解 説

圧縮センシングへの統計力学的アプローチ

樺 島 祥 介

東京工業大学大学院総合理工学研究科*

Statistical-Mechanical Approach to Compressed Sensing

Yoshiyuki Kabashima

Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology*

概要

スパース性(多くの成分がゼロであること)によって特徴づけられる高次元の信号を少ない数の観測データから復元する枠組みは、しばしば、圧縮センシング(compressed (compressive) sensing)と呼ばれる。近年、信号処理の文脈で盛んに研究されているが、その数理の本質は“スパース性”という事前知識を利用したベイズ推論にある、と捉えることもできる。このことは圧縮センシングが同様にベイズの公式を用いて定式化できる通信理論、学習理論の諸問題とも浅からぬ関係があることを意味している。本稿では、最近著者らが行った統計力学の方法に基づいた圧縮センシングに関する性能評価の概要を紹介するとともに、そうした圧縮センシングの数理に関する“横のつながり”について述べる。

1. は じ め に

N 次元の実数値信号 $\mathbf{x}^0 \in \mathbb{R}^N$ が $M \times N$ 実数行列 F を用いた線形変換により M 次元のデータ $\mathbf{y} = F\mathbf{x}^0 \in \mathbb{R}^M$ に変換される状況を考える。 F は既知であるとする。データ $\mathbf{y}$ から原信号 $\mathbf{x}^0$ を復元する問題を考えよう。 $\alpha = M/N < 1$, すなわち、原信号の次元よりデータの次元が少ない場合、方程式 $\mathbf{y} = F\mathbf{x}$ の解は不定になる。そのため $\mathbf{y}$ から $\mathbf{x}^0$ を完全に復元することはこの場合一般には不可能である。

しかしながら、 $\mathbf{x}^0$ がスパースなベクトル、つまり、 $\mathbf{x}^0$ 中の非ゼロ成分の割合 ρ が 1 より小さな場合にはスパース性に関する事前知識を利用することで $\alpha < 1$ の場合でも $\mathbf{y}$ からの $\mathbf{x}^0$ の完全な復元が可能になるかもしれない。このように、事前知識を利用して、圧縮された表現から原信号を復元する枠組みはしばしば圧縮センシング(compressed (compressive) sensing)と呼ばれる¹⁾。圧縮センシングに関する研究の歴史は長く、古くは 1970 年代にまで遡ることができる²⁻⁴⁾。ところが、2000 年代半ばからの Donoho, Candès, Tao による精力的な研究^{5, 8-10)}の影響もあり、その裾野は最近になって急速に拡大している。

本稿では、最近著者らが行った統計力学の方法を用い

た圧縮センシングに関する典型性能評価¹¹⁾の概要を紹介する。本特集の和田山氏の解説でも述べられているが、圧縮センシングは Nyquist-Shannon による標本化(サンプリング)定理の一般化と考えることができる。標本化定理は「最大周波数成分が B よりも小さなことが保証されている任意の原信号に対して、 $2B$ よりも高い周波数 f_s でのサンプリングによって完全復元が可能になる」ことを保証する。これに対し、圧縮センシングはスパース性で特徴付けられるやや狭いクラスの原信号に対し、スパースであるという事前知識を積極的に利用することで標本化定理を超えた高い復元性能を追求する試みと位置づけられる。

事前知識を数学的に表す最も有力な表現法はベイズの公式であろう。我々の解析もこの公式に基づいている。ところで、1990 年代以降、学習理論、通信理論においてもベイズ流の定式化に沿った研究が盛んに行われてきた。こうした共通する数理に目を向けることで圧縮センシングはニューラルネットワーク研究で取り組まれているいくつかの問題とも浅からず関係していることがわかる。こうした“横のつながり”についても言及したい。

以下、問題設定、解析の概要、関連する話題、の順で述べていく。

* 〒 226-8502 横浜市緑区長津田町 4259

2. 問題設定

圧縮センシングの簡単な数理モデルとして以下のよう
な状況を想定する． N 次元の実数値をとる原信号
 $\mathbf{x}^0 \in \mathbb{R}^N$ の各成分 x_i^0 ($i = 1, 2, \dots, N$) が独立に同一
分布 $P(x) = (1 - \rho)\delta(x) + \rho \exp(-x^2/2)/\sqrt{2\pi}$ に従
うとする．ただし， ρ ($0 \leq \rho \leq 1$) は非ゼロ成分の割
合 (信号密度) を表し， $\delta(u)$ は Dirac のデルタ関数で
ある．圧縮された表現 $\mathbf{y}$ は $\mathbf{y} = F\mathbf{x}^0$ によって得られ
る．以下では，圧縮表現 $\mathbf{y}$ と圧縮行列 F がともに既
知であるとして，原信号 $\mathbf{x}^0$ を復元することが可能か，
という問題を考える．

2.1 スパース表現

いきなり抽象的な話になってしまったが，例えば，
“原信号がスパース” というのは次のような状況に対
応している．画像や音声など日常で接する情報の多く
は一般に見かけの次元よりもずっと少ない情報量しか
持っていない．つまり，冗長に表現されている．この
ことは (画像や音声などを想定した) N 次元の情報 $\mathbf{I}$
はある適当な基底 $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N$ を使うと，スパース
な展開係数 x_i^0 ($i = 1, 2, \dots, N$) を用いて

$$\mathbf{I} = \sum_{i=1}^N x_i^0 \mathbf{b}_i \quad (1)$$

のように表現できることを意味している (図 1)．

見かけの表現でゼロ成分の多い信号というのはかな
り特殊な例であるかもしれない．しかしながら，適当
な基底を用いることによりスパースな表現に変換でき
る，ということは実問題で現れる多くの信号に対して
それほど無理なく期待できる性質なのである．

