

# 非凸スパース制約最小化による信号復元： 非凸性制御によるアルゴリズム軌道の安定化

坂田綾香\*, 小淵智之†

Reconstruction of sparse signals by minimization of nonconvex penalties

Ayaka Sakata\* and Tomoyuki Obuchi

本稿では, 圧縮センシングにおける非凸スパース制約を用いた信号復元法について紹介する. ここでは Smoothly Clipped Absolute Deviation(SCAD), Minimax Concave Penalty(MCP) と呼ばれる非凸スパース制約の最小化法を扱う. これらの制約は, 非凸性パラメータと呼ぶ正則化パラメータにより形が変化し, ある極限では  $\ell_1$  制約と一致する. 本稿では, 近似確率伝搬法と呼ばれるアルゴリズムにより非凸制約最小化問題を解くことを考える. 対応する理論解析から, 非凸制約最小化法は  $\ell_1$  制約最小化法よりも高い復元性能を与え, 非凸性パラメータが小さいほどその性能が高くなることが示される. しかし, 非凸性パラメータが小さくなると近似確率伝搬法が収束しない問題が生じる. この問題の背景に, 近似確率伝搬法の固定点への引き込み領域が消失するという現象があることを示し, この現象に由来する収束性の悪さを非凸性制御と呼ぶ方法により部分的に解決する (Sakata and Obuchi (2021)).

In this paper, we introduce a signal recovery method using nonconvex sparse penalties in compressed sensing. We deal with nonconvex sparse penalties called Smoothly Clipped Absolute Deviation (SCAD) and Minimax Concave Penalty (MCP). The form of these penalties change depending on the regularization parameter, which we call the nonconvexity parameter, and SCAD and MCP coincide with the  $\ell_1$  penalty at a certain limit of the nonconvexity parameter. We introduce an approximate message passing (AMP) algorithm to solve the minimization problem of the nonconvex sparse penalties. The corresponding theoretical analysis shows that the nonconvex penalty minimization method gives better recovery performance than the  $\ell_1$ -minimization method as the nonconvexity parameter decreases. However, the small nonconvexity parameter induces the difficulty in the convergence of AMP. We show that this difficulty is caused by the vanishing basin of attraction to the fixed point of AMP, and mitigate the difficulty by introducing a method called non-convexity control (Sakata and Obuchi (2021)).

キーワード: 非凸制約, SCAD, MCP, 確率伝搬法

\* 責任著者, 統計数理研究所 数理・推論研究系: 〒190-8562 東京都立川市緑町 10-3 (E-mail: ayaka@ism.ac.jp)

† 京都大学 大学院情報科学研究科: 〒606-8501 京都府京都市左京区吉田本町 (E-mail: obuchi@i.kyoto-u.ac.jp)

## 1. はじめに

圧縮センシングとは、線形観測により得られた圧縮データから高次元の信号を復元する枠組みの総称である (Candés and Tao (2006), Donoho (2006a)). 数学的には、真の信号を  $\mathbf{x}^0 \in \mathbb{R}^N$ , 観測行列を  $A \in \mathbb{R}^{M \times N}$  として観測  $\mathbf{y} = A\mathbf{x}^0$  を得て、真の信号を観測から復元する問題として定式化される. ここで、観測数  $M$  は信号の次元  $N$  より小さいとし、圧縮率を  $\alpha \equiv M/N < 1$  とする. このような設定は劣決定問題なので、一般に解は一意に決まらない. しかし、真の信号について「有限の割合でゼロ成分がある」というスパース性仮定を導入すると、 $M < N$  であっても真の信号を復元できる場合がある. このスパース性の仮定は自然画像や音声データなど実社会に存在するデータにおいて近似的に成り立つことが多いため、妥当な仮定であると言える (Elad (2010)). また可能な観測数に制限がある場合に、少ない観測回数から対象の信号を復元したいという需要はさまざまな分野に存在し、対象に応じた個別の定式化や数的手法についても研究されてきた (Boche et al. (2015)). 例えば医療診断に用いられる核磁気共鳴画像法 (MRI) (Lustig et al. (2007)) や、複数の電波望遠鏡を統合させたイベントホライズン望遠鏡における画像復元 (The Event Horizon Telescope Collaboration et al (2019)) などが具体的応用例として知られる.

観測結果  $\mathbf{y}$  からスパースな信号を復元する方法の一つは、 $\mathbf{y} = A\mathbf{x}$  という制約を満たす  $\mathbf{x}$  の中で最小の  $\ell_0$  ノルム、すなわち最小の非ゼロ要素数を与える  $\mathbf{x}$  を復元信号とする方法である. これは  $\ell_0$  最小化法と呼ばれ、次のように定式化される.

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, \text{ subject to } \mathbf{y} = A\mathbf{x} \quad (1.1)$$

観測数  $M$  が信号の非ゼロ要素数より大きいとき  $\ell_0$  最小化法は正しく真の信号を復元すると期待されるが、実際に (1.1) を最小化するには可能な非ゼロ要素の位置の組み合わせを全て探索する必要がある. これは NP 困難な問題であり、信号の次元  $N$  が増えるとともに計算量が指数関数的に増大し、実用的ではない. そこで、代わりに (1.1) を凸緩和した  $\ell_1$  最小化法が広く用いられている (Candés and Tao (2006), Chen et al. (2001)).

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ subject to } \mathbf{y} = A\mathbf{x} \quad (1.2)$$

ここで  $\|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i|$  は  $\ell_1$  ノルムである. (1.2) の利点としては、凸最適化法を始めとする様々なアルゴリズムが適用可能であること (Tropp (2007), Donoho et al. (2009a), Chartrand and Yin (2008)), また観測行列に対する条件から (1.1) と (1.2) の解が一致する条件が得られることなどが挙げられる (Gribonval and Nielsen (2003), Candés (2008)).

しかし  $\ell_1$  最小化法の信号復元性能は一般に  $\ell_0$  最小化法に劣る. これは  $\ell_1$  制約が推定量の縮小をもたらすという点に起因する.  $\ell_1$  最小化法の扱いやすさと  $\ell_0$  最小化法の性能の良さ

を両立するため、 $\ell_p$  ( $0 < p < 1$ ) ノルム最小化法なども提案されており、実際  $\ell_p$  ( $0 < p < 1$ ) 最小化法は  $\ell_1$  最小化法よりも少ない観測数で高い信号復元性能を得ることが示唆されている (Chartrand (2007), Chartrand and Staneva (2008)). しかし、 $\ell_p$  ( $0 < p < 1$ ) 最小化による解は入力データに応じて不連続に変化することが知られており、予測性能やアルゴリズムの安定性に懸念がある (Fan and Li (2001)). そこで推定値の縮小の問題と不連続性の問題を同時に解決するため、Smoothly Clipped Absolute Deviation (SCAD) (Fan and Li (2001)) や Minimax Concave Penalty (MCP) (Zhang (2010)) といった区分連続な非凸制約が提案されている. 1.3 節で説明するように、SCAD, MCP は推定量の縮小を軽減し、また入力に対して連続的に振る舞う推定量を与える.

一般に非凸関数を最小化する問題では、多数の局所解の出現に由来して初期条件依存性が生じることが知られる (Agarwal et al. (2016)). 本稿では、圧縮センシングにおける信号復元に対して非凸制約最小化法を用い、近似確率伝搬法 (Approximate Message Passing, AMP) と呼ばれるアルゴリズムを適用して実装する. そして AMP の解析を通して非凸制約特有の問題について考察する. 多数の局所解の出現という問題以外にも、局所解が存在しない領域においてアルゴリズムの収束性に特徴的な難しさが生じることが示し、その解決法である非凸性制御について説明する.

### 1.1 問題設定

本稿では非凸スパース制約を  $J(\mathbf{x}; \eta)$  として示す. 成分ごとに同じ制約が課されているとして  $J(\mathbf{x}; \eta) = \sum_{i=1}^N J(x_i; \eta)$  と表記する.  $\eta$  は非凸制約の形を変えるパラメータであり、ここでは非凸性パラメータと呼ぶ. 非凸制約最小化法による推定値は次のように与えられる.

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} J(\mathbf{x}; \eta), \text{ subject to } \mathbf{y} = A\mathbf{x} \quad (1.3)$$

本稿では  $J(\mathbf{x}; \eta)$  として SCAD, MCP を考える.

### 1.2 SCAD, MCP とは

SCAD 正則化は次のように定義される (Fan and Li (2001)).

$$J(x; \eta) = \begin{cases} \lambda|x| & (|x| \leq \lambda) \\ -\frac{x^2 - 2a\lambda|x| + \lambda^2}{2(a-1)} & (\lambda < |x| \leq a\lambda) \\ \frac{(a+1)\lambda^2}{2} & (|x| > a\lambda) \end{cases} \quad (1.4)$$

非凸性パラメータは  $\eta = \{\lambda, a\}$  ( $a > 1$ ) として与えられ、また  $\lambda \in (0, \infty)$ ,  $a \in (1, \infty)$  である. 図 1 (a) は  $\lambda = 1$ ,  $a = 3$  での SCAD 制約であり、点線は関数形が切り替わる場所 ( $|x| = \lambda$ ,  $|x| = a\lambda$ ,  $|x| = a\lambda$ ) を示している. SCAD 正則化は  $a \rightarrow \infty$  極限で  $\ell_1$  制約と一致し、また  $\lambda \rightarrow \infty$  における SCAD 最小化は  $\ell_1$  最小化と等価である.

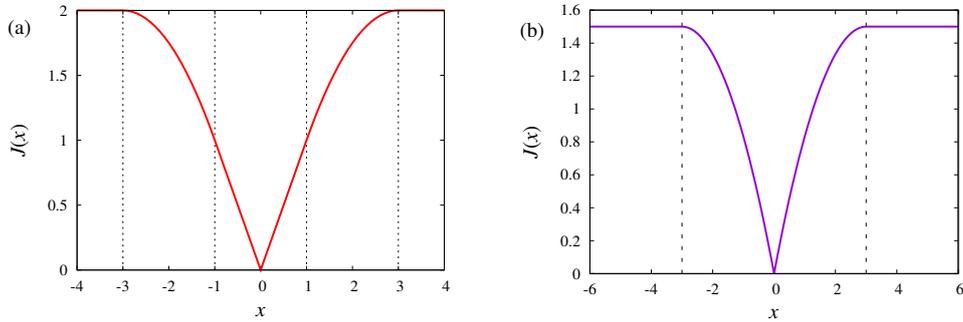


図1  $\lambda = 1$ ,  $a = 3$  での (a) SCAD と (b) MCP. 点線は関数の形が変わるところを示す.

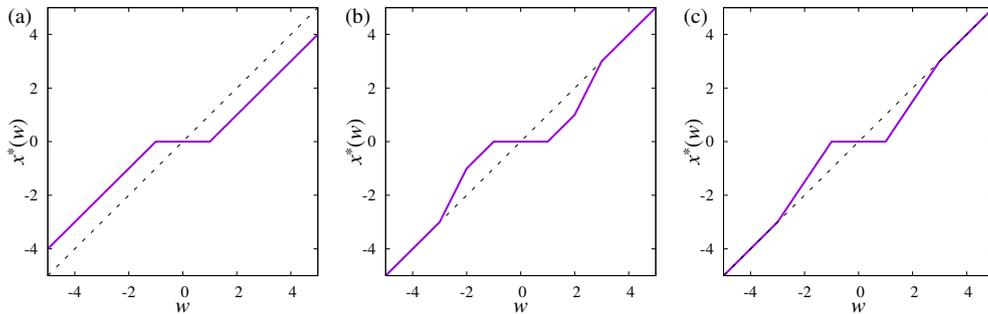


図2 (a)  $\ell_1$  制約下の推定量, (b) SCAD 制約下の推定量, (c) MCP 下の推定量. 点線は最小二乗推定量を表す.

MCP は次のように定義される (Zhang (2010)).

$$J(x; \eta) = \begin{cases} \lambda|x| - \frac{x^2}{2a} & (|x| \leq a\lambda) \\ \frac{a\lambda^2}{2} & (|x| > a\lambda) \end{cases} \quad (1.5)$$

SCAD と同じく  $\eta = \{\lambda, a\}$  であり,  $\lambda \in (0, \infty)$ ,  $a \in (1, \infty)$  である. 図 1 (b) に  $\lambda = 1$ ,  $a = 3$  での MCP を示す. 点線は関数の形が切り替わる点  $|x| = a\lambda$  を示している. SCAD と同様に, MCP も極限  $a \rightarrow \infty$  で  $\ell_1$  制約に一致し, また  $\lambda \rightarrow \infty$  における MCP 最小化は  $\ell_1$  最小化と等価である.

