

L_1 ノルム最小化に基づく圧縮センシング

高橋 昂
January 10, 2021

Abstract

ノイズなしの線形観測に基づいてデータが得られているとする。このとき、 L_1 ノルム最小化に基づく圧縮センシングを行った場合の性能をレプリカ計算で調べてみる。その際の計算メモ。元ネタは[KWT09]の計算で、結果の概要だけ知りたければこちらを読むのがよい。 $\mathbb{E}[Z^n]$ の計算周辺だけ読みたい人はセクション3を飛ばして読んでくれればよい。

1 introduction

$\mathbf{x}_0 = [x_{0,i}]_{1 \leq i \leq N} \in \mathbb{R}^N$ を信号ベクトル、 $A = [a_{\mu i}]_{\substack{1 \leq \mu \leq M \\ 1 \leq i \leq N}} \in \mathbb{R}^{M \times N}$ を観測行列とする。この信号ベクトルをノイズなしの線形観測

$$\mathbf{y} = A\mathbf{x}_0,$$

をもとに復元する問題を考える。信号ベクトルの成分はその $\rho N, \rho \in (0, 1)$ 程度の成分のみが非ゼロであるようなスパースベクトルであるとして、 L_1 ノルムの最小化に基づく推定を考える：

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1, \text{ subject to } \mathbf{y} = A\mathbf{x}. \quad (1)$$

$M, N \rightarrow \infty, M/N \in \alpha \in (0, 1)$ であるような大自由度極限（熱力学極限）において、観測数比 α とスパース率 $\rho \in (0, 1)$ に応じてどのような復元性能が得られるのかに興味がある。特に、 $\lim_{N \rightarrow \infty} \|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2^2 / N$ などのような、パラメータ全体の平均的な振る舞いを表す量について考えることにする。このような量を巨視的な量と呼ぶことにする。巨視的な量が自己平均的、つまり A に関する統計性を考えたときその期待値の周りに分布が集中すると期待し、 $\lim_{N \rightarrow \infty} \mathbb{E}_A[\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2^2 / N]$ といった期待値を評価することにする。このような期待値評価を典型評価と呼んだりもする。

2 setting

理論解析のために、以下のような仮定を置く¹。まず、観測行列は独立に平均ゼロ、分散 $1/M$ のガウス分布に従うものとする：

$$a_{\mu i} \sim_{\text{iid}} \mathcal{N}(0, 1/M).$$

また、信号の要素も非ゼロ成分の分布が q_+ で与えられるような分布に独立に従うものとする：

$$x_{0,i} \sim_{\text{iid}} \rho q_+ + (1 - \rho)\delta, \quad \rho \in (0, 1),$$

¹可解な設定のうちの最も単純なもの1つということで、これ以外が解けないという話ではない。

ここで、 δ はデルタ関数（測度）である²。信号の強度については、 $\mathbb{E}_{x_0}[x_0^2] = \rho$ であるものとする³。また $M, N \rightarrow \infty, M/N \rightarrow \alpha \in (0, 1)$ の熱力学極限での振る舞いを考える。以降、 $M = \alpha N$ として、この熱力学極限を単に $\lim_{N \rightarrow \infty}$ で書くことにする。

3 method

3.1 Boltzmann distribution

以下のようにBoltzmann分布 $p^{(\beta)}(\mathbf{x})$ を定める：

$$p^{(\beta)}(\mathbf{x}) = \frac{1}{Z} \exp(-\beta \|\mathbf{x}\|_1) \delta(\mathbf{y} - \mathbf{A}\mathbf{x}), \quad \beta > 0, \mathbf{x} \in \mathbb{R}^N.$$

ここで、 β は逆温度で Z は分配関数である。このようにBoltzmann分布を定めると、 $\beta \rightarrow \infty$ で分布は最適化問題(1)の解上の一様分布に収束すると考えられる。推定値 $\hat{\mathbf{x}}$ は

$$\hat{\mathbf{x}} = \lim_{\beta \rightarrow \infty} \int \mathbf{x} p^{(\beta)}(\mathbf{x}) d\mathbf{x},$$

と書ける。最適化問題の解の解析が目的であったが⁴、確率モデルの極限の解析として話をすすめる。

3.2 Replica method

一般に適当な関数 $\tilde{\psi} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ を用いて決まる巨視的量

$$\psi = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \tilde{\psi}(x_{0,i}, \hat{x}_i),$$

の振る舞いを調べたいとしよう⁴。先に述べたとおり、これについても自己平均性が成り立つとすれば $\mathbb{E}_A[\psi]$ を調べればよい。これはデルタ関数とそのFourier変換、及び指数関数のテイラー展開を用いて以下のように書き直すことができる：

$$\begin{aligned} \psi &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_A \left[\int \tilde{\psi}(x_{0,i}, s) \delta(s - \hat{x}_i) ds \right] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_A \left[\frac{1}{2\pi} \int \tilde{\psi}(x_{0,i}, s) e^{\sqrt{-1}s\hat{s}} e^{-\sqrt{-1}\hat{s}\hat{x}_i} ds d\hat{s} \right] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_A \left[\frac{1}{2\pi} \int \tilde{\psi}(x_{0,i}, s) e^{\sqrt{-1}s\hat{s}} \sum_{k=0}^{\infty} \frac{(-\hat{s}\sqrt{-1})^k}{k!} (\hat{x}_i)^k ds d\hat{s} \right] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{2\pi} \int \tilde{\psi}(x_{0,i}, s) e^{\sqrt{-1}s\hat{s}} \sum_{k=0}^{\infty} \frac{(-\hat{s}\sqrt{-1})^k}{k!} \mathbb{E}_A[(\hat{x}_i)^k] ds d\hat{s} \right]. \end{aligned}$$