2.2 圧縮＝サンプリング

次は“圧縮”についてである． $\mathbf{I} = (I_k)$ を適当にサ
ンプリングすることで圧縮したい．標準化定理では $\mathbf{b}_i$
は三角関数等の基底関数であり，サンプリングとして
は成分を表す添字 k に関して等間隔で $\mathbf{I}$ の成分 I_k の
値を得ることを想定している．しかしながら，ここ
ではより一般化して線形独立な $M (< N)$ 本のベクトル
 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_M$ との内積の値

$$y_\mu = \mathbf{a}_\mu \cdot \mathbf{I} = \sum_{i=1}^N \mathbf{a}_\mu \cdot \mathbf{b}_i x_i^0 \quad (2)$$

($\mu = 1, 2, \dots, M$) を測定することを“サンプリング”
と考える．以下， $\alpha = M/N$ を圧縮率と呼ぶ． $F = (F_{\mu i}) = (\mathbf{a}_\mu \cdot \mathbf{b}_i)$ を圧縮行列と定義すれば，上述
の圧縮表現 $\mathbf{y} = F\mathbf{x}^0$ とはスパースな展開係数とサン
プリングにより得られる圧縮データを直接つなげる関

図1 スパースな原信号．(a): $i = 1, 2, \dots, N (= 64)$
に対して $x_7^0 = 1$, $x_{13}^0 = 3$, $x_{17}^0 = 2$ とし，その
他の成分はゼロとした．(b): N 次元の基底ベク
トルを $\mathbf{b}_i = (\cos(2\pi i k / N))$ ($k = 1, 2, \dots, N$)
とし $\mathbf{I} = \sum_{i=1}^N x_i^0 \mathbf{b}_i$ によって情報ベクトル $\mathbf{I}$
を構成した． $\mathbf{x}^0 = (x_i^0)$ がスパースであるという
事前知識がない場合， $\mathbf{I}$ に関する N 個すべての
成分をサンプリングしない限り， $\mathbf{x}^0$ を復元する
ことはできない．しかしながら，非ゼロ成分が
 $x_7^0, x_{13}^0, x_{17}^0$ のみである，とわかっている場合
には $\mathbf{I}$ に関する 3 つのデータがあれば済む．圧
縮センシングでは，非ゼロ成分数が少ないこと
が保証されているがその位置まではわからない
状況で効率的な圧縮と復元を試みる．

係式である，と解釈することができる．

実際の応用に際しては，多くの場合，具体的な“表
現基底” $\{\mathbf{b}_i\}$ と“サンプリング基底” $\{\mathbf{a}_\mu\}$ に対して議
論が進められる．一方で，圧縮センシングの枠組み自体
の可能性や限界を理論的に考察する際には，個別性をな
るべく排し単純化した状況を想定する発想も自然であ
る．そうした理論的な考察では圧縮行列をランダム行列
として特徴づけることがしばしば行われている^{5,8~10)}．
このような仮定は表現基底がランダムベクトルにより
構成されている，あるいは，ランダムな基底によりサン
プリングが行われる状況に対応していると考えられる．

以下の我々の解析も，この方針にしたがい， F の各成
分 $F_{\mu i}$ ($\mu = 1, 2, \dots, M$; $i = 1, 2, \dots, N$)，が独立に
平均ゼロ分散 N^{-1} の正規分布に従う状況を仮定する．

2.3 L_p 復元

与えられた $\mathbf{y}$, F から原信号 (スパースな展開係数)
 $\mathbf{x}^0$ を復元したい． F は横長の行列なので， $\mathbf{y} = F\mathbf{x}$
をそのまま解くことはできない．しかしながら，我々
は $\mathbf{x}^0$ がスパースなベクトルであることを知っている．

図2 圧縮センシングの素朴な信号復元法. 非ゼロ成分の数・位置を仮定し, 線形方程式を解くことをすべての可能性に対して行う. 必要な計算量が一般に N に関して指数的に増えるため実際的ではない.

そこで, $\mathbf{x}^0$ のうち非ゼロになる成分を圧縮表現の次元 M よりも少ない K 個だけである (K は M 以下の適当な数), と仮定し N から M を選ぶすべての組み合わせに関して“縦長”の線形方程式 $\mathbf{y} = \mathbf{F}'\mathbf{x}'$ を解くという素朴な戦略が考えられる (図2). ただし, $\mathbf{F}'$ は $\mathbf{F}$ から非ゼロと仮定した成分に対応する K 列を抜き出してできる $M \times K$ の行列であり, $\mathbf{x}'$ は非ゼロ成分に対応する K 次元ベクトルである. K を M から1つつ減らしながら, すべての可能性を試していけば適当な K のところで $\mathbf{y} = \mathbf{F}\mathbf{x}$ を満たす解が唯一定まるはずであり, それが原情報 $\mathbf{x}^0$ に他ならない.

言うまでもなく, この方法の難点は行列のサイズが大きくなるにつれて計算量が膨大になることである. 実用的には非ゼロ成分の選択とそれらの値を $\mathbf{y}$, $\mathbf{F}$ から決めることを同時にできることが望まれる. こうした要求に答える実際的な復元法としてしばしば検討されているのが線形制約下での L_p ノルムの最小化

$$\text{minimize } \|\mathbf{x}\|_p \text{ subject to } \mathbf{F}\mathbf{x} = \mathbf{y}, \quad (3)$$

である. ただし,

$$\begin{aligned} \|\mathbf{x}\|_p &= \lim_{\epsilon \rightarrow +0} \sum_{i=1}^N |x_i|^{p+\epsilon} \\ &= \begin{cases} \sum_{i=1}^N |x_i|^p & (p > 0) \\ \mathbf{x} \text{ の非ゼロ成分の個数} & (p = 0) \end{cases} \quad (4) \end{aligned}$$

はベクトル $\mathbf{x}$ に関する L_p ノルムを表す**. 以下, これを L_p 復元とよび, 特に, $N, M \rightarrow \infty$, $\alpha = M/N \sim O(1)$ の極限での $p = 0, 1$ および 2 に対する復元の可能性について考察する.