### 1.3 SCAD, MCP のもとでの推定値

SCAD, MCP 制約下の推定値の振る舞いを理解するため, 次の 1 次元の問題を考えてみよう.

$$\hat{x}(s, w) = \arg \min_x \left\{ \frac{(x-w)^2}{2s} + J(x; \eta) \right\} \quad (1.6)$$

これは制約  $J(x; \eta)$  のもとで入力  $w$  をガウスモデルに当てはめる問題であり,  $s > 0$  とする. SCAD, MCP とともに上に凸の部分を含むため, 全ての入力データに対して (1.6) が解を持つためには, (1.6) の二次項の係数を考慮して適切な  $s$  および  $a$  の領域を考える必要がある.

SCAD の場合は  $a > 1 + s$  が成立する必要がある, この条件が成立しているとき (1.6) の解は次のように与えられる.

$$\hat{x}(s, w) = \mathcal{V}(s, w/s)\mathcal{W}(s, w/s) \quad (1.7)$$

$\mathcal{W}$  と  $\mathcal{V}$  は次のように与えられる.

$$\mathcal{W}(s, w) = \begin{cases} w - \text{sgn}(w)\lambda & \text{for } \lambda(1 + s^{-1}) \geq |w| > \lambda \\ w - \text{sgn}(w) \frac{a\lambda}{a-1} & \text{for } a\lambda s^{-1} \geq |w| > \lambda(1 + s^{-1}) \\ w & \text{for } |w| > a\lambda s^{-1} \\ 0 & \text{otherwise} \end{cases} \quad (1.8)$$

$$\mathcal{V}(s, w) = \begin{cases} s & \text{for } \lambda(1 + s^{-1}) \geq |w| > \lambda \\ \left(s^{-1} - \frac{1}{a-1}\right)^{-1} & \text{for } a\lambda s^{-1} \geq |w| > \lambda(1 + s^{-1}) \\ s & \text{for } |w| > a\lambda s^{-1} \\ 0 & \text{otherwise} \end{cases} \quad (1.9)$$

$\text{sgn}(w)$  は  $w$  の符号を意味する.  $s = 1$ ,  $\lambda = 1$ ,  $a = 3$  における SCAD 制約下での推定量を, 入力  $w$  の関数として図 2 (b) に示す. 比較のため, 最小二乗推定量 (制約項がない場合の推定量) を点線で示し, また図 2 (a) に  $l_1$  制約のもとでの推定量を示す. SCAD 制約下の推定量は,  $\lambda(1 + s^{-1}) \geq |w| > \lambda$  で  $l_1$  制約下の推定量のように振る舞う. また  $|w| > a\lambda s^{-1}$  では最小二乗推定量のように振る舞う.  $a\lambda s^{-1} \geq |w| > \lambda(1 + s^{-1})$  では,  $l_1$  的な推定量と最小二乗推定量が連続的につながる. この  $|w| > a\lambda s^{-1}$  での最小二乗推定量と同じ振る舞いが SCAD の特徴であり, 十分大きい  $w$  においては損失項  $(\hat{x}(w) - w)^2/2s$  の値がゼロになる. この性質により,  $l_1$  制約下の推定量よりも高いデータ記述性能を持つと考えられる.

MCP の場合は, 全ての入力について (1.6) が解を持つためには  $a > s$  である必要がある. MCP の定義にある  $a > 1$  という条件と合わせると, 満たすべき条件は  $a > \min\{1, s\}$  である. この条件が成立するとき, MCP 下の推定値は SCAD と同様に (1.7) の形で与えられるが,  $\mathcal{W}$  と  $\mathcal{V}$  はそれぞれ次のように与えられる (Sakata (2018)).

$$\mathcal{W}(s, w) = \begin{cases} w - \text{sgn}(w)\lambda & \text{for } a\lambda s^{-1} \geq |w| > \lambda \\ w & \text{for } |w| > a\lambda s^{-1} \\ 0 & \text{otherwise} \end{cases} \quad (1.10)$$

$$\mathcal{V}(s, w) = \begin{cases} (s^{-1} - a^{-1})^{-1} & \text{for } a\lambda s^{-1} \geq |w| > \lambda \\ s & \text{for } |w| > a\lambda s^{-1} \\ 0 & \text{otherwise} \end{cases} \quad (1.11)$$

図 2 (c) は  $s = 1$ ,  $\lambda = 1$ ,  $a = 3$  での MCP 下の推定値の振る舞いである. SCAD と同様に,  $|w| > a\lambda s^{-1}$  において MCP は最小二乗推定量と同様に振る舞い,  $a\lambda s^{-1} \geq |w| > \lambda$  の

領域でゼロから連続的につながる.  $l_1$  的な振る舞いをする領域が存在しない点が SCAD との違いである.

## 2. 確率伝搬法による推定値の評価

### 2.1 準備

SCAD, MCP 制約下での (1.3) を解くにあたり, ベイズ推定の枠組みで問題を定式化しておく. まず尤度と事前分布を次のように与える.

$$f(\mathbf{x}|\mathcal{D}; \tau) = \frac{1}{(\sqrt{2\pi\tau})^M} \exp\left(-\frac{1}{2\tau}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2\right) \quad (2.1)$$

$$\phi(\mathbf{x}; \beta) = \frac{1}{\mathcal{N}_\phi} \exp(-\beta J(\mathbf{x}; \eta)) \quad (2.2)$$

$\mathcal{D} = \{\mathbf{y}, \mathbf{A}\}$  とした. また  $\tau, \beta$  はパラメータであり, 後ほど適切な極限操作を行う. 尤度は  $\mathbf{y} \sim \mathbf{A}\mathbf{x}$  という制約を表し,  $\tau \rightarrow 0$  で厳密に  $\mathbf{y} = \mathbf{A}\mathbf{x}$  を満たす  $\mathbf{x}$  上の一様分布となる. 事前分布は SCAD, MCP 制約を表し,  $\beta \rightarrow \infty$  で最小の制約の値を与える  $\mathbf{x}$  のみサンプリングされる. また  $\mathcal{N}_\phi$  は規格化定数である. ベイズの定理より, 事後分布は次のように与えられる.

$$P(\mathbf{x}|\mathcal{D}; \tau, \beta) = \frac{f(\mathbf{x}|\mathcal{D}; \tau)\phi(\mathbf{x}; \beta)}{Z(\mathcal{D}; \tau, \beta)} \quad (2.3)$$

$Z(\mathcal{D})$  は規格化定数であり, 次のように与えられる.

$$Z(\mathcal{D}; \tau, \beta) = \int d\mathbf{x} f(\mathbf{x}|\mathcal{D}; \tau)\phi(\mathbf{x}; \beta) \quad (2.4)$$

また周辺事後平均を次のように定義しておく.

$$\bar{x}_i(\mathcal{D}; \tau, \beta) = \int d\mathbf{x} x_i P(\mathbf{x}|\mathcal{D}; \tau, \beta) \quad (2.5)$$

周辺事後平均を用いると, (1.3) の推定値の各成分は以下のように与えられる.

$$\hat{x}_i(\mathcal{D}) = \lim_{\beta \rightarrow \infty} \lim_{\tau' \rightarrow 0} \bar{x}_i(\mathcal{D}; \tau' \beta^{-1}, \beta) \quad (2.6)$$

(1.3) では  $\mathbf{y} = \mathbf{A}\mathbf{x}$  という条件の下で SCAD, MCP を最小化することを考えているため,  $\beta \rightarrow \infty$  において  $\tau' \equiv \tau\beta \rightarrow 0$  という極限をとる必要がある.

### 2.2 近似確率伝搬法による近似計算

(2.5) を単純に評価するには, 指数関数オーダーの計算が必要となる (Natarajan (1995)). そこで, ここでは近似確率伝搬法 (Approximate message passing, AMP) と呼ばれるアルゴリズムによって近似評価を行う (Mézard et al. (1987), Mezard and Montanari (2009), MacKay

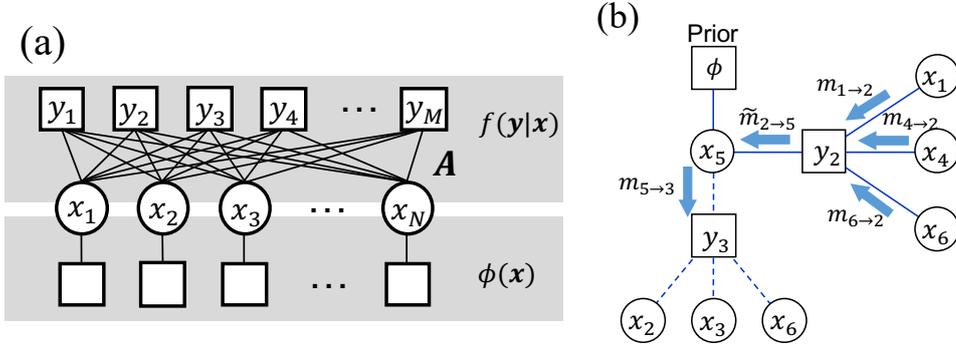


図3 (a) 事後分布のグラフィカルな表現. (b) 局所的ツリー近似とメッセージの例.

(1999), Kabashima and Saad (1998)). AMP の一般系は generalized approximate message passing (GAMP) (Rangan (2011), Kabashima and Uda (2004)) と呼ばれ, GAMP はさまざまな推定問題に対して適用可能である. 本稿では GAMP とその SCAD, MCP 最小化法への応用例を説明する. 導出の詳細は付録 A に譲り, ここでは概要を説明する.

### 2.2.1 確率伝搬法の一般形

AMP とは, 確率伝搬法 (Message Passing, MP) というより一般的なアルゴリズムにおいて, 近似を行うことで計算量を削減したものである. ここでは導入としてまず MP の解説を行う.

MP においては, 変数ノードと因子ノードと呼ばれる二種類のノードと, それらをつなぐエッジによって, 確率モデルを二分グラフとして表現する (図 3 (a)) (Rangan (2011), Krzakala et al. (2012)). 因子ノードと変数ノードはそれぞれデータ  $\mathbf{y}$  と推定される変数  $\mathbf{x}$  を表し, 因子ノードと変数ノードを繋ぐエッジは観測行列  $A$  に対応する. 因子ノード  $y_\mu$  と変数  $x$  およびエッジ  $A_{\mu i}$  ( $i = 1, \dots, N$ ) は  $y_\mu = \sum_{i=1}^N A_{\mu i} x_i$  という関係を表している. 事後分布に対する MP は, 一般に  $2MN$  個の “メッセージ”  $m_{i \rightarrow \mu}(x_i)$  と  $\tilde{m}_{\mu \rightarrow i}(x_i)$  ( $i = 1, 2, \dots, N, \mu = 1, 2, \dots, M$ ) により与えられる. ここでの尤度 (2.1) は成分ごとの積で表現できるため,  $\mu$  番目のデータに関する尤度を  $f(y_\mu | (A\mathbf{x})_\mu)$  と書くことにすると, メッセージは次のように定義される.

$$\tilde{m}_{\mu \rightarrow i}(x_i) = \frac{1}{\tilde{Z}_{\mu \rightarrow i}} \int \prod_{j \neq i} dx_j f(y_\mu | (A\mathbf{x})_\mu) \prod_{j \neq i} m_{j \rightarrow \mu}(x_j) \quad (2.7)$$

$$m_{i \rightarrow \mu}(x_i) = \frac{1}{Z_{i \rightarrow \mu}} \phi(x_i) \prod_{\gamma \neq \mu} \tilde{m}_{\gamma \rightarrow i}(x_i) \quad (2.8)$$

ここで  $\tilde{Z}_{\mu \rightarrow i}$  と  $Z_{i \rightarrow \mu}$  は規格化定数である. これらのメッセージは, グラフのエッジ上で図 3(b) のように定義される. (2.7), (2.8) は, ツリー近似されたグラフ上で評価された一体分布である. たとえば図 3(b) の  $\tilde{m}_{2 \rightarrow 5}(x_2)$  は, 因子ノード  $y_2$  につながる  $x_5$  以外の変数

