メッセージ伝播法入門
第5回講義資料
状態発展法：厳密なアプローチ1

九州大学
平成30年10月10日～10月12日

豊橋技術科学大学
電気・電子情報工学系
准教授　竹内啓悟

状態発展方程式

以下の状態発展方程式[3-4]の厳密な証明を与える。

$$v^t = \sigma^2 + \frac{1}{\delta}\Psi_t(v^{t-1}), \qquad v^0 = 1,$$

$$\Psi_{t+1}(v) = \mathbb{E}[\{x_1 - \eta_t(x_1 + \sqrt{v}\omega_1)\}^2], \qquad \omega_1 \sim \mathcal{N}(0,1).$$

Boltausenの方法[3-4, 5-1]

1. $\boldsymbol{A}$を$\mathcal{N}(0,1/M)$に従う独立な要素を持つ行列とする。

2. 反復t時点までのすべての推定値の履歴を条件として与えたときに、観測行列$\boldsymbol{A}$の条件付き分布を評価する。

3. 上記の条件付き分布を使って、反復$t+1$での推定値の分布を厳密に計算する。

[5-1] E. Bolthausen, “An iterative construction of solutions of the TAP equations for the Sherrington-Kirkpatrick model,” *Commun. Math. Phys*., vol. 325, no. 1, pp. 333–366, Jan. 2014.

誤差の発展方程式

AMP[3-1]

$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\widehat{\boldsymbol{x}}^t + \frac{\langle \eta'_{t-1}(\widehat{\boldsymbol{x}}^{t-1} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^{t-1}) \rangle}{\delta}\boldsymbol{z}^{t-1}, \qquad \boldsymbol{z}^{-1} = \boldsymbol{0},$$

$$\widehat{\boldsymbol{x}}^{t+1} = \eta_t(\widehat{\boldsymbol{x}}^t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^t).$$

誤差の発展方程式

$\boldsymbol{q}_t = \widehat{\boldsymbol{x}}^t - \boldsymbol{x}$ 、$\boldsymbol{h}_t = \widehat{\boldsymbol{x}}^t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^t - \boldsymbol{x}$ と誤差を定義する。
$\boldsymbol{m}_t = \boldsymbol{z}^t$、$\boldsymbol{b}_t = \boldsymbol{w} - \boldsymbol{z}^t$とすると、

$$\boldsymbol{q}_{t+1} = \eta_t(\boldsymbol{x} + \boldsymbol{h}_t) - \boldsymbol{x}, \qquad \boldsymbol{h}_t = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{m}_t + \boldsymbol{q}_t,$$

$$\boldsymbol{m}_t = \boldsymbol{w} - \boldsymbol{b}_t, \qquad \boldsymbol{b}_t = \boldsymbol{A}\boldsymbol{q}_t - \lambda_{t-1}\boldsymbol{m}_{t-1}, \qquad \lambda_t = \frac{\langle \eta'_t(\boldsymbol{x} + \boldsymbol{h}_t) \rangle}{\delta}.$$

誤差の発展方程式

$$\boldsymbol{B}_{\boldsymbol{t}} = (\boldsymbol{b}_{\boldsymbol{0}}, \dots, \boldsymbol{b}_{\boldsymbol{t-1}}), \qquad \boldsymbol{M}_{\boldsymbol{t}} = (\boldsymbol{m}_{\boldsymbol{0}}, \dots, \boldsymbol{m}_{\boldsymbol{t-1}}),$$

$$\boldsymbol{H}_{\boldsymbol{t}} = (\boldsymbol{h}_{\boldsymbol{0}}, \dots, \boldsymbol{h}_{\boldsymbol{t-1}}), \qquad \boldsymbol{Q}_{\boldsymbol{t}} = (\boldsymbol{q}_{\boldsymbol{0}}, \dots, \boldsymbol{q}_{\boldsymbol{t-1}}), \quad (\boldsymbol{0}, \boldsymbol{M}_0 \boldsymbol{\Lambda}_0) = \boldsymbol{0},$$

$$\boldsymbol{\Lambda}_t = \mathrm{diag}\{\lambda_0, \dots, \lambda_{t-1}\}, \qquad \boldsymbol{W}_t = (\boldsymbol{w}, \dots, \boldsymbol{w}) \in \mathbb{R}^{M \times t}.$$

上記の定義を使うと、

$$\boldsymbol{B}_t = \boldsymbol{A}\boldsymbol{Q}_t - (\boldsymbol{0}, \boldsymbol{M}_{t-1}\boldsymbol{\Lambda}_{t-1}), \ \boldsymbol{M}_t = \boldsymbol{W}_t - \boldsymbol{B}_t,$$

$$\boldsymbol{H}_t = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{M}_t + \boldsymbol{Q}_t, \qquad \boldsymbol{q}_{t+1} = \eta_t(\boldsymbol{x} + \boldsymbol{h}_t) - \boldsymbol{x}.$$