つまり、 \hat{x}_i の任意次数のモーメントの典型評価ができればよい構造になっている。一般に $n \in \mathbb{N}$ に対して $d^n \mathbf{x} = d\mathbf{x}_1 d\mathbf{x}_2 \dots d\mathbf{x}_n$ を $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathbb{R}^{N \times n}$ 上の測度と定める。すると、 \hat{x}_i の定義から

$$\mathbb{E}_A[(\hat{x}_i)^k] = \lim_{\beta \rightarrow \infty} \int \prod_{a=1}^k x_a \mathbb{E}_A \left[\prod_{a=1}^k \frac{1}{Z} e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}_a) \right] d^k \mathbf{x}, \quad (2)$$

²[DMM09]とかだと観測数比で δ を使ってるんだけど、デルタ関数と被る記号を使うのはどうかと思う。

³仮説検定の研究とかする際にはこの大きさをコントロールしてトレードオフを見ることもある

⁴この辺の議論は色々やり方があると思う。まあ好きに。[JM14]はLegendre変換を使うちょっとおしゃれな計算をしている。ただ有限温度はちょっと大変か？

である。ここの a はただの添字であって、 A の要素とは異なることに注意してほしい⁵。
(2)の計算を実行する上での困難は Z^{-k} である。これも A, \mathbf{x}_0 に依存しているのだが、 Z^{-k} は計算が容易ではない。そこで恒等式 $\lim_{n \rightarrow 0} Z^{n-k} = Z^{-k}$ を用いて(2)を次の様に書き直す：

$$\mathbb{E}_A[(\hat{x}_i)^k] = \lim_{\beta \rightarrow \infty} \lim_{n \rightarrow 0} \int \prod_{a=1}^k x_{a,i} \mathbb{E}_A \left[Z^{n-k} \prod_{a=1}^k e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}_a) \right] d^k \mathbf{x}.$$

この表式の利点は $n = \{k, k+1, \dots\} \equiv \mathbb{N}_{k+}$ に対しては、分配関数の定義式を用いることで

$$\begin{aligned} & \int \prod_{a=1}^k x_a \mathbb{E}_A \left[Z^{n-k} \prod_{a=1}^k e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}_a) \right] d^k \mathbf{x} \\ & \stackrel{n \in \mathbb{N}_{k+}}{=} \int \prod_{a=1}^k x_a \mathbb{E}_A \left[\prod_{a=1}^n e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}_a) \right] d^n \mathbf{x}, \end{aligned} \quad (3)$$

と書けることである。この表式においては平均を取るべきファクターはすべて自明なかたちをしているため、 Z の負幂を直接扱うよりは遥かに問題が簡単になっている。
 $n \in \mathbb{N}_{k+}$ での計算結果がもしも $n \rightarrow 0$ に外挿できる表現であった場合には、 $n \in \mathbb{N}_{k+}$ での計算結果を $n \rightarrow 0$ に外挿することで所望の結果が得られる。この整数べきからの外挿を行う部分の一般的な数学的正当化は未解決問題であるものの、Parisiの計算スキームに則って計算を行った結果は正しいと考えられている⁶。ここで、実際には $\mathbb{E}_A[\prod_{a=1}^n e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}_a)]$ からの寄与が支配的であるから、

$$\int \mathbb{E}_A \left[\prod_{a=1}^n e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}_a) \right] d^n \mathbf{x} = \mathbb{E}_A[Z^n] \quad (4)$$

が $n \rightarrow 0$ でどのように振る舞うかがわかればその結果から容易に(3)のほうの計算処理もわかるのが普通である。あるいは、(2)式は

$$\mathbb{E}_A[(\hat{x}_i)^k] = \lim_{n \rightarrow 0} \frac{\int \prod_{a=1}^k x_{k,i} \mathbb{E} \left[\prod_{a=1}^n e^{\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}_a) \right] d^n \mathbf{x}}{\int \mathbb{E} \left[\prod_{a=1}^n e^{\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - \mathbf{A}\mathbf{x}_a) \right] d^n \mathbf{x}},$$

⁵慣習として統計力学の文脈ではレプリカの添字に a, b, c, d を用いることが多い。まあここでは誤解を招くことは少ないだろうと思う。

⁶計算しきれたものについては、 $N < \infty$ での数値実験から $N \rightarrow \infty$ に外挿した結果と整合的であるか、あるいは別の手段によって数学的に証明されたものと同じ結果を与えることが知られている。あとは $\mathbb{E}[Z^n], n \in (0, \infty)$ での関数系が実際に外挿されたものでよいことが後に証明されているなどの状況証拠がある。少なくとも可解になるような系においては間違っている例は知られていない（と思う）。が、いつでも正しいのかというと、ひょっとしたら間違った結果を出すような例が存在してもいいんじゃないかと考えている人はいるんじゃないかと思う。

ちなみに、[Tal10]の1.13は"The Physicist's replica method"で、数学者のM. Talagrandは以下のように述べている。

Hundreds of papers have been written relying on the replica method; the authors of these papers seem to have little doubt that this method always gives the correct answer. Its proponents hope that at some point it will be made rigorous. At the present time however it is difficult, at least for this author, to see in it more than a way to guess the correct formulas. Certainly the predictive power of the method is impressive. The future will tell whether this is the case because its experts are guided by a set of intuitions that is correct at a still deeper level, or whether the power comes from the method itself.