が一体分布  $m_{i \rightarrow 2}(x_i)$  に従っているとしたり、観測によりデータ  $y_2$  を得たときの  $x_2$  の周辺事後分布を表す。また  $m_{5 \rightarrow 3}$  は、3 番目の観測結果  $y_3$  を得る前の  $x_5$  の周辺事後分布である。二分グラフが真にツリー構造である場合は、(2.7)、(2.8) に従って適当な端から変数の状態和をとることで全ての  $(i, \mu)$  に対するメッセージを得ることができる。周辺事後分布はメッセージを用いて次のように表される。

$$P_i(x_i) = \frac{1}{Z_i} \phi(x_i) \prod_{\mu=1}^M m_{\mu \rightarrow i}(x_i) \quad (2.9)$$

$Z_i$  は規格化定数である。(2.9) を用いて周辺事後平均を評価すれば、推定値 (1.3) を得ることができる。(2.9) は真にツリーなグラフにおいては厳密な周辺事後分布を与える。一方でツリー構造ではないときは、(2.7) と (2.8) を再帰的に更新し、収束したメッセージを用いて周辺事後分布を近似する。場合によっては収束しないこともあるため、アルゴリズムの挙動を理論的に把握することが重要である。この点については 2.3 節で触れる。

### 2.2.2 AMP の一般形

メッセージ (2.7) と (2.8) は合計で  $2M \times N$  個存在するが、 $N$  が十分大きく、観測行列の各成分が  $A_{\mu i} \sim O(N^{-1/2})$  であり、また成分間の相関が無視できるときは、付録 A に示すように中心極限定理から  $\tilde{m}_{\mu \rightarrow i}$  はガウス分布と考えることができ、そのため、その平均と分散のみを考えれば良い。AMP は、メッセージそのものではなく、その平均と分散を更新するアルゴリズムとして与えられる。AMP では、周辺事後分布  $P_i(x_i)$  は二つのパラメータ  $\Sigma^2$  と  $R$  を用いて次のように与えられる。

$$P_i(x) = \mathcal{M}(x; \Sigma^2, R_i) \quad (2.10)$$

ここで  $\mathcal{M}$  は次のように定義される確率分布である。

$$\mathcal{M}(x; \Sigma^2, R) = \frac{1}{Z_M(\Sigma^2, R)} \phi(x) \frac{1}{\sqrt{2\pi\Sigma^2}} \exp\left(-\frac{(x-R)^2}{2\Sigma^2}\right) \quad (2.11)$$

また  $Z_M(\Sigma^2, R)$  は規格化定数である。(2.11) は、事後分布のうち、変数  $x_i$  に関する事前分布を除いた部分が平均  $R_i$ 、分散  $\Sigma^2$  のガウス分布により表現されていることを意味する。 $R_i$ 、 $\Sigma^2$  は次のように与えられる。

$$\Sigma^2 = \left( \frac{1}{N} \sum_{\mu} (g'_{\text{out}})_{\mu} \right)^{-1} \quad (2.12)$$

$$R_i = a_i + \left( \sum_{\mu} (g_{\text{out}})_{\mu} A_{\mu i} \right) \Sigma^2 \quad (2.13)$$

$a_i$  は周辺事後分布平均であり,  $g_{\text{out}}, g'_{\text{out}}$  は次のように与えられる.

$$(g_{\text{out}})_{\mu} = \frac{\partial}{\partial \omega_{\mu}} \log \left( \int du_{\mu} f(y_{\mu} | u_{\mu}) \exp \left( -\frac{(u_{\mu} - \omega_{\mu})^2}{2V} \right) \right) \quad (2.14)$$

$$(g'_{\text{out}})_{\mu} = -\frac{\partial^2}{\partial \omega_{\mu}^2} \log \left( \int du_{\mu} f(y_{\mu} | u_{\mu}) \exp \left( -\frac{(u_{\mu} - \omega_{\mu})^2}{2V} \right) \right) \quad (2.15)$$

$$V = \frac{1}{N} \sum_{i=1}^N v_i^2 \quad (2.16)$$

$$\omega_{\mu} = \sum_i A_{\mu i} a_i - (g_{\text{out}})_{\mu} V. \quad (2.17)$$

$v_i^2$  は  $x_i$  の周辺事後分布の分散である. 後々の議論のため, 周辺事後分布の平均と分散を次のように定義しておく,

$$f_a(\Sigma^2, R) \equiv \int dx x \mathcal{M}(x; \Sigma^2, R), \quad (2.18)$$

$$f_c(\Sigma^2, R) \equiv \int dx x^2 \mathcal{M}(x; \Sigma^2, R) - f_a^2(\Sigma^2, R). \quad (2.19)$$

$a_i, v_i^2$  は次のように表現される.

$$a_i = f_a(\Sigma^2, R_i) \quad (2.20)$$

$$v_i = f_c(\Sigma^2, R_i) \quad (2.21)$$

以上の一般形から, (2.14), (2.15) に含まれる尤度  $f(y|u)$  と, (2.18) と (2.19) に含まれる事前分布  $\phi(x)$  を適当に設定すれば, AMP をさまざまな問題に応用することができる.

### 2.2.3 SCAD, MCP 制約つき線形回帰における AMP

SCAD, MCP 制約下での (1.3) に対する AMP を一般形から導出してみよう. そのためには, 線形制約  $\mathbf{y} = A\mathbf{x}$  に対応する尤度と, SCAD, MCP 最小化に対応する事前分布を代入すれば良い. まず, 尤度  $f(y_{\mu} | u_{\mu}) \propto \exp \left\{ -\frac{1}{2\tau} (y_{\mu} - u_{\mu})^2 \right\}$  を代入しよう. 後ほど  $\beta$  に関する極限を取ることを考慮して, 変数を  $\hat{V} = \beta V, \hat{\Sigma}^2 = \beta \Sigma^2$  とスケールしておく. これにより  $\hat{V}, \hat{\Sigma}^2$  は  $\beta \rightarrow \infty$  極限において  $O(1)$  となる. 尤度を (2.14) と (2.15) に代入すると,  $(g_{\text{out}})_{\mu}$  と  $(g'_{\text{out}})_{\mu}$  は次のように与えられる.

$$(g_{\text{out}})_{\mu} = \frac{y_{\mu} - \omega_{\mu}}{\beta^{-1} \hat{V} + \tau} \quad (2.22)$$

$$(g'_{\text{out}})_{\mu} = \frac{1}{\beta^{-1} \hat{V} + \tau} \quad (2.23)$$

(2.22) を (2.17) に代入し, また (2.23) を (2.12) に代入すると次の式を得る.

$$\omega_{\mu} = \sum_i A_{\mu i} a_i - \frac{\hat{V}}{\hat{V} + \tau'} (y_{\mu} - \omega_{\mu}) \quad (2.24)$$

$$\hat{\Sigma}^2 = \hat{V} + \tau' \quad (2.25)$$

次に、事前分布  $\phi(x) \propto \exp(-\beta J(x; \eta))$  を (2.11) に代入して  $\beta \rightarrow \infty$  極限をとると次の表現を得る。

$$Z(\hat{\Sigma}^2, R) = \sqrt{\frac{\beta}{2\pi\hat{\Sigma}^2}} \int_{-\infty}^{\infty} \exp \left\{ -\beta J(x; \eta) - \beta \frac{(x-R)^2}{2\hat{\Sigma}^2} \right\} dx \rightarrow \sqrt{\frac{\beta}{2\pi\hat{\Sigma}^2}} \exp \left\{ -\beta \psi^*(\hat{\Sigma}^2, R) \right\} \quad (2.26)$$

ここで以下のように定義した。

$$\psi^*(\hat{\Sigma}^2, R) = \min_x \psi(x; \hat{\Sigma}^2, R), \quad (2.27)$$

$$\psi(x; \hat{\Sigma}^2, R) = J(x; \eta) + \frac{(x-R)^2}{2\hat{\Sigma}^2} \quad (2.28)$$

また (2.22) と (2.25) を (2.13) に代入すると  $R$  は次のように与えられる。

$$R_i = \sum_{\mu} A_{\mu i} (y_{\mu} - \omega_{\mu}) + a_i \quad (2.29)$$

$\beta \rightarrow \infty$  での  $f_a$ (2.18) は  $\psi$  の最小値に対応する  $x$  の値である。

$$f_a(\Sigma^2, R) = \arg \min_x \psi(x; \Sigma^2, R) \quad (2.30)$$

(2.30) は (1.6) と同じ問題である。よって平均と分散は次のように与えられる。

$$f_a(\tilde{\Sigma}^2, R) = \mathcal{V}(\tilde{\Sigma}^2, R/\tilde{\Sigma}^2) \mathcal{W}(\tilde{\Sigma}^2, R/\tilde{\Sigma}^2) \quad (2.31)$$

$$f_c(\tilde{\Sigma}^2, R) = \mathcal{V}(\tilde{\Sigma}^2, R/\tilde{\Sigma}^2) \quad (2.32)$$

以上の変数を再帰的に更新するのが SCAD, MCP 最小化に対する AMP であり、まとめると次の手続きからなる。上つきの  $(t)$  は  $t$  回目の更新時の値を意味する。

$$\hat{V}^{(t)} = \frac{1}{N} \sum_{i=1}^N v_i^{(t-1)} \quad (2.33)$$

$$\omega_{\mu}^{(t)} = \sum_{i=1}^N A_{\mu i} a_i^{(t-1)} - \frac{\hat{V}^{(t)}}{\hat{V}^{(t)} + \tau'} (y - \omega^{(t-1)}) \quad (2.34)$$

$$\Sigma^{(t)2} = \alpha^{-1} (\hat{V}^{(t)} + \tau') \quad (2.35)$$

$$R_i^{(t)} = a_i^{(t-1)} + \sum_{\mu=1}^M A_{\mu i} (y_{\mu} - \omega_{\mu}^{(t)}) \quad (2.36)$$

$$a_i^{(t)} = f_a(\Sigma^{(t)2}, R_i^{(t)}) \quad (2.37)$$

$$v_i^{(t)} = f_c(\Sigma^{(t)2}, R_i^{(t)}) \quad (2.38)$$

ここで  $M/N = \alpha$  とした。以上の表現は有限の  $\beta, \tau$  でのものであるが、(1.3) の評価のためには  $\beta$  と  $\tau$  について極限をとらなくてはならない。  $\tau \ll \beta^{-1}$  であるので  $\tau' \rightarrow 0$  とした AMP を用いることで、SCAD, MCP 最小化法の推定値を得ることができる。

## 2.3 AMP の収束性

AMP は信号のスパース度や観測行列の性質、圧縮率や非凸性パラメータの値によって収束性に差がある。ここでは AMP の性質を理論的に記述する方法についてまとめておく。

### 2.3.1 巨視的な収束性：状態発展法

AMP の軌道はデータに依存する確率変数であるが、 $N$ 、 $M$  が十分に大きく  $A$  の各成分が独立で  $A_{\mu i} \sim O(N^{-1/2})$  のとき、典型的な軌道は二つの統計量の時間発展として表される。以下では  $M/N = \alpha \sim O(1)$  の場合を考える。この統計量の時間発展方程式は状態発展法 (State Evolution, SE) と呼ばれ、AMP の典型的振る舞いを把握する上で有用である。付録 B の方法に従い、AMP の軌道をデータに関して平均化すると、次の二変数からなる状態発展方程式が得られる (Sakata and Xu (2018), Krzakala et al. (2012)).

$$V^{(t+1)} = \int dx^0 \phi(x^0) \int Dz f_c(\alpha^{-1}(V^{(t)} + \tau'), x^0 + z\sqrt{\alpha^{-1}\varepsilon^{(t)}}), \quad (2.39)$$

$$\varepsilon^{(t+1)} = \int dx^0 \phi(x^0) \int Dz \left[ f_a(\alpha^{-1}(V^{(t)} + \tau'), x^0 + z\sqrt{\alpha^{-1}\varepsilon^{(t)}}) - x^0 \right]^2 \quad (2.40)$$

ここで  $\int Dz = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2}$  とした。密度発展法の変数  $V^{(t)}$  と  $\varepsilon^{(t)}$  は、それぞれ AMP の変数と  $V^{(t)} = E_{\mathcal{D}}[\hat{V}^{(t)}]$ ,  $\varepsilon^{(t)} = E_{\mathcal{D}}[\hat{\varepsilon}^{(t)}]$  として対応づけられる。ここで  $\hat{\varepsilon}^{(t)} = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i^{(t)} - x_i^0)^2$  は AMP において  $t$  回の更新を行った際の推定値と真の信号の MSE である。