特異値分解と射影行列

フルランク縦長行列$\boldsymbol{M}$の特異値分解

$$\boldsymbol{M} = \boldsymbol{\Phi}_{\boldsymbol{M}} \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{M}} \\ \boldsymbol{O} \end{pmatrix} \boldsymbol{\Psi}_{\boldsymbol{M}}^{\mathrm{T}}, \qquad \boldsymbol{\Phi}_{\boldsymbol{M}} = \left(\boldsymbol{\Phi}_{\boldsymbol{M}}^{\parallel}, \boldsymbol{\Phi}_{\boldsymbol{M}}^{\perp}\right).$$

一般化逆行列

$$\boldsymbol{M}^{\dagger} = \left(\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}\right)^{-1}\boldsymbol{M}^{\mathrm{T}} = \boldsymbol{\Psi}_{\boldsymbol{M}}\boldsymbol{\Sigma}_{\boldsymbol{M}}^{-1}\left(\boldsymbol{\Phi}_{\boldsymbol{M}}^{\parallel}\right)^{\mathrm{T}}.$$

$\boldsymbol{M}$が張る空間への射影

$$\boldsymbol{P}_{\boldsymbol{M}}^{\parallel} = \boldsymbol{M}\boldsymbol{M}^{\dagger} = \boldsymbol{\Phi}_{\boldsymbol{M}}^{\parallel}\left(\boldsymbol{\Phi}_{\boldsymbol{M}}^{\parallel}\right)^{\mathrm{T}}.$$

直交補空間への射影

$$\boldsymbol{P}_{\boldsymbol{M}}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{M}}^{\parallel} = \boldsymbol{\Phi}_{\boldsymbol{M}}^{\perp}\left(\boldsymbol{\Phi}_{\boldsymbol{M}}^{\perp}\right)^{\mathrm{T}}.$$

補題5.1[3-4]

$\boldsymbol{A}$を$\mathcal{N}(0,1/M)$に従う独立な要素を持つ行列とする。

$\boldsymbol{X} \in \mathbb{R}^{M\times t}$、$\boldsymbol{Y} \in \mathbb{R}^{N\times t'}$、$\boldsymbol{U} \in \mathbb{R}^{N\times t}$、$\boldsymbol{V} \in \mathbb{R}^{M\times t'}$を
縦長行列とし、以下の制約を満たすものとする。

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{U}, \qquad \boldsymbol{Y} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}.$$

$\boldsymbol{U}$と$\boldsymbol{V}$がフルランクならば、上記の制約を満たすという
条件の下で、$\boldsymbol{A}$は以下と統計的に等価である。

$$\begin{aligned}\boldsymbol{A} &\sim \boldsymbol{X}\boldsymbol{U}^{\dagger} + \left(\boldsymbol{V}^{\dagger}\right)^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{U}}^{\perp} + \boldsymbol{\Phi}_{\boldsymbol{V}}^{\perp}\widetilde{\boldsymbol{A}}\left(\boldsymbol{\Phi}_{\boldsymbol{U}}^{\perp}\right)^{\mathrm{T}} \\ &\sim \boldsymbol{P}_{\boldsymbol{V}}^{\perp}\boldsymbol{X}\boldsymbol{U}^{\dagger} + \left(\boldsymbol{V}^{\dagger}\right)^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}} + \boldsymbol{\Phi}_{\boldsymbol{V}}^{\perp}\widetilde{\boldsymbol{A}}\left(\boldsymbol{\Phi}_{\boldsymbol{U}}^{\perp}\right)^{\mathrm{T}}.\end{aligned}$$

ただし、$\widetilde{\boldsymbol{A}}$は$\boldsymbol{A}$と独立で、$\mathcal{N}(0,1/M)$に従う独立な要素を持つ行列である。

補題5.1の証明[5-2]

$$\boldsymbol{U} = \boldsymbol{\Phi}_U \begin{pmatrix} \boldsymbol{\Sigma}_U \\ \boldsymbol{O} \end{pmatrix} \boldsymbol{\Psi}_U^{\mathrm{T}}, \qquad \boldsymbol{V} = \boldsymbol{\Phi}_V \begin{pmatrix} \boldsymbol{\Sigma}_V \\ \boldsymbol{O} \end{pmatrix} \boldsymbol{\Psi}_V^{\mathrm{T}}.$$