あとは確かいつかの数学セミナーでD. Donohoも強力だがとても数学理論とは言えないという位置づけをしていると書いてあった気がする。数学者の立場とそれ以外の人のレプリカ法に対する見方は随分違うように感じられる。

とも書けるのだということに注意すると、 $\log \mathbb{E}[Z^n]$ は n が整数の際の $N \times n$ 自由度の系の
キュムラント生成関数である。したがってこれがわかれば十分というのは自然である
とも言える。

このようにして $\mathbb{E}_A[(\hat{x}_i)^k]$ を求めた後、もとのテイラー展開の式に戻して ψ を調べれば
所望の巨視的な量が得られるというわけである。実際には今回の問題のような長
距離相互作用、かつすべての変数が統計的には均等に寄与しているような系では適
当な一体問題の計算によってモーメントが計算され、 A による平均はエフェクティ
ブな確率変数の平均に置き換えられ、それは i ごとに独立なものになる⁷。結果とし
て $\mathbb{E}_A[(\hat{x}_i)^k] = \mathbb{E}_\xi[(\hat{x}(\xi_i, x_{0,i}))^k]$ というような形になる。ここでの \hat{x} は適当な一体問題の解
を表す記号で、 ξ が A の効果を反映した確率変数である。すると、結局 ψ は

$$\begin{aligned}\psi &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{2\pi} \int \tilde{\psi}(x_{0,i}, s) e^{\sqrt{-1}s\hat{s}} \sum_{k=0}^{\infty} \frac{(-\hat{s}\sqrt{-1})^k}{k!} \mathbb{E}_\xi[(\hat{x}(x_{0,i}, \xi))^k] ds d\hat{s} \right] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\xi \left[\tilde{\psi}(x_{0,i}, \hat{x}(x_{0,i}, \xi)) \right],\end{aligned}$$

だが、 i についての和をとっているところが $x_{0,i}$ の経験分布の平均に置き換えられるか
ら、

$$\psi = \mathbb{E}_{\xi, x_0} \left[\tilde{\psi}(x_0, \hat{x}(\xi, x_0)) \right], \quad (5)$$

というような格好になる。大抵は $\mathbb{E}[Z^n]$ を計算して(5)の結果を使ってサクッと計算した結
果だけ書くのだが、まあこんな感じの理由だということである。

うえのように計算しにくいところを冪の極限で書き直し、整数冪の計算結果から外挿し
て極限を得ることを一般にレプリカ法と呼ぶ⁸。「レプリカ」の由来は整数の n での計算は
同等な相互作用を持つ自由度 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^N$ に基づくことにある。もともとの問題
は N 次元の自由度の系が1つあっただけだが、それを n 個に複製(replicate)したものを扱う
ところから「レプリカ」法となったのだと思う。特に整数冪からの外挿が怪しげである
ことを強調する場合(?)にはレプリカトリックと呼ぶこともあるようだ。

3.3 Helmholtz free energy

レプリカ法では、結局 $\log \mathbb{E}_A[Z^n]$ の $n \rightarrow 0$ での振る舞いがわかれば良さそうだとい
うことがわかる。ただし、これ自体の $n \rightarrow 0$ の極限は自明に1だから、 n^{-1} でスケーリングした極
限に興味がある：

$$\lim_{n \rightarrow 0} \frac{1}{n} \log \mathbb{E}_A[Z^n].$$

これは $\mathbb{E}[\log Z]$ に等しいから、結局はHelmholtzの自由エネルギーの典型評価

$$-\frac{1}{\beta} \mathbb{E}_A[\log Z],$$

を計算すれば良さそうだったという話になるわけである。論文とかでは自由エネルギー
はこうなるので、、と言うだけのことがあるけど、、、。いまの文脈ではこの密度のゼロ温
度極限の熱力学極限での振る舞い

$$f = \lim_{N, \beta \rightarrow \infty} \frac{-1}{\beta N} \mathbb{E}_A[\log Z], \quad (6)$$

に興味がある。

⁷ A の各行が共分散 $\Sigma \in \mathbb{R}^{N \times N}$ のガウス分布から得られている場合とかはこのノイズが i ごとに独
立にならず、相関を持ったものになる。局所場が $\hat{m}\mathbf{x}_0 + \sqrt{\chi}\Sigma^{-1/2}\boldsymbol{\xi}, \boldsymbol{\xi} \sim (0, I_N)$ というような形。
 $\Sigma^{1/2}$ だったかもしれない

⁸これは今回の ψ の計算だけに限らず。

4 replica calculation

ここでは前のセクションの内容を念頭において $\mathbb{E}[Z^n]$, f の計算を行う。

4.1 Z^n の整理

デルタ関数の表現

$$\delta(x) = \lim_{\gamma \rightarrow \infty} \left(\frac{\gamma}{2\pi} \right)^{\frac{1}{2}} \exp \left(-\frac{\gamma}{2} x^2 \right),$$