SE の時間発展 (SE flow) は、データに依存する AMP の典型的軌道であり、あらゆるデータに対する AMP の軌道が必ずしも SE に一致するわけではない。しかし  $N$ 、 $M$  が十分大きく、観測行列  $A$  の各成分が i.i.d. であるなどの良い性質を持つとき、大数の法則から AMP の軌道は SE flow に収束すると期待される。よって、固定されたデータのもとの AMP の振る舞いを理解する上で、SE は理論的根拠となる (Montanari and Bayati (2011)). 特に、 $V$  と  $\varepsilon$  の意味から、 $V = \varepsilon = 0$  が SE の大域的固定点であれば、AMP によって真のパラメータが復元できることを意味する<sup>1)</sup>。よって SE において  $V = \varepsilon = 0$  が解であるかどうかを確認することが、AMP の性質を見る上でまず重要である。

### 2.3.2 微視的な収束性：固定点の線形安定性

SE は  $V^{(t)}$ 、 $\varepsilon^{(t)}$  という統計量を用いて AMP の巨視的性質を記述するが、統計量のレベルで収束していても、微視的なレベル、たとえば  $R_{i \rightarrow \mu}^{(t)}$  などの値そのものが収束しているとは限らない。ここでは AMP の微視的な収束性を議論する。SE における統計量  $V$ 、 $\varepsilon$  の空間に含まれるある微視的固定点があったとして、その周辺での線形安定性解析を考えよう (Kabashima (2003)). 付録 C に示す方法に従うと、次の条件に従うとき、AMP は微視

<sup>1)</sup> 問題によっては  $V = \varepsilon = 0$  以外の固定点も同時に存在することがあり、そのような場合にはあらゆる初期条件から真のパラメータを推定することはできない。しかし本稿で扱う問題設定では、 $V = \varepsilon = 0$  が SE の固定点であるとき、他の解が存在することはない。

的なレベルで安定であると言える.

$$\frac{\alpha}{(V + \tau')^2} \int dx^0 P_0(x^0) \int Dz \left( \frac{\partial f_a(\alpha^{-1}(V + \tau'), x^0 + z\sqrt{\alpha^{-1}\varepsilon})}{\partial(z\sqrt{\alpha^{-1}\varepsilon})} \right)^2 < 1 \quad (2.41)$$

SE の巨視的固定点において (2.41) が満たされているのであれば, その固定点内の 1 点に AMP は収束する. 一方で, SE がある巨視的な固定点に収束するものの, その巨視的な固定点で (2.41) が満たされていないとき, AMP は微視的には収束せず, 統計量  $V, \varepsilon$  の空間内で常に動いている状態である. 微視的収束性を把握するためには各 SE の固定点で (2.41) が満たされているかどうかを確認しなくてはならない. ただし  $V = \varepsilon = 0$  が SE の固定点であるときは (2.41) はただちに満たされる.  $\hat{V}$  および  $\hat{\varepsilon}$  の各成分,  $v_i$  と  $(\hat{x}_i - x_i^0)^2$  は常に非負であるため, その和がゼロになるということは全成分がゼロであることを意味する. したがって  $V = \varepsilon = 0$  であるとき, これを満たす微視的状态は 1 つしかなく,  $V = \varepsilon = 0$  が常に微視的に安定であることと整合する.

微視的な収束性条件 (2.41) は, スピングラス理論における de Almeida-Thouless 条件との対応から, 指数関数個の局所解の出現を意味することが知られている (De Almeida and Thouless (1978), Kabashima (2003)). したがって, 今の問題でも (2.41) が破れた場合, たくさんの局所解が出現していると考えられる.

### 3. AMP と SE の不一致と非凸性制御による改善

本章では, 前節で紹介した AMP とその性能解析法を用いて, 実際に SCAD, MCP 最小化法の性能を評価してみる. 以下で示す性質は SCAD, MCP について共通であるため, 主に SCAD を用いて説明する.

#### 3.1 密度発展方程式から見る引き込み領域の縮小

図 4 は AMP により得た, SCAD 最小化による推定値を示している. 真の信号の各成分は次の分布に従って生成した.

$$P_0(x_i) = (1 - \rho)\delta(x_i^0) + \frac{\rho}{\sqrt{2\pi}} \exp\left(-\frac{x_i^0{}^2}{2}\right) \quad (3.1)$$

また観測行列の各成分は, 平均 0, 分散  $1/N$  のガウス分布に従って生成した. 図 4 の実線と  $\circ$  はそれぞれ真の信号と推定値とした. (a) は  $\lambda = 1, a = 3, \rho = 0.19$ , (b) は  $\lambda = 0.3, a = 3, \rho = 0.28$  の結果である. (a) では高い精度で真のパラメータを推定できている. このパラメータ領域では, SE においても  $V = \varepsilon = 0$  が固定点となることが確認できる. (a) のパラメータ領域では  $\ell_1$  最小化法は失敗することが知られており (Donoho (2006b), Donoho and Tanner (2009), Kabashima et al. (2009)), 図 4(a) は非凸性制御を用いた推定性能改善の一例である. (b) は (a) より密な信号に対して, AMP による 10 回の更新後の推定値

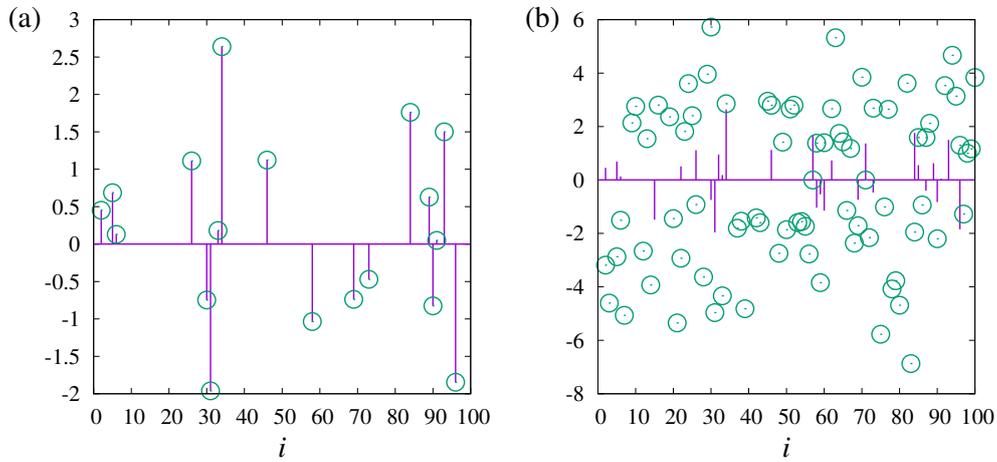


図4  $N = 100$ ,  $M = 50$ での真の信号  $x_i^{(0)}$  (実線) と SCAD 最小化により推定した推定値  $\hat{x}_i$  (○) の比較. (a) は  $\rho = 0.19$ ,  $\lambda = 1$ ,  $a = 3$ . (b) は  $\rho = 0.28$ ,  $\lambda = 0.3$ ,  $a = 3$  の場合. (b) では, AMP による更新を繰り返すと発散するため, 10 回の更新後の結果をプロットしている.

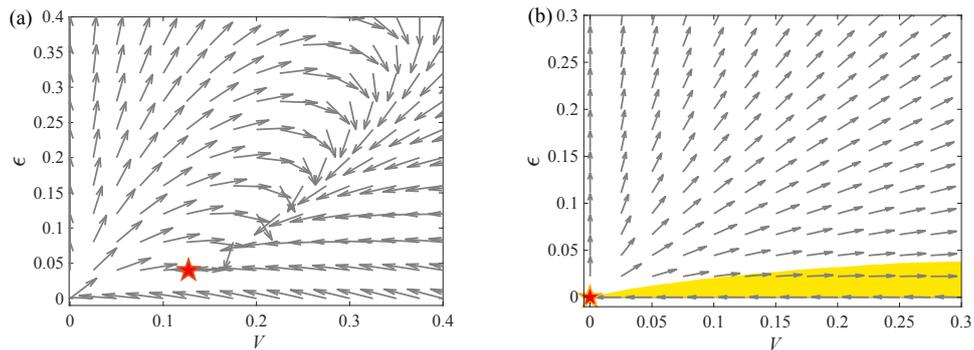


図5  $\alpha = 0.5$ ,  $\rho = 0.28$ での SCAD の SE flow. (a) は  $\lambda = 1$ ,  $a = 3$ , (b) は  $\lambda = 0.3$ ,  $a = 3$ . 星は SE の固定点を示す. また (b) の影付きの領域は原点への引き込み領域を表す.

を示している. 推定値は十分な更新の後でも真の信号と一致せず, 発散してしまう. 一方でこのパラメータ領域 ( $\alpha = 0.5, \rho = 0.28, \lambda = 0.3, a = 3$ ) において SE の固定点を評価すると,  $\varepsilon = V = 0$  は大域的安定解であるため, 原理的には真の信号を復元できるはずである. また微視的安全性も保証されているため, AMP と SE の不一致は局所解の出現に由来するものではない. 図 4(b) の例に限らず, 真の信号が密である場合, SE では  $V = \varepsilon = 0$  が大域的固定点であるにもかかわらず AMP が収束せずに発散する傾向にある.

このような SE と AMP の挙動の不一致性について議論するため,  $\alpha = 0.5$ ,  $\rho = 0.28$ での SE flow を見てみよう. このパラメータ領域では, 図 4(b) に示すように  $\lambda$  が小さい場合 AMP が発散してしまう. 図 5(a) は  $\lambda = 1$ ,  $a = 3$ , (b) は  $\lambda = 0.3$ ,  $a = 3$ での SE flow である. 座標  $(\hat{V}, \hat{\varepsilon})$  上の矢印は規格化されたベクトル  $(V^{(t+1)} - V^{(t)}, \varepsilon^{(t+1)} - \varepsilon^{(t)})$  であり, その

点における SE flow の方向を示す. また SE の固定点を星印で示した. 図 5 (a) に示すように,  $\lambda$  が十分大きいとき SE は有限の  $\varepsilon$ ,  $V$  に固定点を持つ. ただし, この巨視的固定点では微視的安定性条件 (2.41) が破れており, 巨視的には収束するものの微視的には収束しない.  $\lambda$  を下げると, 図 5(b) に示すように  $V = \varepsilon = 0$  が大域的固定点となるが,  $V = \varepsilon = 0$  へ向かう流れは, ほぼ  $V$  軸上にのみ存在する. 十分な回数の更新ののち  $V = \varepsilon = 0$  に収束する初期条件の集合, すなわち引き込み領域 (Basin of Attraction, BOA) を, 図 5(b) の影付きの領域として表している. 矢印は SE による 1 ステップでの変化を表しているので, 矢印が原点を向かない点  $(V, \varepsilon)$  が BOA に含まれることに矛盾はない.  $V - \varepsilon$  平面上の BOA の面積は, 非凸性パラメータ  $\lambda$  が小さくなるにつれて縮小していく. 一方で SE の固定点を見ると,  $\lambda$  が小さくなるほど, 密な信号に対しても  $V = \varepsilon = 0$  が安定であることがわかる. したがって, BOA が縮小するという問題を解決できれば, SCAD, MCP 最小化法は高い性能を発揮できると期待される.  $\lambda \rightarrow \infty$  に対応する  $l_1$  最小化の場合は,  $V = \varepsilon = 0$  への BOA が非常に広いことが数値的に確認できるため, BOA の縮小は  $\lambda$  の減少に起因すると考えられる. 非常に小さい  $\varepsilon$  を AMP の初期条件とすることができれば BOA の縮小は問題にならないが, そもそも  $\varepsilon$  は MSE であるため, 非常に小さい  $\varepsilon$  とは真の信号に極めて近い状態を意味しており, そのような初期条件を用意することは一般に困難である.