上記の特異値分解を二つの制約式に代入すると、

$$\boldsymbol{\Phi}_V^{\mathrm{T}} \boldsymbol{X} \boldsymbol{\Psi}_U = \widehat{\boldsymbol{A}} \begin{pmatrix} \boldsymbol{\Sigma}_U \\ \boldsymbol{O} \end{pmatrix}, \qquad \boldsymbol{\Phi}_U^{\mathrm{T}} \boldsymbol{Y} \boldsymbol{\Psi}_V = \widehat{\boldsymbol{A}}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{\Sigma}_V \\ \boldsymbol{O} \end{pmatrix}, \qquad \widehat{\boldsymbol{A}} = \boldsymbol{\Phi}_V^{\mathrm{T}} \boldsymbol{A} \boldsymbol{\Phi}_U.$$

$\widehat{\boldsymbol{A}} = \begin{pmatrix} \widehat{\boldsymbol{A}}_{11} & \widehat{\boldsymbol{A}}_{12} \\ \widehat{\boldsymbol{A}}_{21} & \widehat{\boldsymbol{A}}_{22} \end{pmatrix}$と分解すると、

$$\begin{pmatrix} \widehat{\boldsymbol{A}}_{11} \\ \widehat{\boldsymbol{A}}_{21} \end{pmatrix} = \boldsymbol{\Phi}_V^{\mathrm{T}} \boldsymbol{X} \boldsymbol{\Psi}_U \boldsymbol{\Sigma}_U^{-1}, \qquad \left(\widehat{\boldsymbol{A}}_{11}, \widehat{\boldsymbol{A}}_{12} \right) = \boldsymbol{\Sigma}_V^{-1} \boldsymbol{\Psi}_V^{\mathrm{T}} \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\Phi}_U.$$

[5-2] K. Takeuchi, “Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements,” in *Proc. 2017 IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 501–505.

補題5.1の証明

$\boldsymbol{A} \sim \widehat{\boldsymbol{A}}$なので、$\widehat{\boldsymbol{A}}$の未知の要素$\widehat{\boldsymbol{A}}_{22}$は$\mathcal{N}(0,1/M)$に従う独立な要素を持つ。

$$\boldsymbol{A} = \boldsymbol{\Phi}_V \widehat{\boldsymbol{A}} \boldsymbol{\Phi}_U^{\mathrm{T}} = \boldsymbol{\Phi}_V \begin{pmatrix} \boldsymbol{\Phi}_V^{\mathrm{T}} \boldsymbol{X} \boldsymbol{\Psi}_U \boldsymbol{\Sigma}_U^{-1} & \boldsymbol{\Sigma}_V^{-1} \boldsymbol{\Psi}_V^{\mathrm{T}} \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\Phi}_U^{\perp} \\ & \widehat{\boldsymbol{A}}_{22} \end{pmatrix} \boldsymbol{\Phi}_U^{\mathrm{T}}$$

$$= \boldsymbol{\Phi}_V \begin{pmatrix} \multicolumn{2}{c}{\boldsymbol{\Sigma}_V^{-1} \boldsymbol{\Psi}_V^{\mathrm{T}} \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\Phi}_U} \\ \left(\boldsymbol{\Phi}_V^{\perp}\right)^{\mathrm{T}} \boldsymbol{X} \boldsymbol{\Psi}_U \boldsymbol{\Sigma}_U^{-1} & \widehat{\boldsymbol{A}}_{22} \end{pmatrix} \boldsymbol{\Phi}_U^{\mathrm{T}}$$

前者の表現を使って、

$$\boldsymbol{A} = \boldsymbol{X} \boldsymbol{\Psi}_U \boldsymbol{\Sigma}_U^{-1} \left(\boldsymbol{\Phi}_U^{\parallel}\right)^{\mathrm{T}} + \boldsymbol{\Phi}_V^{\parallel} \boldsymbol{\Sigma}_V^{-1} \boldsymbol{\Psi}_V^{\mathrm{T}} \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{P}_U^{\perp} + \boldsymbol{\Phi}_V^{\perp} \widehat{\boldsymbol{A}}_{22} \left(\boldsymbol{\Phi}_U^{\perp}\right)^{\mathrm{T}}$$

$$= \boldsymbol{X} \boldsymbol{U}^{\dagger} + \left(\boldsymbol{V}^{\dagger}\right)^{\mathrm{T}} \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{P}_U^{\perp} + \boldsymbol{\Phi}_V^{\perp} \widehat{\boldsymbol{A}}_{22} \left(\boldsymbol{\Phi}_U^{\perp}\right)^{\mathrm{T}}.$$

もう一つの表現は後者から得られる。 ■

補題5.1の注意

次ページの主定理の証明で必要な性質は、ベクトル$\widetilde{\boldsymbol{A}}\boldsymbol{v}$が平均$\boldsymbol{0}$の$K$体独立なガウス分布に従うという性質である。

例えば、$\boldsymbol{A}$をハール直交行列[5-3]に取ると、この性質を満たす。

- AMPはハール直交する観測行列を使用すると、収束性が保証されて、i.i.d.ガウス観測行列と同じ性能を達成する？
- ベイズ最適な性能[5-4]は$\boldsymbol{A}$の経験特異値分布に依存するため、この場合のAMPは常に最適でない。

[5-3] S. Chatterjee and E. Meckes, “Multivariate normal approximation using exchangeable pairs,” *ALEA Latin. Amer. J. Prob. Math. Stat.*, vol. 4, pp. 257–283, 2008.