と、 $\mathbf{y} = A\mathbf{x}_0$ を用いると

$$\begin{aligned} Z^n &= \int \prod_{a=1}^n e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{y} - A\mathbf{x}_a) d^n \mathbf{x} \\ &= \lim_{\gamma \rightarrow \infty} \int \prod_{a=1}^n e^{-\beta \|\mathbf{x}_a\|_1} \gamma^{\frac{Nn\alpha}{2}} \prod_{a=1}^n \exp \left(-\frac{\gamma}{2} \|A(\mathbf{x}_a - \mathbf{x}_0)\|_2^2 \right) d^n \mathbf{x}. \end{aligned} \quad (7)$$

である。ここで、 2π の寄与は定数なので無視した。こうやって定数の部分が無視したり、あとで鞍点法を使うことを見越して $o(N)$ の寄与を無視したりして=で結んじゅうのは、(特にこの分野の)統計力学の人たちの(悪しき)慣習である(でも慣れると止められない)。

4.2 A 平均

(7)式から、 $\mathbb{E}[Z^n]$ のうち A に関する平均操作に関わる部分は

$$\mathbb{E}_A \left[\exp \left(-\frac{\gamma}{2} \|A(\mathbf{x}_a - \mathbf{x}_0)\|_2^2 \right) \right],$$

である。この計算をするには $u_{\mu,a} \equiv \mathbf{a}_{\mu}^{\top}(\mathbf{x}_a - \mathbf{x}_0)$, $\mu = 1, 2, \dots, N\alpha$; $a = 1, 2, \dots, n$ と定めると便利である。 A の各要素は独立なガウス分布だから、その線形結合もガウス分布に従う。ガウス分布は平均と分散で完全に指定できるが、平均については

$$\mathbb{E}[u_{\mu,a}] = 0, \quad \mu = 1, 2, \dots, N\alpha; a = 1, 2, \dots, n,$$

であり、二次モーメントは

$$\mathbb{E}[u_{\mu,a} u_{\nu,b}] = \delta_{\mu,\nu} \frac{1}{N} (\mathbf{x}_a^{\top} \mathbf{x}_b - \mathbf{x}_a^{\top} \mathbf{x}_0 - \mathbf{x}_b^{\top} \mathbf{x}_0 + \|\mathbf{x}_0\|_2^2), \quad \mu, \nu = 1, 2, \dots, N\alpha; a, b = 1, 2, \dots, n,$$

である。したがって、 $\bar{Q} \in \mathbb{R}^{n \times n}$ を

$$\bar{Q}_{ab} \equiv \frac{1}{N} (\mathbf{x}_a^{\top} \mathbf{x}_b - \mathbf{x}_a^{\top} \mathbf{x}_0 - \mathbf{x}_b^{\top} \mathbf{x}_0 + \|\mathbf{x}_0\|_2^2),$$

とし、さらに

$$\mathbf{u}_{\mu} = [u_{\mu,a}]_{1 \leq a \leq n}, \quad \mu = 1, 2, \dots, N\alpha,$$

とすれば、ある特定の \mathbf{x}_a , $a = 1, 2, \dots, n$ に対しては、 A 平均処理は以下ようになる:

$$\begin{aligned} &\mathbb{E}_A \left[\exp \left(-\frac{\gamma}{2} \|A(\mathbf{x}_a - \mathbf{x}_0)\|_2^2 \right) \right] \\ &\stackrel{(a)}{=} \prod_{\mu=1}^{N\alpha} \int \exp \left(-\frac{\gamma}{2} \mathbf{u}_{\mu}^{\top} \mathbf{u}_{\mu} \right) \exp \left(-\frac{1}{2} \mathbf{u}_{\mu}^{\top} \bar{Q}^{-1} \mathbf{u}_{\mu} \right) \frac{1}{(\det \bar{Q})^{1/2}} d\mathbf{u}_{\mu} \\ &\stackrel{(b)}{=} \exp \left(-\frac{N\alpha}{2} \log \det(I_n + \gamma \bar{Q}) \right). \end{aligned}$$

(a)では各 μ ごとに u が独立な確率変数になっていることを用いて整理。(b)では単に多次元のガウス積分をした。

ここまでで、 Z^n の平均は以下の様になる:

$$\mathbb{E}_A[Z^n] = \int \gamma^{\frac{Nn\alpha}{2}} \exp \left(-\frac{N\alpha}{2} \log \det(I_n + \gamma \bar{Q}) \right) \prod_{a=1}^n e^{-\beta \|\mathbf{x}_a\|_1} d^n \mathbf{x}.$$

4.3 秩序変数の導入

\bar{Q} は $\mathbf{x}_a^\top \mathbf{x}_b, \mathbf{x}_a^\top \mathbf{x}_0$ のみを通じて \mathbf{x}_a に依存している。 \mathbf{x}_a の積分のなかに恒等式

$$1 = \prod_{1 \leq a \leq b \leq n} \int \delta(NQ_{ab} - \mathbf{x}_a^\top \mathbf{x}_b) dQ_{ab},$$

$$1 = \prod_{a=1}^n \int \delta(Nm_a - \mathbf{x}_a^\top \mathbf{x}_0) dm_a,$$

を挟んで秩序変数 $Q \in \mathbb{R}^{n \times n}, \mathbf{m} \in \mathbb{R}^n$ を導入する。 $\|\mathbf{x}_0\|_2^2$ は $N\rho$ の周りからのゆらぎが $o(N)$ だろうと期待されるので、 $N\rho$ ですり替えてしまうことにする。このとき、 $\bar{Q}_{ab} = Q_{ab} - 2m_a + \rho$ となる。