### 3.1.1 非凸性制御

BOA の縮小問題は, 非凸性制御と呼ぶ方法によって部分的に解決される. 非凸性制御においては, 図 5(a) のように大きい  $\lambda$  で現れる, 有限の  $V, \varepsilon$  の SE の解を利用する. 有限の  $V, \varepsilon$  が巨視的な解であるとき, AMP の微視的収束性 (2.41) は満たされないため微視的には収束していない. しかし,  $\varepsilon, V$  という統計量でみる範囲では収束していることから, AMP の軌道は統計量  $V, \varepsilon$  の空間に閉じ込められており, 小さい  $\lambda$  で見られたような AMP の軌道の発散は生じない. このように十分大きい  $\lambda$  で軌道が発散しない性質を用いて,  $\lambda$  を下げながら AMP を更新する方法を非凸性制御と呼ぶ.

AMP の更新と同時に  $\lambda$  を下げることで  $V = \varepsilon = 0$  に到達するためには, SE の巨視的固定点が  $\lambda$  に対して連続的になっている必要がある. 図 6(a) は,  $\alpha = 0.5$ ,  $a = 3$  で,  $\rho = 0.28$  での巨視的固定点の  $\lambda$  に関する系列と原点への引き込み領域を示したものである. 実線は  $\lambda$  を連続的に変化させた時の SE の巨視的固定点の系列であり, 線上の点は各  $\lambda$  における固定点を例として示している. 原点は微視的にも収束しているが, その他の点は微視的には収束していない. 図 6 (a) の影付きの領域は  $\lambda = 0.3$ ,  $a = 3$  での  $V = \varepsilon = 0$  への BOA である.  $\lambda = 0.3$ ,  $a = 3$  で真の信号を復元するには,  $\varepsilon \sim O(10^{-2})$  の初期条件を用意しなくてはならず, 非現実的である. 一方で  $\lambda = 1$  などの大きい値では SE の固定点への BOA が十分大きく, 初期条件を fine tuning せずに巨視的固定点に収束することが可能である.

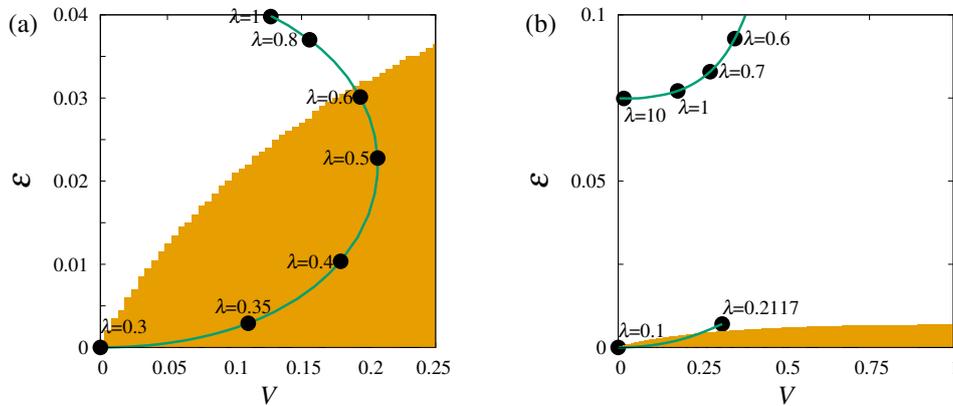


図6 SCADにおいて  $\alpha = 0.5$ ,  $a = 3$  で固定して  $\lambda$  を減少させたときの巨視的固定点の系列 (実線).  $\bullet$  は各  $\lambda$  における固定点を表す. 影付きの領域は十分小さい  $\lambda$  における原点への引き込み領域で, (a) は  $\lambda = 0.3$ ,  $\rho = 0.28$  の場合, (b) は  $\lambda = 0.1$ ,  $\rho = 0.32$  の場合.

この  $\lambda = 1$  から  $\lambda$  の値を徐々に下げることで, 各  $\lambda$  の SE の固定点を辿って  $\lambda = 0.3$  での  $V = \varepsilon = 0$  の BOA に入ることができる, というのが非凸性制御の方法である. しかし, この方法は真の信号がより密な場合には必ずしも有効ではない. 図 6 (b) に  $\alpha = 0.5$ ,  $\rho = 0.32$  での  $\lambda = 0.1$ ,  $a = 3$  における原点への引き込み領域と固定点の系列を示す. このとき, SE の巨視的固定点の系列が不連続になることがある. 図 6(b) の場合は  $\lambda \in (0.2117, 0.5530)$  で解が存在せず,  $\lambda$  を  $\lambda = 0.2117$  から 0.1 まで下げれば固定点の系列が  $V = \varepsilon = 0$  につながる. しかしこの  $\lambda = 0.2117$  では既に BOA が狭くなっており, 現実的な初期条件から巨視的固定点に到達することが難しい.  $\lambda$  に対する解の不連続性は, 1.3 節で述べた非凸性パラメータ  $a$  に対する制約から生じる. 信号が密な場合にはこの制約を満たす解が存在しない  $\lambda$  領域があるため, 非凸性制御による性能改善が難しくなってしまう.

以上の SE における議論を頼りに, 一つのデータの実現値に対する AMP における非凸性制御のプロトコルを考えてみる. ここでは ‘平衡’ 的アプローチを考える. すなわち, 各  $\lambda$  において AMP が巨視的に収束するまで十分待ち, その後  $\lambda$  を  $d\lambda$  だけ減らす, というものである. SE では典型軌道を  $V$  と  $\varepsilon$  により記述していたが,  $\varepsilon$  については, その 1 サンプルの値  $\varepsilon^{(t)}$  を評価することができない. そこで,  $\hat{V}^{(t)}$  と  $\hat{D}^{(t)} \equiv \frac{1}{N} \sum_{i=1}^N (\hat{x}_i^{(t+1)} - \hat{x}_i^{(t)})^2$  により巨視的な収束を評価する. 十分な AMP のステップ数ののち  $\hat{V}^{(t)}$ ,  $\hat{D}^{(t)}$  がある値の周囲で揺らぐようになれば, 巨視的に収束していると見なすことにする. その上で,  $\lambda$  を  $d\lambda$  だけ下げるが, 次にその  $d\lambda$  の値を考えよう. 非凸性制御がうまくいくためには,  $\lambda$  における SE の巨視的固定点が  $\lambda - d\lambda$  における巨視的固定点の BOA 内にある必要がある. そこで, その条件が成立する最大の  $d\lambda$  を  $d\lambda_{\max}$  として評価したのが図 7(a) である. ここで  $\alpha = 0.5$ ,  $\rho = 0.28$  とした. このパラメータ領域では  $\lambda < 0.3$  で完全再構成が可能なた

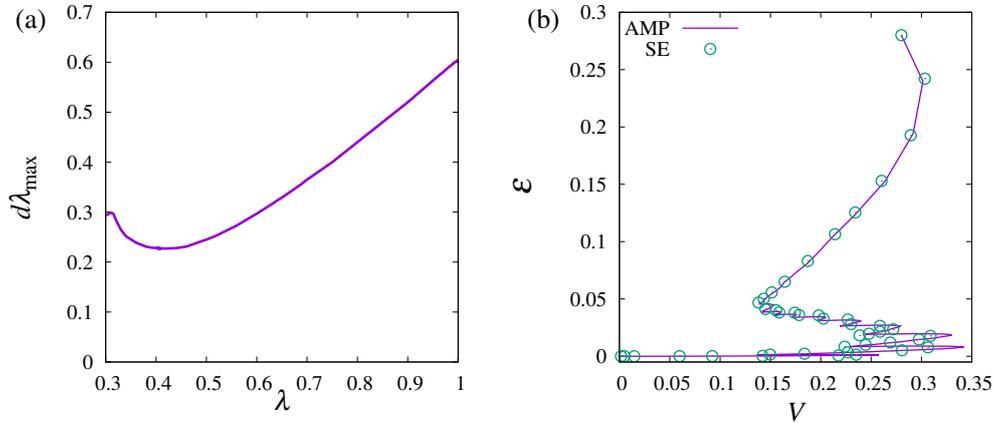


図7 (a)  $\alpha = 0.5, \rho = 0.28, a = 3$ での  $d\lambda_{\max}$  の  $\lambda$  依存性. (b)  $V - \varepsilon$  平面上での SCAD における非凸性制御を用いた軌道.  $N = 10^5$  における一つの  $\mathcal{D}$  に対する AMP の軌道を実線で, 密度発展法の軌道を  $\circ$  で示す.

め,  $\lambda < 0.3$  での  $d\lambda$  は自明に  $d\lambda \sim \lambda$  となっている.

図7(a)の  $d\lambda_{\max}$  の値は SE から評価しているが, 実際の AMP での  $d\lambda_{\max}$  の値はデータの実現値に応じて揺らぐことが想定される. そこで安全のため, ここでは全ての  $\lambda$  において  $d\lambda = 0.1$  として AMP における非凸性制御を行ってみる. 実際に AMP において非凸性制御を行った例が図7(b)である. また, 非凸性制御を行わない場合と行った場合の推定値を図8に示す. ここでは  $\alpha = 0.5, \rho = 0.28$  としており, このパラメータ領域で適当な初期条件から AMP を実行すると推定値が発散してしまう. そのため図8(a)では AMP による10回の更新後の推定値を示している. 図7(b)では, ひとつのデータに対する  $N = 10^5$  での AMP の軌道を実線で, 対応する SE の軌道を  $\circ$  で示している. AMP の初期条件は  $\mathbf{x} = \mathbf{0}$ ,  $\mathbf{v}$  の全成分は1としており, 対応する巨視的変数の初期条件は  $V = \varepsilon = \rho$  である. 非凸性パラメータの初期値は  $\lambda = 1, a = 3$  として,  $d\lambda = 0.1$  ずつ下げて最終的に  $\lambda = 0.3$  とした. 非凸性制御下の AMP の振る舞いは SE により精度良く記述されており, また図5(b)の SE flow と比較すると, AMP の軌道は  $V = \varepsilon = 0$  への引き込み領域に入っていくことがわかる.  $V = \varepsilon = 0$  への到達時の推定値は図8(b)に示す. このように非凸性制御がうまくいくパラメータ領域では, 適当な初期条件からの AMP の更新が失敗するような場合でも, 初期条件を fine tuning することなく解くことができる.

図9は, (a)SCAD 最小化法と (b)MCP 最小化法において, 非凸性制御による信号復元が可能なパラメータ領域を表している. ‘NCC limit’ の線は非凸性制御により信号復元できる限界を示している. この線より上の領域であれば,  $\lambda$  に関する巨視的固定点の系列が連続的に  $V = \varepsilon = 0$  に繋がるため, 非凸性制御によって信号を復元することが可能である. また ‘Principle limit’ の線は, これより上の領域であれば SE において  $V = \varepsilon = 0$  が大域的安

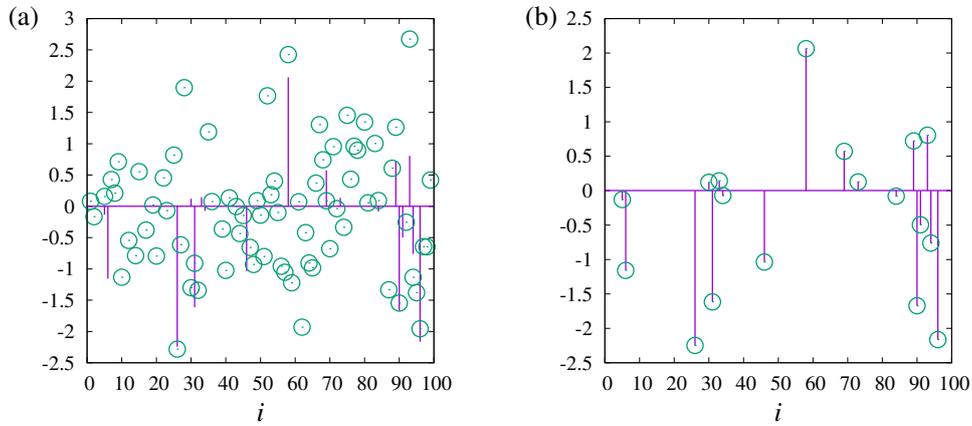


図8  $N = 1000, M = 500, \rho = 0.28, a = 3$  における, 推定値  $\hat{x}_i$  ( $i = 1, \dots, 100, \circ$ ) と真の信号  $x_i^{(0)}$  ( $i = 1, \dots, 100$ , 実線) の比較. (a) は  $\lambda = 0.3$  で固定して 10 ステップ更新した場合. (b) は非凸性制御により  $\lambda = 1$  から  $\lambda = 0.3$  まで下げながら AMP を更新した場合.