3. 解析の概要

3.1 L_p 復元とベイズ推論

L_p 復元の典型的な性能を評価するために, 線形制約下の最小化問題 (3) に基づいてパラメータ β を含む「事後分布」

$$P(\mathbf{x}|\mathbf{y}) = \frac{e^{-\beta\|\mathbf{x}\|_p} \delta(\mathbf{F}\mathbf{x} - \mathbf{y})}{Z(\beta; \mathbf{y})}, \quad (5)$$

を構成する. $e^{-\beta\|\mathbf{x}\|_p}$ は原信号 $\mathbf{x}$ がスパースであるという事前知識を表す「事前分布」の役割を果たしている. 一方, $\delta(\mathbf{F}\mathbf{x} - \mathbf{y})$ は $\mathbf{x}$ が与えられた下で圧縮表現 $\mathbf{y}$ が得られる「条件付き確率」である. 事前知識とサンプリングによって得られる情報をベイズの公式を用いて組み合わせることにより式 (5) が得られる. 規格化定数 $Z(\beta; \mathbf{y}) = \int d\mathbf{x} e^{-\beta\|\mathbf{x}\|_p} \delta(\mathbf{F}\mathbf{x} - \mathbf{y})$ は統計力学における分配関数に対応し, 性能評価を行う際に鍵となる量である.

事後分布 (5) は $\mathbf{x}$ が $e^{-\beta\|\mathbf{x}\|_p}$ に比例した割合で生成されている, という仮定の下で $\mathbf{y}$ を得た際に最良の推論を行うための基盤を与える. 特に, $\beta \rightarrow \infty$ の極限を考えると (5) は一般に (3) の解をサポートとする一様分布に収束する. このことは, L_p 復元が, L_p ノルムにより定まる事前分布を用いたベイズ推論の特別な場合, として位置づけられることを示している.

3.2 解析の方針: 何を評価すれば良いのか

我々の問題設定では $\mathbf{x}^0$ や $\mathbf{F}$ はランダムに生成されると仮定している. つまり, 統計的なゆらぎに左右され, 一般には, 具体的なサンプルに応じて復元は成功したり, 失敗したりする. そういった問題に対しては, 大きく分けて最悪評価と典型評価の2種類の評価法がある. ここでは, ベイズの枠組みに基づいて典型評価の立場から L_p 復元の性能評価を行おう. 最悪評価については, 和田山氏の解説をご覧ください.

我々が吟味したいのは, 典型的な $\mathbf{x}^0$ と $\mathbf{F}$ に対して, L_p 復元による復元結果が原情報 $\mathbf{x}^0$ を正しく表しているか, ということである. これは1自由度あたりの平均自乗誤差

** (和田山氏の解説も含めて) 通常は式 (4) を $1/p$ 乗したものを L_p ノルムと呼ぶ. ただし, いずれの定義を用いても最小化条件は変わらないため, 本稿では後の計算が容易になる表現 (4) を用いる.

$$\text{MSE} = N^{-1} [\langle |\mathbf{x} - \mathbf{x}^0|^2 \rangle] = Q - 2m + \rho \quad (6)$$

を用いて特徴づけることができる[†]。ただし、 $\langle \cdots \rangle$ は $\beta \rightarrow \infty$ における (5) に関する平均操作、 $[\cdots]$ は F , $\mathbf{x}^0$ に関する平均操作をそれぞれ表し $Q = N^{-1} [\langle |\mathbf{x}|^2 \rangle]$, $m = N^{-1} [\mathbf{x}^0 \cdot \langle \mathbf{x} \rangle]$ とした。また、 $N, M \rightarrow \infty$, $\alpha = M/N \sim O(1)$ の極限では対数の法則から $\rho = N^{-1} [\langle |\mathbf{x}^0|^2 \rangle]$ となることを用いた。

復元が成功するのは $\langle \mathbf{x} \rangle = \mathbf{x}^0$ が成立する場合である。このことは、典型的に復元が成功する場合には $Q = m = \rho$ となり、その結果 $\text{MSE} = 0$ となることを意味する。十分大きな α に対しては、 $\text{MSE} = 0$ となるであろう。一方、 α の値を下げていくと、信号密度 ρ に依存したある臨界的な値 $\alpha_c(\rho)$ で典型的に復号が失敗するようになり MSE は正の値を取るようになるはずである。つまり、復元の成功/失敗の転移を表す臨界値 $\alpha_c(\rho)$ は各 ρ に対して $Q = m = \rho$ が成立する α の下限として特徴づけられる。この臨界値を評価することでそれぞれの L_p ノルムに関する復元性能の限界を特徴づけることができる。

3.3 2重の平均とレプリカ法

式 (6) を求めるためには、ランダムに与えられる $\mathbf{x}^0$, F に依存して定まる事後分布 (5) に関する平均を求め、さらに、その結果を $\mathbf{x}^0$ および F について平均する、という2重の平均評価が必要になる。レプリカ法はこのようなタイプの問題を解決するために統計力学で発達した評価法である。煩雑な計算を必要とするため、その詳細は文献¹¹⁾ にゆずり、ここではその大まかな方針について述べる。