デルタ関数をFourier変換し、その共役変数を \tilde{Q}, \tilde{m} で書く。さらに、 $\mathbf{x}_i = (x_{1,i}, x_{2,i}, \dots, x_{n,i})^\top \in \mathbb{R}^n$ と定めて整理し鞍点法を使うと

$$\begin{aligned} \mathbb{E}_A[Z^n] &\stackrel{(a)}{=} \int \exp\left(\frac{Nn\alpha}{2}\gamma + \frac{N}{2}\text{Tr}(Q\tilde{Q}) - N\mathbf{m}^\top \tilde{\mathbf{m}} - \frac{N\alpha}{2}\log \det(I_n + \gamma\bar{Q})\right) \\ &\quad \times \prod_{i=1}^N \int \exp\left(\frac{Nn\alpha}{2}\gamma + -\frac{1}{2}\mathbf{x}_i^\top \tilde{Q}\mathbf{x}_i + \tilde{\mathbf{m}}^\top \mathbf{x}_i - \beta\|\mathbf{x}_i\|_1\right) d\mathbf{x}_i dQ d\mathbf{m} d\tilde{Q} d\tilde{\mathbf{m}} \\ &\stackrel{(b)}{=} \int \exp\left(\frac{Nn\alpha}{2}\gamma + \frac{N}{2}\text{Tr}(Q\tilde{Q}) - N\mathbf{m}^\top \tilde{\mathbf{m}} - \frac{N\alpha}{2}\log \det(I_n + \gamma\bar{Q})\right) \\ &\quad + \sum_{i=1}^N \log \int_{\mathbf{x} \in \mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{x}^\top \tilde{Q}\mathbf{x} + \tilde{\mathbf{m}}^\top \mathbf{x} - \beta\|\mathbf{x}\|_1\right) d\mathbf{x} dQ d\mathbf{m} d\tilde{Q} d\tilde{\mathbf{m}} \\ &\stackrel{(c)}{=} \exp\left(N \underset{Q, \tilde{Q}, \mathbf{m}, \tilde{\mathbf{m}}}{\text{extr}} \left[\frac{n\alpha}{2}\gamma + \frac{1}{2}\text{Tr}(Q\tilde{Q}) - N\mathbf{m}^\top \tilde{\mathbf{m}} - \frac{\alpha}{2}\log \det(I_n + \gamma\bar{Q}) \right. \right. \\ &\quad \left. \left. + \frac{1}{N} \sum_{i=1}^N \log \int_{\mathbf{x} \in \mathbb{R}^n} \exp\left(-\frac{1}{2}\mathbf{x}^\top \tilde{Q}\mathbf{x} + \tilde{\mathbf{m}}^\top \mathbf{x} - \beta\|\mathbf{x}\|_1\right) d\mathbf{x} \right] \right) \end{aligned}$$

である。(a)では各 i ごとの積分が独立に行えることを利用した。(b)では各 i ごとの積分の結果が等しいことを使い、さらに x 積分の結果を指数関数の肩に上げた。最後に(c)では $N \gg 1$ で鞍点評価した。

4.4 レプリカ対称仮定

極値評価をやる必要があるが、このうちの最も単純なものとしてレプリカ対称性を満たしたものを考えることにする:

$$Q = \frac{\chi}{\beta} I_n + q \mathbf{1}_n,$$

$$\tilde{Q} = \beta \hat{Q} I_n - \beta^2 \hat{\chi} \mathbf{1}_n$$

$$m_a = m,$$

$$\tilde{m}_a = \beta \hat{m}.$$

ここで、 $\mathbf{1}_n \in \mathbb{R}^{n \times n}$ はその要素がすべて1の行列である。また、 β の選び方は恣意的な感じがするが、 Q_{ab} の対角と非対角要素の差が分散であり $\beta \rightarrow \infty$ で帯磁率が $(0, \infty)$ にあると仮定し、さらに各項が自由エネルギーに $O(1)$ の寄与を与えると仮定しながら構成していけばこのような選択になる。

上のレプリカ対称な構造を仮定し、さらに恒等式 $e^{\frac{a^2}{2}} = \int e^{a\xi} D\xi$, $D\xi = e^{-\frac{\xi^2}{2}}/\sqrt{2\pi}d\xi$ を利用して整理し、 $\gamma \rightarrow \infty$ の極限をとれば⁹

$$\begin{aligned} \lim_{N, \gamma \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[Z^n] &= n\beta \operatorname{extr}_{Q, \hat{Q}, \hat{m}, \hat{\chi}} \left[\frac{1}{2} \left(q + \frac{\chi}{\beta} \right) \hat{Q} - \frac{1}{2} \chi \hat{\chi} - m\hat{m} - \alpha \frac{q - 2m + \rho}{\chi} \right. \\ &\quad \left. + \frac{1}{\beta} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\xi} \left[\log \int \exp \left(-\frac{\beta \hat{Q}}{2} x^2 + \beta(\hat{m}x_{0,i} + \sqrt{\hat{\chi}}\xi) - \beta|x| \right) dx \right] \right] + \mathcal{O}(n^2) \\ &= n\beta \operatorname{extr}_{Q, \hat{Q}, \hat{m}, \hat{\chi}} \left[\frac{1}{2} \left(q + \frac{\chi}{\beta} \right) \hat{Q} - \frac{1}{2} \chi \hat{\chi} - m\hat{m} - \alpha \frac{q - 2m + \rho}{\chi} \right. \\ &\quad \left. + \frac{1}{\beta} \mathbb{E}_{x_0, \xi} \left[\log \int \exp \left(-\frac{\beta \hat{Q}}{2} x^2 + \beta(\hat{m}x_0 + \sqrt{\hat{\chi}}\xi) - \beta|x| \right) dx \right] \right] + \mathcal{O}(n^2). \end{aligned}$$