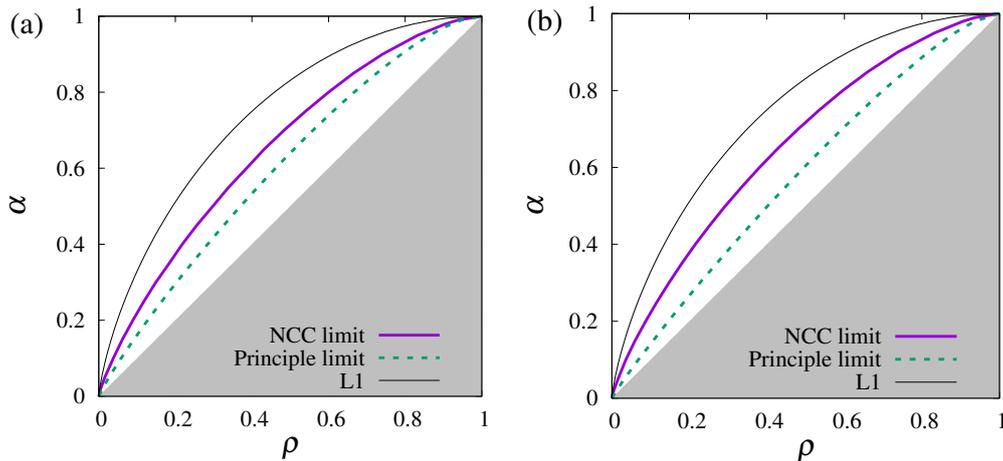


図9  $a = 3$  において (a) SCAD 最小化法 (b) MCP 最小化法を用いた場合の信号復元に関する相図. NCC limit は非凸性制御により信号復元ができる限界点で, Principle limit は SE において  $V = \varepsilon = 0$  が固定点ではなくなる限界点である. 比較のために  $\ell_1$  最小化法の結果を示している. 影つきの領域は, 信号復元が自明に不可能である領域である.

定解になることを示している. Principle limit と NCC limit の間にはギャップがあり, この領域では本稿で提案する非凸性制御法では信号を復元することはできない. しかし, 非凸性制御により信号復元が可能なパラメータ領域は,  $\ell_1$  最小化法より広がっており, より少ない観測でより密な信号を復元できることがわかる. また図9において (a)SCAD と (b)MCP を比較すると, Principle limit については SCAD より MCP の方が内側にあるが, 一方で NCC limit については SCAD の方が僅かに内側にあり, 非凸性制御による復元性能は SCAD の方が高い. このことから, 非凸制約のデザイン方法によっては, さらに広い

領域で非凸性制御を有効にできる可能性があると考えられる。

本項で紹介した非凸性制御は、多数の局所解がないパラメータ領域において BOA の縮小を解決するための手法である。したがって、局所解により推定が失敗する場合には、別の解決方法を考える必要がある。しかし SE の解析から、局所解の出現以外の理由で非凸制約最小化が困難になる場合があることがわかった。非凸性制御は部分的な解決策ではあるが、非凸制約最小化問題に対する AMP の軌道を安定化させる妥当な方法の一つであると考えている。

#### 4. まとめ

本稿では SCAD, MCP 最小化法を用いた圧縮センシングにおける信号復元について紹介した。非凸性パラメータが十分小さいとき、 $\ell_1$  最小化法よりも密な信号を少ない観測数で復元することが理論的に予想されるが、実際に AMP を用いて解くと、解が発散してしまう現象が見られた。SE flow の観察の結果、この現象の背後には引き込み領域の消失という SCAD, MCP 特有の現象があることがわかった。この問題点は非凸性制御と呼ぶ方法により一部解決される。十分大きい非凸性パラメータを初期値とし、AMP での推定変数の更新と同時に非凸性パラメータを下げることにより、初期条件の fine tuning をしなくても固定点に収束できることが示された。

SCAD, MCP は、非凸性パラメータについて適切な極限をとった場合に推定量がオラクル性を持つようにデザインされている (Fan and Li (2001), Zhang (2010))。一方で圧縮センシングにこれらの制約を用いる場合には、引き込み領域をいかに消失させないかという観点で制約をデザインする必要があるということが本研究から示された。このような引き込み領域の消失と、非凸性制御による AMP の安定化が、i.i.d 以外の場合についても有効であるかどうかを検討することも必要であろう。たとえば回転不変行列などはその候補である (Schniter et al. (2017), Gerbelot et al. (2020), Takahashi and Kabashima (2020))。

本稿ではデータがノイズを含まない場合を扱ったが、ノイズありの場合にも適用することができる。たとえば次のようなデータ生成過程を考える。

$$\mathbf{y} = A\mathbf{x}^0 + \Delta \quad (4.1)$$

ここで  $\Delta \in \mathbb{R}^M$  はノイズベクトルである。このデータのもとでの線形回帰問題を考えると、推定値は次のように与えられる。

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|_2^2 + J(\mathbf{x}; \eta) \right\} \quad (4.2)$$

係数  $1/2$  は数学的取り扱いやすさのために導入した。2章の表記に従うと、推定値は次の

ように与えられる.

$$\hat{x}_i(\mathcal{D}) = \lim_{\tau \rightarrow 0} \bar{x}_i(\mathcal{D}; \tau, \beta = \tau^{-1}) \quad (4.3)$$

したがって, 3章以降の議論で  $\tau' = 1$  とすれば, ノイズありデータに対する線形回帰問題に対応する. ノイズありの場合についても, 局所解の出現と対応する微視的安定性条件を得ることができ, この性質を利用した適切な非凸性パラメータの選択法なども提案できる. 詳しくは Sakata and Xu (2018), Obuchi and Sakata (2019) を参照されたい.

## 付録

### A SCAD, MCP 最小化に対する AMP の導出

ここではまず一般の尤度と事前分布に対する AMP である GAMP を導出し (Rangan (2011), Kabashima and Uda (2004), Xu et al. (2014)), SCAD, MCP 最小化問題に対する AMP を導出する.

#### A1 GAMP の導出

まず  $u_\mu = (\mathbf{A}\mathbf{x})_\mu$  という制約を表す次の恒等式を導入する.

$$1 = \int du_\mu \delta(u_\mu - (\mathbf{A}\mathbf{x})_\mu) = \int du_\mu \frac{1}{2\pi} \int d\hat{u}_\mu \exp \left\{ -i\hat{u}_\mu \left( u_\mu - \sum_{i=1}^N A_{\mu i} x_i \right) \right\} \quad (A.1)$$

これを (2.7) に代入すると次の表式を得る.

$$\begin{aligned} \tilde{m}_{\mu \rightarrow i}(x_i) &= \frac{1}{2\pi \tilde{Z}_{\mu \rightarrow i}} \int du_\mu f(y_\mu | u_\mu) \int d\hat{u}_\mu \exp \left\{ -i\hat{u}_\mu (u_\mu - A_{\mu i} x_i) \right\} \\ &\quad \times \prod_{j \neq i} \left\{ \int dx_j m_{j \rightarrow \mu}(x_j) \exp \left\{ i\hat{u}_\mu A_{\mu j} x_j \right\} \right\} \end{aligned} \quad (A.2)$$

ここで  $j \neq i$  に対する  $\exp(i\hat{u}_\mu A_{\mu j} x_j)$  について,  $A_{\mu j}$  に関して二次まで展開する. 二次での打ち切りは,  $A_{\mu j} \sim O(N^{-1/2})$  であることから正当化される. そして  $j \neq i$  について  $x_j$  積分を行い, 結果として得られる  $\hat{u}_\mu$  のガウス積分を実行すると, 次の表式を得る.

$$\tilde{m}_{\mu \rightarrow i}(x_i) = \frac{1}{\tilde{Z}_{\mu \rightarrow i} \sqrt{2\pi(V_\mu - A_{\mu i}^2 v_{i \rightarrow \mu})}} \int du_\mu f(y_\mu | u_\mu) \exp \left\{ -\frac{(u_\mu - \omega_\mu - A_{\mu i}(x_i - a_{i \rightarrow \mu}))^2}{2(V_\mu - A_{\mu i}^2 v_{i \rightarrow \mu})} \right\} \quad (A.3)$$

ここで  $\omega_\mu$  と  $V_\mu$  はそれぞれ次のように与えられる.

$$\omega_\mu = \sum_{i=1}^N A_{\mu i} a_{i \rightarrow \mu} \quad (A.4)$$

$$V_\mu = \sum_{i=1}^N A_{\mu i}^2 v_{i \rightarrow \mu} \quad (A.5)$$

$a_{i \rightarrow \mu}$ ,  $v_{i \rightarrow \mu}$  はメッセージ  $m_{i \rightarrow \mu}$  の平均と分散である。再度, (A.3) を  $A_{\mu i}$  について2次まで展開すると次の表式を得る。

$$\tilde{m}_{\mu \rightarrow i}(x_i) = \frac{1}{\tilde{Z}_{\mu \rightarrow i}} \exp \left\{ -\frac{A_{\mu \rightarrow i}}{2} x_i^2 + B_{\mu \rightarrow i} x_i \right\} \quad (\text{A.6})$$

ここで  $\tilde{Z}_{\mu \rightarrow i} = \sqrt{\frac{2\pi}{A_{\mu \rightarrow i}}} e^{\frac{B_{\mu \rightarrow i}^2}{2A_{\mu \rightarrow i}}}$  とした。また  $A_{\mu \rightarrow i}$  と  $B_{\mu \rightarrow i}$  は次のように与えられる。

$$A_{\mu \rightarrow i} = (g'_{\text{out}})_{\mu} A_{\mu i}^2 \quad (\text{A.7})$$

$$B_{\mu \rightarrow i} = (g_{\text{out}})_{\mu} A_{\mu i} + (g'_{\text{out}})_{\mu} A_{\mu i}^2 a_{i \rightarrow \mu} \quad (\text{A.8})$$

$g_{\text{out}}$ ,  $g'_{\text{out}}$  は次のように与えられる。

$$(g_{\text{out}})_{\mu} = \frac{\partial}{\partial \omega_{\mu}} \log \left( \int du_{\mu} f(y_{\mu} | u_{\mu}) \exp \left( -\frac{(u_{\mu} - \omega_{\mu})^2}{2(V_{\mu} - A_{\mu i}^2 v_{i \rightarrow \mu})} \right) \right) \quad (\text{A.9})$$

$$(g'_{\text{out}})_{\mu} = -\frac{\partial^2}{\partial \omega_{\mu}^2} \log \left( \int du_{\mu} f(y_{\mu} | u_{\mu}) \exp \left( -\frac{(u_{\mu} - \omega_{\mu})^2}{2(V_{\mu} - A_{\mu i}^2 v_{i \rightarrow \mu})} \right) \right) \quad (\text{A.10})$$

(A.6) を (2.8) に代入すると, (2.8) についても同様の表現を得る。

$$\begin{aligned} m_{i \rightarrow \mu}(x_i) &\propto \phi(x_i) \exp \left\{ -\frac{x_i^2}{2} \sum_{\gamma \neq \mu} \mathcal{A}_{\gamma \rightarrow i} + x_i \sum_{\gamma \neq \mu} \mathcal{B}_{\gamma \rightarrow i} \right\} \\ &= \frac{1}{\tilde{Z}_{i \rightarrow \mu}} \phi(x_i) \exp \left\{ -\frac{(x_i - R_{i \rightarrow \mu})^2}{2\Sigma_{i \rightarrow \mu}^2} \right\} \end{aligned} \quad (\text{A.11})$$

ここで  $\tilde{Z}_{i \rightarrow \mu}$  は規格化定数である。  $R_{i \rightarrow \mu}$ ,  $\Sigma_{i \rightarrow \mu}^2$  は次のように定義した。

$$R_{i \rightarrow \mu} = \frac{\sum_{\gamma \neq \mu} \mathcal{B}_{\gamma \rightarrow i}}{\sum_{\gamma \neq \mu} \mathcal{A}_{\gamma \rightarrow i}}, \quad (\text{A.12})$$

$$\Sigma_{i \rightarrow \mu}^2 = \left( \sum_{\gamma \neq \mu} \mathcal{A}_{\gamma \rightarrow i} \right)^{-1} \quad (\text{A.13})$$

$a_{i \rightarrow \mu}$ ,  $v_{i \rightarrow \mu}$  の定義から, 次の式が成立する。

$$a_{i \rightarrow \mu} = f_a(\Sigma_{i \rightarrow \mu}^2, R_{i \rightarrow \mu}), \quad (\text{A.14})$$

$$v_{i \rightarrow \mu} = f_c(\Sigma_{i \rightarrow \mu}^2, R_{i \rightarrow \mu}). \quad (\text{A.15})$$

周辺事後分布の平均と分散は, (A.14) と (A.15) の  $\mu$  に依存する部分を足し合わせると次のように得られる。

$$a_i = f_a(\Sigma_i^2, R_i), \quad (\text{A.16})$$

$$v_i = f_c(\Sigma_i^2, R_i), \quad (\text{A.17})$$

ここで次のように定義した.

$$\Sigma_i^2 = \left( \sum_{\mu} \mathcal{A}_{\mu \rightarrow i} \right)^{-1} \quad (\text{A.18})$$

$$R_i = \frac{\sum_{\mu} \mathcal{B}_{\mu \rightarrow i}}{\sum_{\mu} \mathcal{A}_{\mu \rightarrow i}}. \quad (\text{A.19})$$

以上の表現はどのようなデータ  $\mathcal{D}$  に対しても適用できる.