問題の核心は大自由度極限 $N, M \rightarrow \infty$, $\alpha = M/N \sim O(1)$ でどのように Q や m を求めるか、である。レプリカ法ではこの評価に際して

$$Q = \lim_{n \rightarrow 0} N^{-1} \frac{[\langle Z^n(\beta; \mathbf{y}) \times |\mathbf{x}|^2 \rangle]}{[\langle Z^n(\beta; \mathbf{y}) \rangle]} \quad (7)$$

$$m = \lim_{n \rightarrow 0} N^{-1} \frac{[\mathbf{x}^0 \cdot \langle Z^n(\beta; \mathbf{y}) \times \mathbf{x} \rangle]}{[\langle Z^n(\beta; \mathbf{y}) \rangle]} \quad (8)$$

という恒等式を利用する。これらの恒等式を評価するために $n = 1, 2, \dots$ の場合に成立する表現 $[Z^n(\beta; \mathbf{y})] = \int \prod_{a=1}^n d\mathbf{x}^a e^{-\beta \sum_{a=1}^n \|\mathbf{x}^a\|_p} [\prod_{a=1}^n \delta(F\mathbf{x}^a - \mathbf{y})]$ に着目する。統計的な対称性からこの右辺の積分は積分変数 $\mathbf{x}^a$ ($a = 1, 2, \dots, n$) と $\mathbf{x}^0$ の相対的な位置関係を

表す巨視的な変数

$$q_{ab} = N^{-1} \mathbf{x}^a \cdot \mathbf{x}^b \quad (a \geq b = 1, 2, \dots, n) \quad (9)$$

$$m_a = N^{-1} \mathbf{x}^0 \cdot \mathbf{x}^a \quad (a = 1, 2, \dots, n) \quad (10)$$

を用いて

$$[Z^n(\beta; \mathbf{y})] = \int e^{N g_n(\{q_{ab}\}, \{m_a\})} \prod_{a \geq b} dq_{ab} \prod_{a=1}^n dm_a \quad (11)$$

という形に書き直すことができる。煩雑になるので具体的な表現は省略しているが、 $g_n(\{q_{ab}\}, \{m_a\})$ は (9), (10) の拘束下で $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^n$ に関する多重積分を評価することにより求まる N に依存しない関数である。積分変数 $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^n$ のことを一般にレプリカと呼ぶ。

式 (11) の右辺の表現は、重み $e^{-\beta \sum_{a=1}^n \|\mathbf{x}^a\|_p} \times [\prod_{a=1}^n \delta(F\mathbf{x}^a - \mathbf{y})]$ に比例してレプリカが生成される場合その相対位置を表す巨視的な変数 q_{ab} や m_a は $N \rightarrow \infty$ において $g_n(\{q_{ab}\}, \{m_a\})$ を最大にする値に収束することを意味している。ところで、 n 個のレプリカは互いに同等なので $g_n(\{q_{ab}\}, \{m_a\})$ を最大化する値はレプリカを区別する置換に関して不変な

$$q_{aa} = Q(n) \quad (a = 1, 2, \dots, n) \quad (12)$$

$$q_{ab} = q(n) \quad (a > b = 1, 2, \dots, n) \quad (13)$$

$$m_a = m(n) \quad (a = 1, 2, \dots, n) \quad (14)$$

という形になる、と仮定するのは自然であろう。これをレプリカ対称仮定と呼ぶ。興味深いことに、レプリカ対称仮定の下で $g_n(\{q_{ab}\}, \{m_a\})$ を最大化することにより得られる $Q(n)$, $q(n)$, $m(n)$ は一般に $n = 1, 2, \dots$ に限らず実数 $n \in \mathbb{R}$ に対しても定義可能な関数になる。そこで、それらの関数が $n \in \mathbb{R}$ に対しても成り立っていると“信じて” (7), (8) をそれぞれ $Q = \lim_{n \rightarrow 0} Q(n)$, $m = \lim_{n \rightarrow 0} m(n)$ と置き換えて評価する。この一連の手続きが (レプリカ対称仮定の下での) レプリカ法である。

3.4 結果

$p = 0, 1, 2$ に対し、レプリカ対称仮定の下でレプリカ法を適用することにより以下の結果が得られる。

3.4.1 $p = 0$

成功解 $Q = m = \rho$ は $\alpha > \rho$ ならば、またその場合に限り、安定である。よって、 $\alpha_c(\rho) = \rho$ となる。 $\mathbf{x}^0$ の非ゼロ成分の位置がすべてあらかじめ明らかにされている場合でも $\alpha > \rho$ は $\mathbf{y} = F\mathbf{x}$ が唯一の解を持つための必要条件である。このことは L_0 復元が理論的

[†] 細かいことを言うと、実際には $N \rightarrow \infty$ の極限での解析を行うので、 $O(N)$ 以下の誤差がある場合でも「成功」と判定する規範を用いていることに注意。文献^{6,7)} では誤差がゼロでないと成功と認めない規範で解析を行っている。

な限界性能を漸近的に達成することを意味している。ただし、残念ながら局所的な安定性を吟味するとレプリカ対称仮定により得られる L_0 復元の評価値は統計力学的に妥当でない解であることが示される。こうした場合、詳細な性能評価を行うためにはより複雑な解を仮定するレプリカ非対称仮定に基づく解析が必要となる。

3.4.2 $p = 1$

臨界値は

$$\alpha_c(\rho) = 2(1 - \rho)H(\hat{\chi}^{-1/2}) + \rho. \quad (15)$$

で与えられる。ただし、 $\hat{\chi}$ は $H(x) = \int_x^\infty Dt$ としたとき、方程式

$$\hat{\chi} = \alpha^{-1} \left[2(1 - \rho) \left((\hat{\chi} + 1)H(\hat{\chi}^{-1/2}) - \hat{\chi}^{1/2} \frac{e^{-1/(2\hat{\chi})}}{\sqrt{2\pi}} \right) + \rho(\hat{\chi} + 1) \right], \quad (16)$$