ここで、 i の和を取ることで $x_{0,i}$ の経験平均にすり替わることを用いた¹⁰。 $n\beta$ で割って $n \rightarrow 0, \beta \rightarrow \infty$ の極限をとり、 q を Q と書き直せば

$$\begin{aligned} f &= \operatorname{extr}_{Q, \hat{Q}, \hat{m}, \hat{\chi}} \left[\frac{1}{2} Q \hat{Q} - \frac{1}{2} \chi \hat{\chi} - m\hat{m} - \alpha \frac{q - 2m + \rho}{\chi} - \phi(\hat{Q}, \hat{m}, \hat{\chi}) \right], \\ \phi &= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \mathbb{E}_{x_0, \xi} \left[\log \int \exp \left(-\frac{\beta \hat{Q}}{2} x^2 + \beta(\hat{m}x_0 + \sqrt{\hat{\chi}}\xi) - \beta|x| \right) dx \right] \\ &= -\mathbb{E}_{x_0, \xi} \left[\frac{(|h| - 1)^2}{2\hat{Q}} \mathbf{1}(|h| > 1) \right], \quad h = \hat{m}x_0 + \sqrt{\hat{\chi}}\xi. \end{aligned}$$

極値条件は下記の通りである：

$$\hat{Q} = \hat{m} = \alpha \frac{1}{\chi}, \quad (8)$$

$$\hat{\chi} = \alpha \frac{q - 2m + \rho}{\chi^2}, \quad (9)$$

$$Q = \mathbb{E}_{x_0, \xi} [\langle x^2 \rangle], \quad (10)$$

$$m = \mathbb{E}_{x_0, \xi} [x_0 \langle x \rangle], \quad (11)$$

$$\chi = \mathbb{E}_{x_0, \xi} \left[\frac{\partial}{\partial h} \langle x \rangle \right]. \quad (12)$$

ここで、一般に関数 $\omega(x)$ に対して

$$\begin{aligned} \langle \omega(x) \rangle &= \lim_{\beta \rightarrow \infty} \frac{\int \omega(x) \exp \left(-\frac{\beta \hat{Q}}{2} x^2 + \beta(\hat{m}x_0 + \sqrt{\hat{\chi}}\xi) - \beta|x| \right) dx}{\int \exp \left(-\frac{\beta \hat{Q}}{2} x^2 + \beta(\hat{m}x_0 + \sqrt{\hat{\chi}}\xi) - \beta|x| \right) dx} \\ &= \omega(\hat{x}), \\ \hat{x} &= \arg \max_{x \in \mathbb{R}} \left[-\frac{\hat{Q}}{2} x^2 + hx - |x| \right] \equiv S(h; \hat{Q}), \end{aligned}$$

である。 $\beta \rightarrow \infty$ で積分に使っている測度が 1 点に集中することを用いた。この極値条件(8)-(12)を状態方程式と呼ぶこともある。状態というのは共役変数の組 $(\hat{Q}, \hat{m}, \hat{\chi})$ である。要するに x_i の周辺分布の振る舞いを特徴づけるのに十分な変数の組である。

⁹ このときに γ の発散する項はちゃんとキャンセルされる。

¹⁰ 自由エネルギーを得るだけなら、最初っから x_0 について平均とっておいても同じ結果が出てくる。

5 performance analysis based on the equation of state

5.1 巨視的量の表現

3の方針に沿って $(\hat{x}_i)^k$ がどのようなになるかを計算する。これは前節の自由エネルギーの計算に $\prod_{a=1}^k x_{a,i}$ をつけて計算をたどって行けば容易にわかるように

$$(\hat{x}_i)^k = S(h; \hat{Q})^k,$$

である。ここで、 h 内のパラメータや \hat{Q} 内のパラメータは極値条件(8)-(12)を満たすものである。したがって、所望の評価関数は以下のように評価できる:

$$\psi = \mathbb{E}_{x_0, \xi} \left[\psi \left(x_0, S(h; \hat{Q}) \right) \right]. \quad (13)$$

もともとは高次元の多体問題の話だったが、少数自由度の計算で話が済むということである。特に、二乗和誤差(MSE) $\|x_0 - \hat{x}\|_2^2/N$ は、 $Q - 2m + \rho$ で計算できる。

5.2 完全復元可能/不可能相転移

$(\rho, \alpha) \in [0, 1]^2$ の範囲で極値条件を数値的に解くと、以下のことがわかる:

- (ρ, α) がある境界線より上側にある場合、 $\hat{Q} = \hat{m} \rightarrow \infty, \hat{\chi} \rightarrow (0, \infty), Q = m = \rho, \chi = 0$ である。このとき、 $Q - 2m + \rho = 0$ となる
- (ρ, α) が同じ境界線の下側にある場合、 $\hat{Q} \rightarrow (0, \infty), \hat{\chi} \rightarrow (0, \infty)$ である。このとき、 $Q - 2m + \rho > 0$ となる