## A2 GAMP の TAP 形式

前節の形式では,  $i = 1, \dots, N$ ,  $\mu = 1, \dots, M$  に対する (A.7), (A.8), (A.14), (A.15) を更新することから, 1ステップ内で  $2M \times N$  個のメッセージを更新する必要がある. この手続きの計算量は, 1回の更新あたり  $O(M^2 \times N + M \times N^2)$  であり, 小さいサイズの問題にしか適用できない. しかし, 行列  $\mathbf{A}$  の成分がスパースでなく, 全ての要素が  $O(1/\sqrt{N})$  であるとする,  $2M \times N$  のメッセージを  $M + N$  個のメッセージで書き換えることができる. これにより, 1回の更新あたりの計算量は  $O(M \times N)$  となる. このメッセージを削減した更新式は, 統計物理学におけるスピングラス理論で知られる Thouless–Anderson–Palmer (TAP) 方程式 (Thouless et al. (1977)) と等価であり, また AMP と呼ばれている (Donoho et al. (2009b)).

ここでは TAP 形式の導出について説明する. まず, (A.3) に現れる  $A_{\mu i}^2 v_{i \rightarrow \mu} \sim O(N^{-1})$  が十分小さいとして無視すると, 次の式を得る.

$$(g_{\text{out}})_{\mu} = \frac{\partial}{\partial \omega_{\mu}} \log \left( \int du_{\mu} f(y_{\mu} | u_{\mu}) \exp \left( -\frac{(u_{\mu} - \omega_{\mu})^2}{2V_{\mu}} \right) \right) \quad (\text{A.20})$$

$$(g'_{\text{out}})_{\mu} = -\frac{\partial^2}{\partial \omega_{\mu}^2} \log \left( \int du_{\mu} f(y_{\mu} | u_{\mu}) \exp \left( -\frac{(u_{\mu} - \omega_{\mu})^2}{2V_{\mu}} \right) \right) \quad (\text{A.21})$$

対応する  $\Sigma_i^2$ ,  $R_i$  は次のように与えられる.

$$\Sigma_i^2 = \left( \sum_{\mu} (g'_{\text{out}})_{\mu} A_{\mu i}^2 \right)^{-1} \quad (\text{A.22})$$

$$R_i = a_i + \left( \sum_{\mu} (g_{\text{out}})_{\mu} A_{\mu i} \right) \Sigma_i^2 \quad (\text{A.23})$$

十分大きい  $N$  ではメッセージ  $a_{i \rightarrow \mu}$  と  $v_{i \rightarrow \mu}$  はほとんど  $\mu$  に依存せず,  $a_i$ ,  $v_i$  に近くなる. しかし, 補正項の扱いには注意しなくてはならない. これは Onsager 反跳場として知られる (Thouless et al. (1977)). (A.14) に  $R_i$  の周辺でのテイラー展開を適用すると,  $a_{i \rightarrow \mu}$  は次のようになる.

$$\begin{aligned} a_{i \rightarrow \mu} &= f_a \left( \frac{1}{\sum_{\gamma} \mathcal{A}_{\gamma \rightarrow i} - \mathcal{A}_{\mu \rightarrow i}}, \frac{\sum_{\gamma} \mathcal{B}_{\gamma \rightarrow i} - \mathcal{B}_{\mu \rightarrow i}}{\sum_{\gamma} \mathcal{A}_{\gamma \rightarrow i} - \mathcal{A}_{\mu \rightarrow i}} \right) \\ &\simeq a_i + \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R_i} (-\mathcal{B}_{\mu \rightarrow i} \Sigma_i^2) + O(N^{-1}), \end{aligned} \quad (\text{A.24})$$

ここで  $A_{\mu \rightarrow i} \propto A_{\mu i}^2 \sim O(N^{-1})$  であることから  $\sum_{\gamma} \mathcal{A}_{\gamma \rightarrow i} - A_{\mu \rightarrow i}$  を  $\sum_{\gamma} \mathcal{A}_{\gamma \rightarrow i} = \Sigma_i^{-2}$  と近似し, また  $B_{\mu \rightarrow i} \sim O(N^{-1/2})$  であることを利用した. (A.24) に  $A_{\mu i}$  をかけて, 得られた式を  $i$  について和をとると, (A.4) は次のようになる.

$$\omega_{\mu} = \sum_i A_{\mu i} a_i - (g_{\text{out}})_{\mu} V_{\mu}, \quad (\text{A.25})$$

ここで,  $v_i = f_c(\Sigma_i^2, R_i) = \Sigma_i^2 \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R_i}$  であることを利用した.

$V_{\mu}$  は次のように近似される.

$$V_{\mu} = \sum_i A_{\mu i}^2 v_i. \quad (\text{A.26})$$

ここで  $v_{i \rightarrow \mu}$  は  $A_{\mu i}^2$  倍されているため,  $v_{i \rightarrow \mu}$  と  $v_i$  の差は  $N \rightarrow \infty$  で無視できるとした. 以上の手続きで最も時間がかかる部分は, (A.23) にある行列とベクトルの積  $\sum_{\mu} (g_{\text{out}})_{\mu} A_{\mu i}$  と (A.25) にある  $\sum_i A_{\mu i} a_i$  部分である. したがって, 1回の更新あたりの計算量は  $O(NM)$  である.

### A3 ランダム行列に対するさらなる一般化

$N, M$  が十分大きく, 行列  $A$  の各成分が i.i.d で平均ゼロ, 分散  $1/N$  であるとき, TAP 方程式はさらに簡単化される. 行列  $A$  の各成分の i.i.d. 性より, 大数の法則に従って次のように書き換えることができる.

$$V_{\mu} = \sum_i A_{\mu i}^2 v_i \rightarrow \frac{1}{N} \sum_{i=1}^N v_i \quad (\text{A.27})$$

以下では  $V \equiv \frac{1}{N} \sum_{i=1}^N v_i$  と定義する. 同様の議論から,  $\Sigma_i^2$  は  $i$  によらず典型的に次のように与えられる定数  $\Sigma^2$  に一致することが示される.

$$\Sigma^2 = \left( \frac{1}{M} \sum_{\mu} g'_{\text{out}}(\omega_{\mu}, V) \right)^{-1} \quad (\text{A.28})$$

対応する  $R_i$  は次の通りである.

$$R_i = \left( \sum_{\mu} g_{\text{out}}(\omega_{\mu}, V) A_{\mu i} \right) \Sigma^2 + a_i \quad (\text{A.29})$$

AMP は (A.27), (A.28), (A.29) と次の変数から成る.

$$\omega_{\mu} = \sum_i A_{\mu i} a_i - g_{\text{out}}(\omega_{\mu}, V) V, \quad (\text{A.30})$$

$$a_i = f_a(\Sigma^2, R_i), \quad (\text{A.31})$$

$$v_i = f_c(\Sigma^2, R_i), \quad (\text{A.32})$$

したがって AMP の 1 ステップは  $2(M + N + 1)$  個の変数の更新で与えられる.

## B 状態発展方程式の導出

ここでは発見的な説明から密度発展方程式を導出するが、その厳密性については Bayati and Montanari (2011), Bayati and Montanari (2012) などに詳しい。(A.19), (A.7), (A.8), (2.22), (2.23), (A.4) を組み合わせると、分布を特徴付ける平均値  $R_i^{(t)}$  について次の式が得られる。

$$R_i^{(t)} = \frac{1}{\sum_{\mu} A_{\mu i}^2} \sum_{\mu} \left\{ A_{\mu i} y_{\mu} - A_{\mu i} \sum_{j \neq i} A_{\mu j} a_{j \rightarrow \mu}^{(t)} \right\} \quad (\text{B.1})$$

上つきの  $(t)$  は、 $t$  回の更新時の値を意味する。ここで、 $N \rightarrow \infty$  において、異なる  $j, \mu$  に対する  $a_{j \rightarrow \mu}^{(t)}$  が互いに独立であると仮定する。すると  $\{R_i^{(t)}\}$  を  $A_{\mu i}$  および  $y_{\mu}$  に依存する確率変数と見なすことができる。さらに (B.1) は乱数の和として与えられることから、中心極限定理に従いガウス乱数と見なすことができる。これらの考察のもと、 $R_i^{(t)}$  の平均は次のように与えられる。

$$\begin{aligned} \mathbb{E}_{\mathcal{D}}[R_i^{(t)}] &= \mathbb{E}_{\mathcal{D}} \left[ \frac{1}{\sum_{\mu} A_{\mu i}^2} \left\{ \sum_{\mu} A_{\mu i} y_{\mu} - \sum_{\mu} \sum_{j \neq i} A_{\mu i} A_{\mu j} a_{j \rightarrow \mu}^{(t)} \right\} \right] \\ &\simeq \sum_{\mu} \mathbb{E}_{\mathcal{D}} [A_{\mu i} y_{\mu}] - \mathbb{E}_{\mathcal{D}} \sum_{\mu} \sum_{j \neq i} [A_{\mu i} A_{\mu j} a_{j \rightarrow \mu}^{(t)}] = x_i^{(0)} \end{aligned} \quad (\text{B.2})$$

また  $R_i^{(t)}$  の分散は次のように与えられる。

$$\mathbb{E}_{\mathcal{D}}[(R_i^{(t)} - x_i^{(0)})^2] \simeq \mathbb{E}_{\mathcal{D}} \left[ \left\{ \sum_{\mu} A_{\mu i} y_{\mu} - \sum_{\mu} \sum_{j \neq i} A_{\mu i} A_{\mu j} a_{j \rightarrow \mu}^{(t)} - x_i^{(0)} \right\}^2 \right] \simeq \alpha^{-1} \varepsilon^{(t)} \quad (\text{B.3})$$

ここで (A.24) における  $O(1/\sqrt{N})$  の項は無視し、 $(a_{j \rightarrow \mu}^{(t)})^2$  を  $(a_j^{(t)})^2$  と置き換え、また  $\{A_{\mu j}^2\}$  を  $N^{-1}$  で置き換えた。 $R_i^{(t)}$  の統計性 (B.2), (B.3) から、 $t+1$  回の更新時における周辺分布  $m_i^{(t+1)}(x_i)$  は次のように与えられる。

$$m_i^{(t+1)}(x_i) \simeq \frac{1}{\hat{Z}_i^{(t+1)}} \phi(x_i) \exp \left\{ -\beta \frac{(x_i - (x_i^0 + z\sqrt{\alpha^{-1}\varepsilon^{(t)}}))^2}{2\alpha^{-1}(\tau' + V^{(t)})} \right\} \quad (\text{B.4})$$

ここで  $z$  は平均ゼロ、分散 1 のガウス乱数で、 $\hat{Z}_i^{(t+1)}$  は規格化定数とした。(B.4) から、 $N$  が十分大きいとき、 $t$  ステップにおける周辺事後分布は 2 つのパラメータ  $V^{(t)}$  と  $\varepsilon^{(t)}$  により表現されることがわかる。最後に (2.18), (2.19), (B.4) より、密度発展方程式は (2.39), (2.40) のように得られる。

### C 微視的安定性条件の導出

AMP を十分長いステップ数更新し,  $R_{i \rightarrow \mu}^{(t)}$  が固定点  $R_{i \rightarrow \mu}$  に到達したとする. このとき (A.12) と (A.24) から次の式を得る.

$$R_{i \rightarrow \mu} = \frac{1}{\sum_{\gamma \neq \mu} \mathcal{A}_{\gamma \rightarrow i}^t} \left\{ \sum_{\gamma \neq \mu} \frac{A_{\gamma i} y_{\mu} - A_{\gamma i} \sum_{j \neq i} A_{\gamma j} a_j}{\alpha^{-1}(V_{\mu} + \tau')} \right\} + \sum_{\gamma \neq \mu} \frac{A_{\gamma i} \sum_{j \neq i} A_{\gamma j} \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R}}{\alpha^{-1}(V_{\mu} + \tau')} R_{j \rightarrow \gamma} \quad (\text{C.1})$$

固定点の線形安定性を見るため, 固定点から  $\delta R_{i \rightarrow \mu}$  だけずらしたときに, そのずれが AMP の更新によりどのように発展するかを考える. (C.1) より,  $\delta R_{i \rightarrow \mu}$  は次の式を満たす.

$$\delta R_{i \rightarrow \mu}^{(t+1)} = \frac{\alpha}{V + \tau'} \sum_{\gamma \neq \mu} A_{\gamma i} \sum_{j \neq i} A_{\gamma j} \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R} \delta R_{j \rightarrow \gamma}^{(t)} \quad (\text{C.2})$$

ここで  $N$  は十分大きいとして  $V_{\mu}$  を平均値  $V$  で置き換えた. ずれの各成分が相関しないとする,  $\{A_{\mu i}\}$  はガウス乱数なので, (C.2) 右辺の和は中心極限定理からガウス分布に従う. このことから, (C.2) の一次と二次のモーメントを観察することで, 線形安定性を議論することができる.