により定まる解である。また、この解は統計力学的に妥当であることも示される。

3.4.3 $p = 2$

成功解は $\alpha \geq 1$ ならば、またその場合に限り、安定である。つまり、 $\alpha_c(\rho) = 1$ となる。この解は統計力学的に妥当である。ただし、 $\alpha \geq 1$ であれば、スパース性を用いなくても方程式 $\mathbf{y} = \mathbf{F}\mathbf{x}$ を掃きだし法などで解くことにより $\mathbf{x}^0$ を復元することが可能である。このことから、 L_2 復元には圧縮センシングの能力がないことが結論付けられる。

3.4.4 数値的検証

レプリカ法を用いるとレプリカ対称性に関する適当な仮定の下、性能評価指標に関する具体的な数値を求めることができる。図 3(a) に $p = 0, 1, 2$ に対して得られた臨界値 $\alpha_c(\rho)$ を示す。参考のため、和田山氏の解説で述べられている Candès と Tao による $p = 1$ に対する十分条件を挿入図に示している。評価の立場や復元成功の基準に関して異なる点はあるもののこれらの評価値の違いは Candès と Tao による最悪評価には改善の余地が残されている可能性を示唆している。

ただし、レプリカ法による評価値は数学的に正しいことが保証された数値ではなく、あくまでもある仮定の下で得られた「仮説」に過ぎないことを忘れてはならない。仮説は実験と比較することによって、その妥当性を「検証」する必要がある。こうした議論の進め方は、自然科学ではごく一般的である一方で、定理・証明形式による数学的アプローチが主流の情報科学の理論研究では従来あまり用いられてこなかった。そう

図3 (a): $p = 0, 1$ および 2 に対する L_p 復元の典型的な復元限界。 (b): L_1 復元に対する $\alpha_c(\rho = 0.5)$ の実験的評価。

した「仮説 & 実験による検証」による議論の進め方こそが統計力学的アプローチの最大の特徴である。

我々は、上述の結果のうち最も自明でない $p = 1$ の結果について $N = 10, 12, \dots, 30$ に対する数値実験によりその妥当性を検証した。図 3(b) は $\rho = 0.5$ に対し、各 N に関する 10^6 回の実験に基づいて評価した実験的臨界値 $\alpha_c(\rho = 0.5, N)$ を $1/N$ を横軸として示している。2 次フィッティングにより得られる $N \rightarrow \infty$ に対する外挿値は $\alpha_c(\rho = 0.5) \simeq 0.8317$ である。この値はレプリカ法による評価値 $\alpha_c(\rho = 0.5) = 0.83129\dots$ と小数点以下 3 桁目まで一致しており、我々の評価が妥当であることを実験的に支持している。

4. 関連する話題

以上が我々が圧縮センシングに対して行った解析の概要である。しかしながら、スパース性やバイズの公式に基づいた推論の話題は、ニューラルネットワークに関連する研究において著者が知るだけでも 1980 年代から繰り返し議論されている。以下ではそうした研究とのつながりについて述べる。

4.1 正則化とベイズ推論

ネットワークの構造決定は工学的な用途にニューラルネットワークを用いる際に頭の痛い問題である。入出力関係を表す M 個のデータ $D^M = \{(\mathbf{x}^1, y^1), (\mathbf{x}^2, y^2), \dots, (\mathbf{x}^M, y^M)\}$ の背後にあるルールをニューラルネットワーク $y = f(\mathbf{x}; \mathbf{w})$ で学習することを考えよう。ここで、 $\mathbf{w}$ は結合や閾値などの学習によって定めるネットワークパラメータであり、入力 $\mathbf{x}$ に対するネットワークの出力を一般的に $f(\mathbf{x}; \mathbf{w})$ と書いた。

単純過ぎるネットワークでは学習すべきルールを十分に表現できないかもしれない。とはいえ、必要以上に複雑なネットワークは例題を過学習し、未学習データに関する汎化性能の低下につながるものが経験的にも理論的にも示されている。

そこでしばしば行われるのが大きめのネットワークを用意した上で、データへの適合度を表すコストに罰金項を加えた関数

$$E(\mathbf{w}|D^M) + \lambda R(\mathbf{w}) \quad (17)$$

の最小化によりネットワークパラメータ $\mathbf{w}$ の学習を行う正則化である。ただし、 $E(\mathbf{w}|D^M)$ はデータへの適合度を表すコストであり、自乗誤差 $E(\mathbf{w}|D^M) = \sum_{\mu=1}^M (y^\mu - f(\mathbf{x}^\mu; \mathbf{w}))^2$ などがしばしば用いられる。一方、 $R(\mathbf{w})$ は罰金項でありネットワークパラメータ $\mathbf{w}$ の各成分 w_i の絶対値に関して増加する適当な関数が仮定される。 $\lambda > 0$ は罰金項の強さを調節するハイパーパラメータである。

1980 年代末、Levin らはデータに関するコスト $E(\mathbf{w}|D^M)$ から可変パラメータ $\mathbf{w}$ により特徴付けられる条件付き確率

$$P(\{y^\mu\}|\{\mathbf{x}^\mu\}; \mathbf{w}) \propto \exp(-\beta E(\mathbf{w}|D^M)) \quad (18)$$