つまり、ある程度スパースかつ観測数があるのなら完全にMSEが0であるような完全復元が可能で、逆にスパースさが低い(ρ が大きい)か、観測数が少ないならば、MSEは非ゼロになってしまうという境界が存在するというわけである。この境界線の表現は完全復元解 $Q = m = \rho, \hat{Q} = \hat{m} \rightarrow \infty$ の安定性解析によって決めることが可能である。完全復元解では、 $\hat{\chi}$ だけが非自明な値になっているわけだが、これを決める自己無撞着条件式を出しその反復に関する安定性を調べる。具体的には、まず $\hat{Q} \gg 1$ であるとして、極値条件のうち Q, m の式を \hat{Q}^{-2} までで展開する。それを $\hat{\chi}$ を決める式に代入することで自己無撞着条件式を出す。特に、 $q_+ = \mathcal{N}(0, 1)$ の場合には

$$\hat{\chi} = \frac{1}{\alpha} \left(2(1 - \rho) \left((1 + \hat{\chi}) H \left(\frac{1}{\sqrt{\hat{\chi}}} \right) - \sqrt{\hat{\chi}} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2\hat{\chi}} \right) \right) + \rho(1 + \hat{\chi}) \right),$$

$$H(x) = \int_x^\infty D\xi,$$

という条件式が得られる。この固定点を $\hat{\chi}^*$ として、それに微小な摂動を加えて1反復した際にその摂動が拡大しない条件を書くと以下を得る:

$$\alpha > 2(1 - \rho) H \left(\frac{1}{\sqrt{\hat{\chi}^*}} \right) + \rho.$$

同じ信号モデルで実際に計算して得られるMSEを図示すると図1のようになる。さらに、 $\rho = 0.1$ の場合に完全復元不可能な相で α をコントロールした場合における各パラメータの振る舞いを図2に示しておく。

5.3 相転移について

自由エネルギーは $f = \lim_{N \rightarrow \infty} \frac{1}{N} \|\hat{x}\|_1$ なのだが、 $q_+ = \mathcal{N}(0, 1)$ かつ、 $\rho = 0.1$ での振る舞いを書いたのが3である。これを見ると、自由エネルギーははっきりと折れていて一次相転

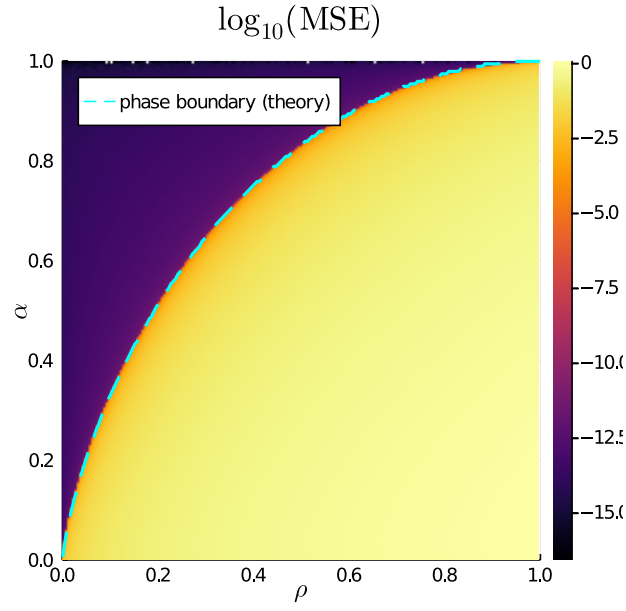


Figure 1: 平均二乗誤差をプロットしたもの。水色の破線で書いた境界線より下側ではMSEが大きくなっている。これは極値条件を反復計算して解いたもの（反復ごとのMSEの変化が 10^{-14} を下回ったら反復終了とした）だが、適当な反復回数で打ち切っているので、復元可能相でも微妙に値が残っている。

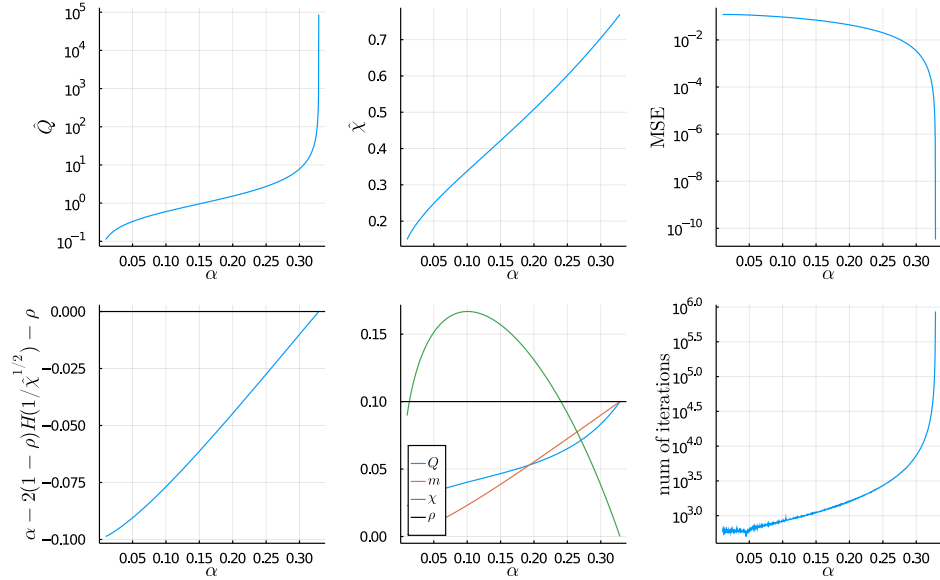


Figure 2: $\rho = 0.1$ の場合における諸量のふるまい。右下の図は極値条件を反復代入で解く際に必要とした反復回数。復元可能相に向けて \hat{Q} や反復回数が発散している。MSEは連続的にゼロに向かっていく。 Q, m, χ は転移点で完全復元相での値（ $\rho, \rho, 0$ ）となる。

