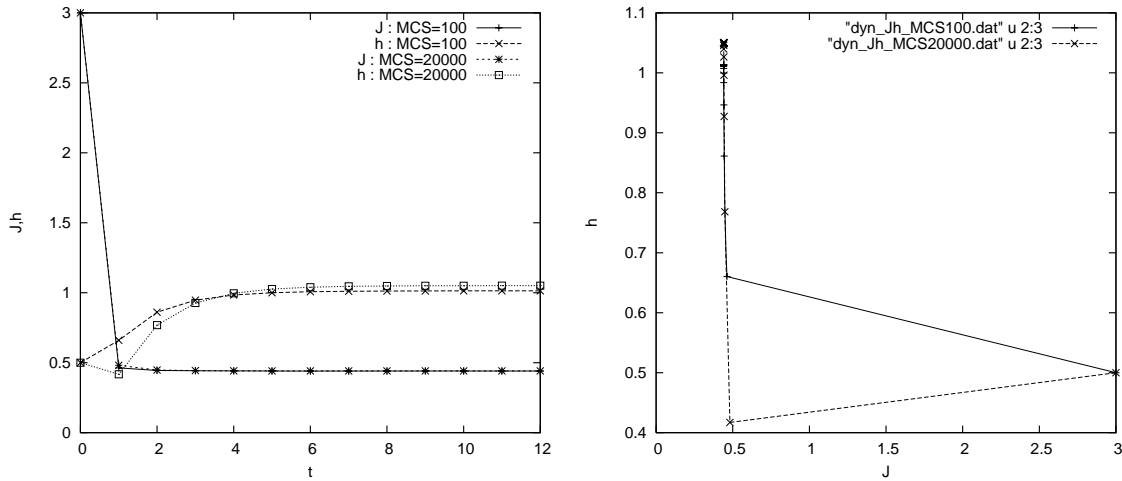


図 14: EM アルゴリズムの処理過程. $J_* = 0.465$ で与えられる事前確率 (50) からのスナップショットを原画像に選んだ場合であり, 劣化率は $p = 0.1$ ($h_* = 1.1$). 右図はパラメータ空間 (J, h) でみた EM アルゴリズム処理過程の軌跡.

な印鑑等の自然画像に対しても MCMC 法に基づく EM アルゴリズムを適用することができる. そのパラメータの時間発展と最終的に得られる修復画像とをそれぞれ図 13 (右) と図 16 (b3) に載せておく. 図 17 に $J_* = 0.465$ の事前分布 (50) からのスナップショットを原画像とした場合に EM アルゴリズムで決定されたパラメータに対して, シミュレーテッド・アニーリングを実行した際のエネルギー E の時間変化 (左), 及び, ビット誤り率 P_b の時間変化 (右) をプロットしたのも載せておこう.

2.5 平均場アルゴリズムによる計算量の低減

前節で有限温度推定を導入したが, 局所磁化での計算には $\{\sigma\}$ の自由度にわたる計算量, つまり, 2^N 個の和の計算が必要である. 前出の画像のように $200 \times 200 = 4 \times 10^4$ 程度の画像サイズでは $\sim 2^{10^4}$ もの和を計算する必要がある. これを全ての画素 σ_i について行うわけであるから, 事実上, この計算は不可能ということになってしまう. このとき, 次に説明する平均場近似 (mean-field approximation) に基づくアルゴリズム — 平均場アルゴリズム (mean-field algorithm) — の適用が有効である.

まず, 各画素が $L_1 \times L_2 = N$ の正方格子に乗っていると, 番地 (i, j) の画素を σ_{ij} で表現する (図 18(左) 参照). すると, エネルギー関数は

$$E_2(\{\sigma\}) \equiv -J \sum_{ij} \sigma_{ij} (\sigma_{ij+1} + \sigma_{ij-1} + \sigma_{i+1j} + \sigma_{i-1j}) - h \sum_{ij} \tau_{ij} \sigma_{ij} \quad (63)$$

となる. このとき, 平均場近似では, 着目する画素 σ_{ij} の周りの画素をその期待値で置き換える (図 18(右) 参照). つまり

$$\sigma_{ij+1} \simeq \langle \sigma_{ij+1} \rangle_{J,h} \equiv m_{ij+1} \quad (64)$$

$$\sigma_{ij-1} \simeq \langle \sigma_{ij-1} \rangle_{J,h} \equiv m_{ij-1} \quad (65)$$

$$\sigma_{i+1j} \simeq \langle \sigma_{i+1j} \rangle_{J,h} \equiv m_{i+1j} \quad (66)$$

$$\sigma_{i-1j} \simeq \langle \sigma_{i-1j} \rangle_{J,h} \equiv m_{i-1j} \quad (67)$$