を構成し、罰金項 $R(\mathbf{w})$ を事前分布

$$P(\mathbf{w}) \propto \exp(-\beta \lambda R(\mathbf{w})) \quad (19)$$

に対応させれば、式 (17) に基づく学習がベイズ推論の枠組みで議論できることを指摘した¹²⁾。さらに、彼らはベイズ流の定式化により、ニューラルネットワークの学習に確率的複雑度や予測誤差などの情報理論的な概念が自然に導入されること、また、スピングラス模型など不規則系の統計力学の問題との形式的な対応が明確になり統計力学由来の計算技術が活用できることも言及している。

前者の言及に基づき、渡邊はニューラルネットワークを特異な（正則でない）統計モデルと捉え、代数幾

何の方法を用いて確率的複雑度や予測誤差の漸近評価を可能にする理論を発展させている¹³⁾。また、本稿で紹介したレプリカ法に基づく性能評価は後者の言及の延長上にある成果であるといえる。

4.2 忘却付き構造学習と lasso

L_2 ノルム $\|\mathbf{w}\|_2$ を罰金項 $R(\mathbf{w})$ とする方法は計算の容易さから広く用いられている。これに対し、石川は L_1 ノルム $\|\mathbf{w}\|_1$ に比例した罰金項を用いることで余分なパラメータを除去し自動的にネットワーク構造を学習する効果を有する忘却付き構造学習を 1980 年代末に提案している¹⁴⁾。

L_1 ノルムを用いてスパースな構造を獲得する類似の方法としては Tibshirani による lasso が有名である¹⁵⁾。上述の文脈に当てはめると、lasso では入出力間に線形の関係 $y = \mathbf{w} \cdot \mathbf{x}$ を仮定し不等式制約下での最小化

$$\begin{aligned} & \text{minimize } \sum_{\mu=1}^M (y^\mu - \mathbf{w} \cdot \mathbf{x}^\mu)^2 \\ & \text{subject to } \|\mathbf{w}\|_1 \leq a \end{aligned} \quad (20)$$

によってネットワークパラメータ $\mathbf{w}$ を決定する。 $a > 0$ は制約の強さを調整する制御パラメータである。データ数 M がパラメータの次元 N より小さな場合には例題 D^M を完全に再現する $\mathbf{w}$ は不定になるが L_1 の不等式制約を課すことで、不要な成分は自動的に刈り込まれ質の良いパラメータを学習することができるというわけである。

4.3 スパースコーディング

第 2 節で、スパースな信号の例として画像や音声ที่เหมาะสมな基底とスパースな係数ベクトルによって展開できると述べた。こうした観点からニューラルネットワークの情報処理を議論している研究に Olshausen と Field による第 1 次視覚野のモデルがある¹⁶⁾。

第 2 節の記法を用いると、Olshausen と Field のモデルでは第 1 次視覚野で観測されるスパースなニューロンの応答は (1) のようなスパースな表現が得られるように基底 $\phi = \{\mathbf{b}_1, \mathbf{b}_2, \dots\}$ が学習された結果であると考えられる。これを実現する学習則はベイズ推論の枠組みに基づき、以下のような方針に沿って導出される。

展開係数 $\mathbf{x}$ と基底 ϕ を与えたとき、画像 $\mathbf{I}$ が、たとえば、条件付き確率

$$P(\mathbf{I}|\mathbf{x}, \phi) \propto \exp\left(-\beta \|\mathbf{I} - \sum_i x_i \mathbf{b}_i\|^2\right) \quad (21)$$

($\beta > 0$ は適当なハイパーパラメータ) にしたがって生成されるモデルを仮定する。ただし、 $\mathbf{x}$ はスパースな

係数ベクトルである．これをスパースな事前分布

$$P(\mathbf{x}) \propto \exp\left(-\lambda \sum_i S(x_i)\right) \quad (22)$$

によって特徴づける．ただし， $\lambda > 0$ はハイパーパラメータであり， $S(x)$ はゼロ付近で最小になり $|x|$ とともに増加する適当な関数である．式 (21), (22) を用いると外界から提示された画像 $\mathbf{I}$ に関する基底 ϕ の尤度

$$P(\mathbf{I}|\phi) = \int P(\mathbf{I}|\mathbf{x}, \phi) P(\mathbf{x}) d\mathbf{x} \quad (23)$$

を評価することができる．

ところで，生理学的に解釈すると ϕ は外界からの視覚情報と第 1 次視覚野のニューロンとの間のシナプス効率を表す変数であり，その変化はゆっくりしていると考えるのが自然である．そういった場合には， ϕ が少し変化する間にも多数の画像が提示されているはずである．Olshausen と Field はこうした考察に基づき ϕ は平均対数尤度

$$[\ln P(\mathbf{I}|\phi)] \quad (24)$$

を最大化するように学習されると仮定し，それを実現するための学習則を導いた．ただし， $[\cdot]$ は外界から提示される画像 $\mathbf{I}$ に関する平均を表す．さらに彼らは導出した学習則を画像データベースに適用し，第 1 次視覚野としてもっともらしいスパースな符号化が実現されることを実験的に示した．

圧縮センシングの立場から見直すと，以上の研究は，基底が未知の場合に信号だけでなく適当な圧縮行列 F も推定する状況に対応している．こうした課題はしばしばブラインド圧縮センシングと呼ばれている¹⁷⁾．

4.4 線形ベクトル通信路

無線通信の研究では，線形ベクトル通信路と呼ばれる数理モデル

$$\mathbf{y} = W\mathbf{x} + \mathbf{n} \quad (25)$$

がしばしば考察される．ただし， $\mathbf{x}$ は N 次元の入力ベクトルであり， $\mathbf{y}$ は M 次元の出力ベクトル， $\mathbf{n}$ は通信路ノイズの影響を表した M 次元のノイズベクトルをそれぞれ意味している． $M \times N$ 行列 $W = (W_{\mu i})$ は通信路行列と呼ばれる．