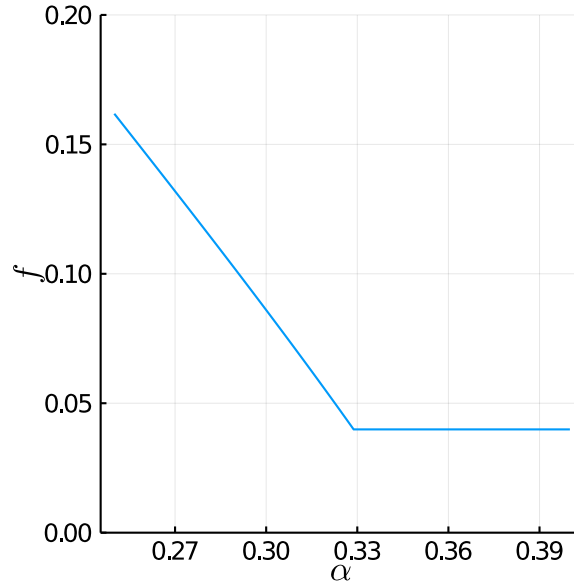


Figure 3: $\rho = 0.1$ での自由エネルギーの振る舞い。1階微分に不連続性が存在することが伺える。

移であるということがわかる。MSEだけ見てみると秩序変数が連続的に立ち上がる連続転移のような気分になるんだけど、そうではないということである。この設定では帯磁率 χ も臨界発散はしない。

5.4 サポート選択の問題について

完全復元相では二乗誤差がゼロになるんだけど、どの変数が非ゼロになるのかという問題を考える。 $q_+ = \mathcal{N}(0, 1)$ の場合に true positive rate と false positive rate を書いたのが図4である。これを見ると、完全復元相に入っても fpr が残ることがわかる。とはいえ、 $\hat{Q} \rightarrow \infty$ なのでこれらからの寄与は無限小であるから、MSEには効かない。これはまあ無限に小さいシグナルがあるからかもしれない。こういうのを刈り取る程の力はL1ノルムにはないということがわかる。真の非ゼロ成分は見つけられるが、偽陽性が多く出てしまうのはL1ノルムを使ったスパース推定一般の問題であって[BVDG11]、これを抑え込むための手法が色々研究されている。

6 summary

L1ノルム最小化に基づく圧縮センシングのレプリカ計算を紹介した。観測行列をランダムにした問題にすることで比較的容易に解析が可能かつ、ほぼ任意の巨視的統計量が計算可能となる。実際に統計量を決定する方程式を調べることで、完全復元可能／不可能相の相転移があることが明らかとなり、さらにL1ノルムを用いた変数選択の性能の一端を調べるのが可能となることを見た。

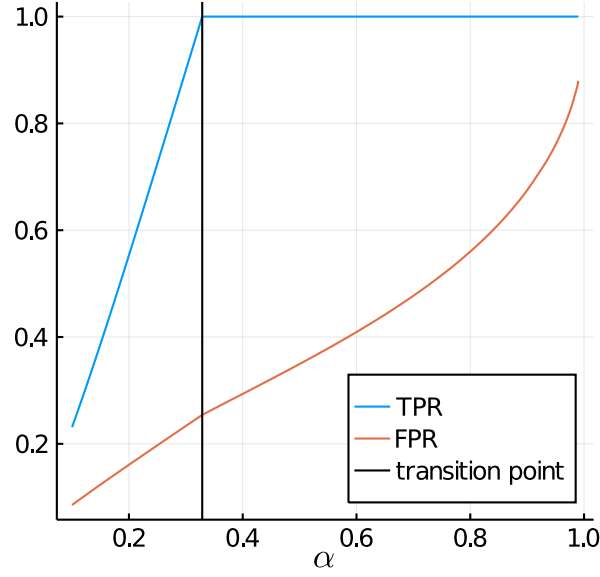


Figure 4: $\rho = 0.1$ でのTPR, FPRの振る舞い。完全復元が可能なところでTPRが1になっており、推定値は真の非ゼロ成分を必ず含むということがわかる。一方、FPRはずっと非ゼロである。

参考文献

- [BVDG11] Peter Bühlmann and Sara Van De Geer, *Statistics for high-dimensional data: methods, theory and applications*, Springer Science & Business Media, 2011.
- [DMM09] David L Donoho, Arian Maleki, and Andrea Montanari, *Message-passing algorithms for compressed sensing*, Proceedings of the National Academy of Sciences **106** (2009), no. 45, 18914–18919.
- [JM14] Adel Javanmard and Andrea Montanari, *Hypothesis testing in high-dimensional regression under the gaussian random design model: Asymptotic theory*, IEEE Transactions on Information Theory **60** (2014), no. 10, 6522–6554.
- [KWT09] Yoshiyuki Kabashima, Tadashi Wadayama, and Toshiyuki Tanaka, *A typical reconstruction limit for compressed sensing based on l_p -norm minimization*, Journal of Statistical Mechanics: Theory and Experiment **2009** (2009), no. 09, L09003.
- [Tal10] Michel Talagrand, *Mean field models for spin glasses: Volume i: Basic examples*, vol. 54, Springer Science & Business Media, 2010.