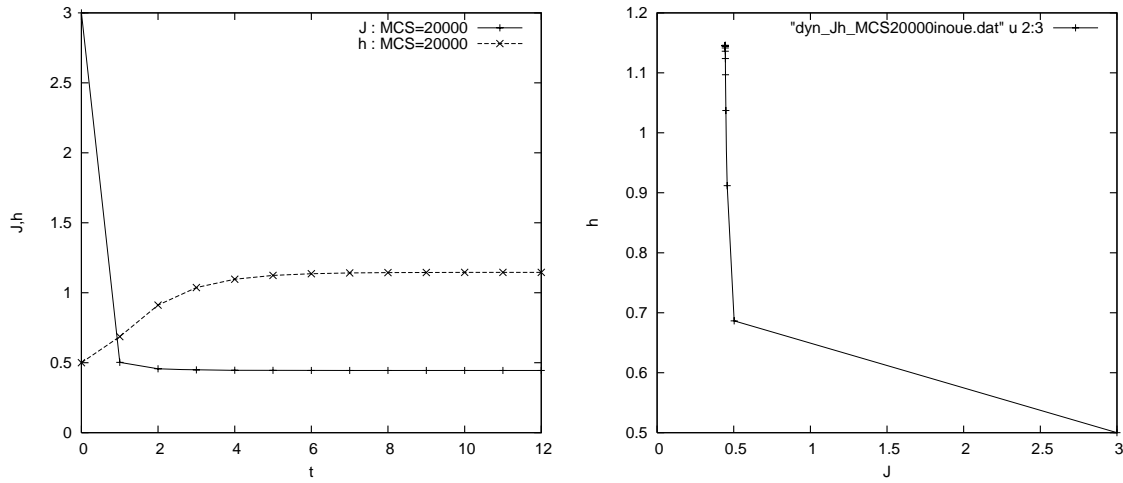


図 15: EM アルゴリズムの処理過程. 図 16 (b1) の印鑑画像を原画像としたものである. 劣化率は $p = 0.1$ ($h_* = 1.1$). 右図はパラメータ空間 (J, h) でみた EM アルゴリズム処理過程の軌跡.

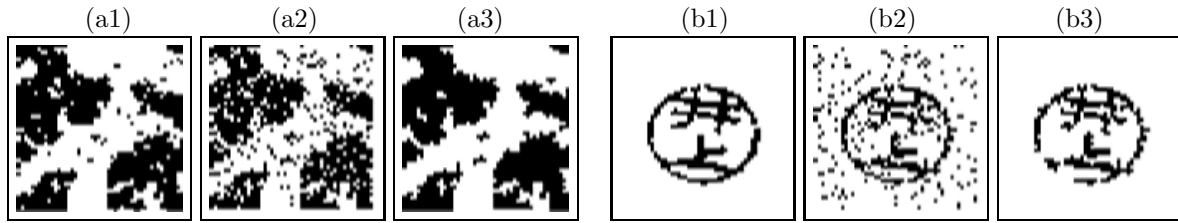


図 16: 原画像 (a1)(b1) ((a1) は $J_* = 0.465$ の事前分布 (50) からのスナップショット), 劣化画像 (a2)(b2) (劣化率: $p = 0.1$) $N_{MCS} = 2 \times 10^2$ での EM アルゴリズムで求められたパラメータ (図 13 (右) の収束点) で温度制御 $T = 3/\sqrt{t}$ でのシミュレーテッド・アニーリング法による最大事後確率推定により復元された画像 (a3)(b3).

この下で, 着目する画素 σ_{ij} の期待値は

$$m_{ij} = \langle \sigma_{ij} \rangle_{J,h} = \frac{\sum_{\sigma} \sigma_{ij} e^{-E_2^{(\text{mean-field})}(\{\sigma\})}}{\sum_{\sigma} e^{-E_2^{(\text{mean-field})}(\{\sigma\})}} \quad (68)$$

$$E_2^{(\text{mean-field})}(\{\sigma\}) \equiv -J \sum_{ij} (m_{ij+1} + m_{ij-1} + m_{i+1j} + m_{i-1j}) \sigma_{ij} - h \sum_{ij} \tau_{ij} \sigma_{ij} \quad (69)$$

となり, 実際に (68) の和は簡単に実行できて全ての番地 (i, j) にある画素の平均値に関し,

$$m_{ij} = \tanh [J(m_{ij+1} + m_{ij-1} + m_{i+1j} + m_{i-1j}) + h\tau_{ij}] \quad \forall (i,j) \quad (70)$$

なる連立式が得られる.