符号分割多重接続 (Code Division Multiple Access: CDMA) 通信路と多入力多出力 (Multiple Input Multiple Output: MIMO) 通信路は線形ベクトル通信路 (25) でモデル化される代表的なシステムである．CDMA では単一の受信アンテナに向けて N 人のユーザが同時に送信する N 個のシンボルが入力ベクトル

$\mathbf{x}$ に対応する．その際，それぞれのユーザ i はシンボルを W の i 番目の列ベクトルに対応する長さ M の拡散系列 ($W_{\mu i}$) に変調して無線で送信する．各ユーザが完全に同期して送信し，かつ，伝送波の強度が同一である場合，電磁波に関する重ね合わせの原理から受信側では各ユーザからの信号の総和にノイズ $\mathbf{n}$ が重畳した M 次元のデータが受信される．一方，MIMO の場合には，単一のユーザが通信の効率化を目的として N 本の送信アンテナから同時に送信する N 個のシンボルが $\mathbf{x}$ であると考ええる．これらのシンボルは送受信アンテナ間の通信効率を表す行列 $M \times N$ 行列で変換され，さらに通信路ノイズ $\mathbf{n}$ の影響を受けた M 次元のデータが M 本の受信アンテナによって受信される．

簡単のため $\mathbf{n}$ の各成分が独立に平均ゼロ，分散 σ^2 の正規分布に従うと仮定する．また， $\mathbf{x}$ の各成分も独立に同一の適当な事前分布 $P(x)$ に従うとする．このとき $\mathbf{y}$ を受信した下での $\mathbf{x}$ の事後分布は

$$P(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma^2}|\mathbf{y} - W\mathbf{x}|^2\right) \times \prod_{i=1}^N P(x_i) \quad (26)$$

によって与えられる． $\sigma^2 \rightarrow 0$, $P(x_i) \propto \exp(-\beta|x|^p)$ とすると，このことは我々の圧縮センシングの解析が線形ベクトル通信路の特別な場合に対応していることを意味している．

線形ベクトル通信路の統計力学的解析は田中や著者らによって進められており¹⁸⁻²⁰⁾，実際，圧縮センシングに関する結果もそれらの先行研究の知見に基づいて得られている．また，線形ベクトル通信路に関しては本稿で紹介した性能評価に関する解析だけでなく，与えられた個別の受信データ $\mathbf{y}$ から送信したベクトル $\mathbf{x}$ を具体的に推定するアルゴリズムの開発も著者，田中・岡田により統計力学の平均場近似のアイデアに基づいて行われている^{21, 22)}．圧縮センシングにこうした近似アルゴリズムを導入する研究は最近 Montanari らによって進められている²³⁾．

4.5 パーセプトロン学習

圧縮センシングが線形ベクトル通信路モデルの特別な場合に対応していると述べた．実は線形ベクトル通信路モデル自体も単純パーセプトロンに関するベイズ学習の特別な場合に対応している．小節 4.1 の最後でも触れたが解析の技術的側面についてはこれらの研究はいずれも以下の形式によるパーセプトロン学習に関する成果を起源としているのである．

再び M 個の入出力例 $D^M = \{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots\}$

$\dots, (\mathbf{x}^M, y^M)\}$ からのニューラルネットワークの学習に戻る．学習を行うモデルとして N 次元入力ベクトル $\mathbf{x}$ に対し，出力 y がシナプス結合 $\mathbf{w}$ との内積に依存した条件付き確率

$$P(y|\mathbf{w} \cdot \mathbf{x}) \quad (27)$$

にしたがって生成される確率的パーセプトロンを仮定する．「確率的パーセプトロン」と呼んだのはこのモデルが内積 $\mathbf{w} \cdot \mathbf{x}$ の正負に応じて 2 値出力 $y = \pm 1$ を返す単純パーセプトロンの一般化と見なすことができるからである．

Levin らの指摘に従うと， D^M からの学習を事後分布

$$P(\mathbf{w}|D^M) = \frac{P(\mathbf{w}) \prod_{\mu=1}^M P(y^\mu|\mathbf{w} \cdot \mathbf{x}^\mu)}{\int P(\mathbf{w}) \prod_{\mu=1}^M P(y^\mu|\mathbf{w} \cdot \mathbf{x}^\mu) d\mathbf{w}} \quad (28)$$

に基づいて議論することができる．ここで，因数分解された事前分布 $P(\mathbf{w}) = \prod_{i=1}^N P(w_i)$ を仮定し， $M \times N$ 行列 $(\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M)^T$ (T は行列の転置) を通信路行列 W ， $\mathbf{w}$ を線形ベクトル通信路における入力ベクトル $\mathbf{x}$ ，条件付き確率 (27) を $\mathbf{w} \cdot \mathbf{x}$ を中心とした分散 σ^2 の正規分布とすれば線形ベクトル通信路モデルに関する事後分布 (26) が得られる．

D^M を構成する M 個の入力ベクトルがある確率分布からランダムに生成される，と想定する場合，事後分布 (28) に基づいた学習の性能評価は小節 3.3 で述べた 2 重平均の評価を必要とする問題になる．この認識が広まった 1990 年代の初頭，パーセプトロンの統計力学的解析は，不規則系の統計力学の研究者の間で大流行し，大量の論文が生産された^{24, 25)}．伝達関数，シナプス結合の制約，ネットワークの形状などニューラルネットワークを特徴づける性質のあらゆる組み合わせについて，ニューラルネットワークの研究としての意義など顧みられることなく，とにかく片っ端から“計算された”，と言っても過言ではない．当時を振り返り一体何のための研究だったのかかわからない，といった批判もしばしば耳にする．