ここで十分大きい  $N$  において, ずれ  $\delta R_{i \rightarrow \mu}$  がデータについて期待値をとった値に収束する, という大数の法則が成立すると考える. このとき, 微視的安定性は密度発展法における巨視的変数  $V$  と  $\varepsilon$  を使って表現される. まず  $\delta R_{i \rightarrow \mu}$  の一次モーメントは,  $A_{\mu i}$  の期待値がゼロなので十分大きい  $N$  で無視できる. 次に二次モーメントは次のように与えられる.

$$\begin{aligned} (\delta R_{i \rightarrow \mu}^{(t+1)})^2 &\simeq \frac{\alpha^2}{(V + \tau')^2} E_{\mathcal{D}} \left[ \left( \sum_{\gamma \neq \mu} A_{\gamma i} \sum_{j \neq i} A_{\gamma j} \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R} \delta R_{j \rightarrow \gamma}^{(t)} \right)^2 \right] \\ &= \frac{\alpha}{(V + \tau')^2} \frac{1}{M} \sum_{\gamma \neq \mu} \left( \frac{1}{N} \sum_{j \neq i} E_{\mathcal{D}} \left[ \left( \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R} \right)^2 \right] (\delta R_{j \rightarrow \gamma}^{(t)})^2 \right) \\ &\simeq \frac{\alpha}{(V + \tau')^2} \frac{1}{M} \sum_{\gamma \neq \mu} \left( \frac{1}{N} \sum_{j \neq i} E_{\mathcal{D}} \left[ \left( \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R} \right)^2 \right] \right) \left( \frac{1}{N} \sum_{j \neq i} (\delta R_{j \rightarrow \gamma}^{(t)})^2 \right) \end{aligned} \quad (\text{C.3})$$

ここで積のサンプル平均をサンプル平均の積で書き換えた. さらに, 巨視的変数  $E_{\mathcal{D}} \left[ \left( \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R} \right)^2 \right]$  は  $i$  によらず次のように表現される.

$$E_{\mathcal{D}} \left[ \left( \frac{\partial f_a(\Sigma_i^2, R_i)}{\partial R} \right)^2 \right] \simeq \int dx^0 \phi(x^0) \int Dz \left( \frac{\partial f_a(\alpha^{-1}(V + \tau'), x^0 + z\sqrt{\alpha^{-1}\varepsilon})}{\partial(z\sqrt{\alpha^{-1}\varepsilon})} \right)^2 \quad (\text{C.4})$$

(C.3) に (C.4) を代入すると以下の表現を得る.

$$(\delta R_{i \rightarrow \mu}^{(t+1)})^2 \simeq \frac{\alpha}{(V + \tau')^2} \int dx^0 \phi(x^0) \int Dz \left( \frac{\partial f_a(\alpha^{-1}(V + \tau'), x^0 + z\sqrt{\alpha^{-1}\varepsilon})}{\partial(z\sqrt{\alpha^{-1}\varepsilon})} \right)^2 (\delta R_{i \rightarrow \mu}^{(t+1)})^2 \quad (\text{C.5})$$

ここで, 大数の法則により, 十分大きい  $N$  で  $\frac{1}{N} \sum_{j \neq i} (\delta R_{j \rightarrow \gamma}^{(t)})^2$  が  $(\delta R_{i \rightarrow \mu}^{(t)})^2$  そのものと一致することを利用した. (C.5) において  $(\delta R_{i \rightarrow \mu}^{(t+1)})^2 = 0$  が安定となるためには係数が 1 未満でなくてはならず, これはまさに条件 (2.41) である.

### 参考文献

- Agarwal, N., Allen-Zhu, Z., Bullins, B., Hazan, E. and Ma, T. (2016). Finding Approximate Local Minima for Nonconvex Optimization in Linear Time, *arXiv preprint arXiv:1611.01146*, 2016.
- The Event Horizon Telescope Collaboration et al (2019). First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole, *The Astrophysical Journal Letters*, **875**, p. L1.
- Bayati, M. and Montanari, A. (2011). The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing, *IEEE Transaction on information theory*, **57** (2), 764–785.
- Bayati, M. and Montanari, A. (2012). The Lasso risk for Gaussian matrices, *IEEE Transaction on information theory*, **58** (4), 1997–2017.
- Boche, H., Calderbank, R., Kutyniok, G. and Vybíral, J. (2015). *Compressed Sensing and its Applications*, Springer.
- Candés, E. J. (2008). The restricted isometry property and its implications for compressed sensing, *Comptes Rendus Mathématique*, **346**, 589–592.
- Candés, E. J. and Tao, T. (2006). Decoding by linear programming, *IEEE transactions on information theory*, **52** (12), 5406–5425.
- Chartrand, R. (2007). Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Process. Lett.*, **14**, 707–710.
- Chartrand, R. and Staneva, S. (2008). Restricted isometry properties and nonconvex compressive sensing, *Inverse Problems*, **24** (3), p. 035020.

- Chartrand, R. and Yin, W. (2008). Iteratively reweighted algorithms for compressive sensing, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3869–3872, IEEE.
- Chen, S. S., Donoho, D. L. and Saunders, M. A. (2001). Atomic Decomposition by Basis Pursuit, *SIAM Review*, **43** (1), 129–159.
- De Almeida, J. and Thouless, D. J. (1978). Stability of the Sherrington-Kirkpatrick solution of a spin glass model, *Journal of Physics A: Mathematical and General*, **11** (5), p. 983.
- Donoho, D. L. (2006a). Compressed sensing, *IEEE Transactions on information theory*, **52** (4), 1289–1306.
- Donoho, D. L. (2006b). High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension, *Discrete & Computational Geometry*, **35** (4), 617–652.
- Donoho, D. L. and Tanner, J. (2009). Counting faces of randomly-projected polytopes when the projection radically lowers dimension, *Journal of American Mathematical Society*, **22**, 1–53.
- Donoho, D. L., Maleki, A. and Montanari, A. (2009a). Message-passing algorithms for compressed sensing, *Proceedings of the National Academy of Sciences*, **106** (45), 18914–18919.
- Donoho, D. L., Maleki, A. and Montanari, A. (2009b). Message-passing algorithms for compressed sensing, *Proceedings of the National Academy of Sciences*, **106** (45), 18914–18919.
- Elad, M. (2010). *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American statistical Association*, **96** (456), 1348–1360.
- Gerbelot, C., Abbara, A. and Krzakala, F. (2020). Asymptotic Errors for High-Dimensional Convex Penalized Linear Regression beyond Gaussian Matrices, *Proceedings of Machine Learning Research, 33rd Annual Conference on Learning Theory*, 1–32, MLR press.

- Gribonval, R. and Nielsen, M. (2003). Sparse representation in unions of bases, DC Approximation Approach for  $\ell_0$ -minimization in Compressed Sensing, *IEEE Transaction on Information Theory*, **49** (12), p. 3320.
- Kabashima, Y. (2003). A CDMA multiuser detection algorithm on the basis of belief propagation, *J Phys A: Math & Gen*, **36**, 11111–11121.
- Kabashima, Y. and Saad, D. (1998). Belief propagation vs. TAP for decoding corrupted messages, *EPL (Europhysics Letters)*, **44** (5), p. 668.
- Kabashima, Y. and Uda, S. (2004). A BP-based algorithm for performing Bayesian inference in large perceptron-type networks, *International Conference on Algorithmic Learning Theory*, 479–493, Springer.
- Kabashima, Y., Wadayama, T. and Tanaka, T. (2009). A typical reconstruction limit for compressed sensing based on  $\ell_1$ -norm minimization, *Journal of Statistical Mechanics: Theory and Experiment*, **2009** (09), p. L09003.
- Krzakala, F., Mézard, M., Sausset, F., Sun, Y. and Zdeborová, L. (2012). Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices, *Journal of Statistical Mechanics: Theory and Experiment*, **2012** (08), p. P08009.
- Lustig, M., Donoho, D. and Pauly, J. M. (2007). Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging, *Magnetic Resonance in Medicine*, **58**, 1182–1195.
- MacKay, D. J. (1999). Good error-correcting codes based on very sparse matrices, *IEEE transactions on Information Theory*, **45** (2), 399–431.
- Mezard, M. and Montanari, A. (2009). *Information, physics, and computation*, Oxford University Press.
- Mézard, M., Parisi, G. and Virasoro, M. (1987). *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, **9**, World Scientific Publishing Co Inc.
- Montanari, A. and Bayati, M. (2011). The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing, *IEEE Transaction on Information Theory*, **57**, p. 764–785.

- Natarajan, B. K. (1995). Sparse approximate solutions to linear systems, *SIAM journal on computing*, **24** (2), 227–234.
- Obuchi, T. and Sakata, A. (2019). Cross validation in sparse linear regression with piecewise continuous nonconvex penalties and its acceleration, *J Phys A: Math & Theor*, **52**, p. 414003.
- Rangan, S. (2011). Generalized approximate message passing for estimation with random linear mixing, *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 2168–2172, IEEE.
- Sakata, A. (2018). Estimator of prediction error based on approximate message passing for penalized linear regression, *Journal of Statistical Mechanics: Theory and Experiment*, **2018** (6), p. 063404.
- Sakata, A. and Obuchi, T. (2021). Perfect reconstruction of sparse signals with piecewise continuous nonconvex penalties and nonconvexity control, *Journal of Statistical Mechanics: Theory and Experiment*, **2021** (9), p.093401.
- Sakata, A. and Xu, Y. (2018). Approximate message passing for nonconvex sparse regularization with stability and asymptotic analysis, *Journal of Statistical Mechanics: Theory and Experiment*, **2018** (3), p. 033404.
- Schniter, P., Rangan, S. and Fletcher, A. K. (2017). Vector Approximate Message Passing, *Information Theory Proceedings (ISIT), 2017 IEEE International Symposium on*, 1588–1592, IEEE.
- Takahashi, T. and Kabashima, Y. (2020). Macroscopic Analysis of Vector Approximate Message Passing in a Model Mismatch Setting, *Information Theory Proceedings (ISIT), 2020 IEEE International Symposium on*, 1403–1408, IEEE.
- Thouless, D. J., Anderson, P. W. and Palmer, R. G. (1977). Solution of 'solvable model of a spin glass', *Philosophical Magazine*, **35** (3), 593–601.
- Tropp, J. A. (2007). Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit, *IEEE Transaction on Information Theory*, **53** (12), 4655–4666.
- Xu, Y., Kabashima, Y. and Zdeborová, L. (2014). Bayesian signal reconstruction for 1-bit compressed sensing, *Journal of Statistical Mechanics: Theory and Experiment*, **2014** (11), p. P11015.

Zhang, C.-H. (2010). Nearly Unbiased Variable Selection Under Minimax Concave Penalty, *The Annals of Statistics*, **38** (2), 894–942.