[5-4] K. Takeda, S. Uda, and Y. Kabashima, “Analysis of CDMA systems that are characterized by eigenvalue spectrum,” *Europhys. Lett.*, vol. 76, no. 6, pp. 1193–1199, 2006.

主定理[3-4]

大システム極限で、以下が成り立つ。

(a) $\boldsymbol{b}_t \sim \boldsymbol{B}_t \boldsymbol{\beta}_t + \boldsymbol{\xi}_t, \qquad \boldsymbol{\xi}_t \sim \mathcal{N}(\boldsymbol{0}, M^{-1}\|\boldsymbol{q}_t^{\perp}\|^2 \boldsymbol{I}_M).$

(b) $\boldsymbol{h}_t \sim \boldsymbol{H}_t \boldsymbol{\alpha}_t + \boldsymbol{\zeta}_t, \qquad \boldsymbol{\zeta}_t \sim \mathcal{N}(\boldsymbol{0}, M^{-1}\|\boldsymbol{m}_t^{\perp}\|^2 \boldsymbol{I}_N).$

(c) $\frac{1}{N}\boldsymbol{h}_t^{\mathrm{T}}\boldsymbol{h}_{t'} - \frac{1}{M}\boldsymbol{m}_t^{\mathrm{T}}\boldsymbol{m}_{t'} \to 0, \quad \frac{1}{M}\boldsymbol{b}_t^{\mathrm{T}}\boldsymbol{b}_{t'} - \frac{1}{\delta N}\boldsymbol{q}_t^{\mathrm{T}}\boldsymbol{q}_{t'} \to 0.$

(d) $\frac{1}{N}\boldsymbol{h}_t^{\mathrm{T}}\boldsymbol{q}_{t'} - \frac{\delta\lambda_{t'-1}}{N}\boldsymbol{h}_t^{\mathrm{T}}\boldsymbol{h}_{t'-1} \to 0.$

(e) $\frac{1}{N}\boldsymbol{b}_t^{\mathrm{T}}\boldsymbol{w} \to 0, \quad \frac{1}{N}\boldsymbol{h}_t^{\mathrm{T}}\boldsymbol{x} \to 0.$

$$\boldsymbol{\alpha}_t = \boldsymbol{M}_t^{\dagger}\boldsymbol{m}_t, \qquad \boldsymbol{\beta}_t = \boldsymbol{Q}_t^{\dagger}\boldsymbol{q}_t = (\beta_0, \dots, \beta_{t-1})^{\mathrm{T}},$$

$$\boldsymbol{m}_t^{\perp} = \boldsymbol{P}_{\boldsymbol{M}_t}^{\perp}\boldsymbol{m}_t, \qquad \boldsymbol{q}_t^{\perp} = \boldsymbol{P}_{\boldsymbol{Q}_t}^{\perp}\boldsymbol{q}_t.$$

状態発展方程式の導出

性質(b)から、誤差$\boldsymbol{h}_t$はガウス分布に従うi.i.d.成分からなることが言える。

以下を仮定して、$\mathrm{N}^{-1}\boldsymbol{q}_{t+1}^{\mathrm{T}}\boldsymbol{q}_{t+1}$を評価する。

$$\frac{1}{N}\boldsymbol{q}_t^{\mathrm{T}}\boldsymbol{q}_t \to \Psi_t(v^{t-1}), \qquad v^t = \frac{1}{N}\boldsymbol{h}_t^{\mathrm{T}}\boldsymbol{h}_t.$$

性質(c)(e)、定義$\boldsymbol{m}_t = \boldsymbol{w} - \boldsymbol{b}_t$、大数の強法則を使って、

$$v^t \to \frac{1}{M}\boldsymbol{m}_t^{\mathrm{T}}\boldsymbol{m}_t \to \sigma^2 + \frac{1}{M}\boldsymbol{b}_t^{\mathrm{T}}\boldsymbol{b}_t \to \sigma^2 + \frac{1}{\delta}\Psi_t(v^{t-1}).$$

$$\frac{1}{N}\boldsymbol{q}_{t+1}^{\mathrm{T}}\boldsymbol{q}_{t+1} = \frac{1}{N}\sum_{n=1}^{N}\{\eta_t(x_n + [\boldsymbol{h}_t]_n) - x_n\}^2 \to \Psi_{t+1}(v^t).$$