しかしながら，こうした一見無意味に思える網羅的研究の中に，その 10 年以上後に脚光を浴びることになる線形無線通信路や圧縮センシングの研究に直結する成果²⁶⁾が既に得られていることを見過ごすべきではない．2006 年に発表された報告書「忘れられた科学—数学」²⁷⁾を契機として，国策としての数学振興のあり方が近年注目されている．ここに述べたさまざまな話題の関連は，一見関係のなさそうな問題たちを共通の

数理で一網打尽にできる数学の力強さを例証すると同時に，数理的な技術を深めるとともに「適度な距離にある」多様な話題に接し常に関心を持ちながら数学研究を進めることの重要性を示しているように著者には思える．こうした観点からの研究振興が今後一層促進されることを期待したい．

5. お わ り に

著者たちが最近行った圧縮センシングの性能解析の概要を紹介するとともにニューラルネットワークの周辺にある話題との関連について述べた．話題が広がりすぎるので，ここでは触れなかったが線形制約の下で 0, 1 の 2 値変数に関してベイズ推論を行う低密度パリティ検査符号²⁸⁾やランダム制約充足問題²⁹⁾も広い意味で圧縮センシングとつながる話題である．

本稿ではベイズ流の定式化から圧縮センシングを眺めた．もちろん，他の視点から見た風景は違ったものになるだろう．一つは和田山氏が解説している低次元空間へのランダムな線形射影を施した際のベクトルの長さの変化を評価する Candès と Tao らの議論である⁸⁻¹⁰⁾．この他にも L_p 復元 (3) の成功の可能性を高次元空間でランダムに生成される多胞体の性質に還元することで議論する Donoho と Tannner による方法がある^{6, 7)}．驚くべきことに，Donoho-Tanner の結果は全く異なるアプローチで得られる我々の結果と一致する．

こうした異なるアプローチ間の関係の解明も今後検討すべき興味深い課題である．

謝辞

本研究の一部は文部科学省科学研究費補助金特定領域研究「情報統計力学の深化と展開」の補助を受けて行われた．

参 考 文 献

- 1) Candès, E.J. and Wakin, M.B. (2008): IEEE Signal Processing Magazine, Vol.25, p.21
- 2) Claerbout, J.F. and Muir, F. (1973): Geophys. Mag., Vol.38, p.826
- 3) Santosa, F. and Symes, W.W. (1986): SIAM J. Sci. Stat. Comput., Vol.7, p.1307
- 4) Donoho, D.L. and Stark, P.B. (1989): SIAM J. Appl. Math., Vol.49, p.906
- 5) Donoho, D. (2006): IEEE Trans. Inf. Theory, Vol.52, p.1289
- 6) Donoho, D.L. (2006): Discrete and Computational Geometry, Vol.35, p.617
- 7) Donoho, D.L. and Tanner, J. (2009): J. Amer.

- Math. Soc., Vol.22, p.1
- 8) Candès, E.J. and Tao, T. (2005): IEEE Trans. Inf. Theory, Vol.51, p.4203
 - 9) Candès, E.J., Romberg, J. and Tao, T. (2006): IEEE Trans. Inf. Theory, Vol.52, p.489
 - 10) Candès, E.J. and Tao, T. (2006): IEEE Trans. Inf. Theory, Vol.52, p.5406
 - 11) Kabashima, Y., Wadayama, T. and Tanaka, T. (2009): J. Stat. Mech., L09003
 - 12) Levin, E., Tishby, N. and Solla, S.A. (1990): Proc. of IEEE, Vol.78, p.1568
 - 13) Watanabe, S. (2001): Neural Comput., Vol.13, p.899
 - 14) Ishikawa, M. (1995): Artificial Intelligence, Vol.75, p.51
 - 15) Tibshirani, R. (1996): J. Roy. Stat. Soc. Ser. B, Vol.58, p.267
 - 16) Olshausen, B.A. and Field, D.J. (1996): Nature, Vol.381, p.607
 - 17) Gleichman, S. and Eldar, Y.C. (2010): Blind Compressed Sensing, arXiv:1002.2586
 - 18) Tanaka, T. (2002): IEEE Trans. Inf. Theory, Vol.48, p.2888
 - 19) Takeda, K., Uda, S. and Kabashima, Y. (2006): Europhys. Lett., Vol.76, p.1193
 - 20) Guo, D. and Verdú, S. (2005): IEEE Trans. Inf. Theory, Vol.51, p.1983
 - 21) Kabashima, Y. (2003): J. Phys. A, Vol.36, p.11111
 - 22) Tanaka, T. and Okada, M. (2005): IEEE Trans. Inf. Theory, Vol.51, p.700
 - 23) Donoho, D.L., Maleki, A. and Montanari, A. (2009): PNAS, Vol.106, p.18914
 - 24) Watkin, T.L.H., Rau, A. and Biehl, M. (1993): Rev. Mod. Phys., Vol.65, p.499
 - 25) Engel, A. and van den Broeck, C. (2001): Statistical Mechanics of Learning, Cambridge, UK: Cambridge University Press
 - 26) Seung, H.S., Sompolinsky, H. and Tishby, N. (1992): Phys. Rev. A, Vol.45, p.6056
 - 27) 細坪護孝, 伊藤裕子, 桑原輝隆 (2006): 「忘れられた科学—数学 主要国の数学研究を取り巻く状況及び我が国の科学における数学の必要性」<http://www.nistep.go.jp/achiev/ftx/jpn/po1012j/idx012j.html>
 - 28) MacKay, D.J.C. (1999): IEEE Trans. Inf. Theory, Vol.45, p.399
 - 29) Mézard, M., Parisi, G. and Zecchina, R. (2002): Science, Vol.297, p.812