従って, 劣化画像 $\{\tau\}$ に対し, 式 (70) を

$$m_{ij}^{(t+1)} = \tanh [J(m_{ij+1}^{(t)} + m_{ij-1}^{(t)} + m_{i+1j}^{(t)} + m_{i-1j}^{(t)}) + h\tau_{ij}] \quad \forall (i,j) \quad (71)$$

という「連立漸化式」とみなし, 全ての画素に関して反復計算し, 引き続き 2 ステップ間の誤差 $|m_{ij}^{(t+1)} - m_{ij}^{(t)}|$ がある ϵ 以内に収まったときの m_{ij} を解とし, 原画像の画素 ξ_{ij} の推定値 $\hat{\xi}_{ij}$ を (55) に従って

$$\hat{\xi}_{ij} = \text{sgn}(m_{ij}) \quad (72)$$

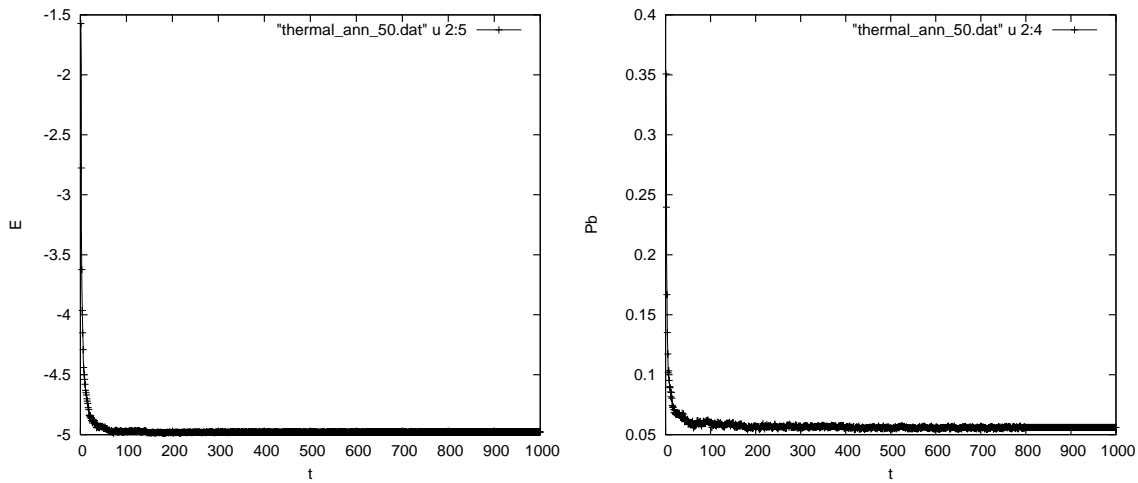


図 17: $J_* = 0.465$ の事前分布 (50) からのスナップショットを原画像とした場合に EM アルゴリズムで決定されたパラメータに対して、シミュレーテッド・アニーリングを実行した際のエネルギー E の時間変化 (左), 及び、ビット誤り率 P_b の時間変化 (右) 横軸の時間単位は 1 モンテカルロ・ステップに選んである。

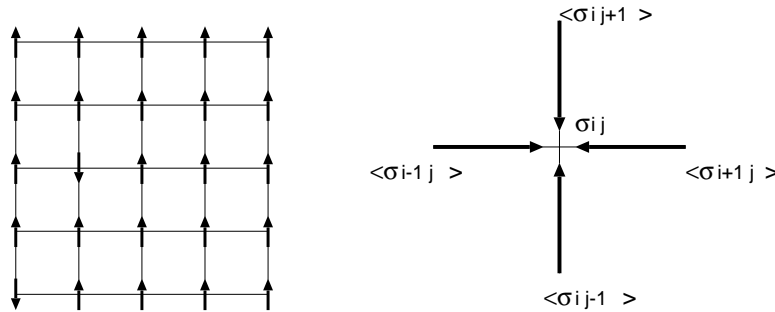


図 18: 2 次元正方格子に乗った画素 (左). と平均場近似の概念図 (右).

とすれば良い。

問題 2 :

原画像 ($J_* = 0.465$ の事前分布 (50) からのスナップショット) として用いるデジタル画像を

http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/KONTON2005/konton2005.html

からダウンロードし、この画像を 20% 程度の 2 元対称通信路で劣化後、平均場近似を用いた有限温度推定を用いて復元せよ。上記 URL に置く画像ファイルの大きさは $L_1 \times L_2 = 50 \times 50$ であるが、計算時間や手持ちの計算機への負荷等を考え、各自 GIMP, PaintShop 等の画像編集ソフトで画像サイズを変更すること。