

メッセージ伝播法入門

第3回講義資料

近似的メッセージ伝播法

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

ベイズ最適AMP

更新式

$$\mathbf{z}^t = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^t + \frac{\langle \eta'(\hat{\mathbf{x}}^{t-1} + \mathbf{A}^T \mathbf{z}^{t-1}; v^{t-1}) \rangle}{\delta} \mathbf{z}^{t-1},$$

$$v^t = \sigma^2 + \frac{1}{\delta} \text{MMSE}(1/v^{t-1}),$$

$$\hat{\mathbf{x}}^{t+1} = \eta(\hat{\mathbf{x}}^t + \mathbf{A}^T \mathbf{z}^t; v^t).$$

AWGN観測モデル $\bar{x}_n = x_n + s^{-1/2} \omega_n, \quad \omega_n \sim \mathcal{N}(0, 1).$

事後平均 $\eta(\bar{x}_n; s) = \mathbb{E}[x_n | \bar{x}_n, s].$

最小平均二乗誤差 (MMSE) $\text{MMSE}(s) = \mathbb{E}[\{x_n - \eta(\bar{x}_n; s)\}^2].$

信号の事前分布が既知である必要がある。

AMP[3-1]

更新式

$$\mathbf{z}^t = \mathbf{y} - A\hat{\mathbf{x}}^t + \frac{\langle \eta'_{t-1}(\hat{\mathbf{x}}^{t-1} + A^T \mathbf{z}^{t-1}) \rangle}{\delta} \mathbf{z}^{t-1},$$

$$\hat{\mathbf{x}}^{t+1} = \eta_t(\hat{\mathbf{x}}^t + A^T \mathbf{z}^t).$$

初期値

$$\mathbf{x}^0 = \mathbf{0}, \quad \mathbf{z}^{-1} = \mathbf{0}.$$

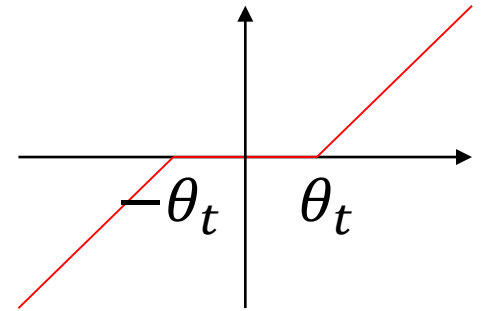
閾値関数を関数列 $\{\eta_t\}$ に一般化

[3-1] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.

閾値関数の例

Soft thresholding (ST)

$$\eta_t(x) = \eta_{\text{ST}}(x; \theta_t) = \begin{cases} x - \theta_t & \text{for } x > \theta_t \\ 0 & \text{for } |x| \leq \theta_t \\ x + \theta_t & \text{for } x < -\theta_t. \end{cases}$$



ミニマックス最適性[3-2]

任意の $p, q > 0$ に対して、

$$\lim_{a \downarrow 0} \frac{r_{\text{ST}}(a)}{r_{\text{opt}}(a)} = 1.$$

$$r_{\text{ST}}(a) = \inf_{\theta} \sup_{p(x_n): (\mathbb{E}[|x_n|^p])^{1/p} \leq a} \mathbb{E} \left[|x_n - \eta_{\text{ST}}(x_n + \sqrt{v}\omega_n; \theta)|^q \right],$$

$$r_{\text{opt}}(a) = \inf_{\eta} \sup_{p(x_n): (\mathbb{E}[|x_n|^p])^{1/p} \leq a} \mathbb{E} \left[|x_n - \eta(x_n + \sqrt{v}\omega_n)|^q \right].$$

[3-2] D. L. Donoho and I. M. Johnstone, “Minimax risk over l_p -balls for l_q -error,”
Probab. Theory Relat. Fields, vol. 99, no. 2, pp. 277-303, Jun. 1994.

ミニマックス最適性の意義

モーメント制約

$$p(x_n) = (1 - \rho)\delta(x_n) + \rho\pi(x_n), \quad \int |x|^p \pi(x) dx = M_p < \infty.$$

$\rho < a^p / M_p$ を満たす分布は、モーメント制約を満たす。

上記の信号分布のクラスでは信号密度 ρ が0に収束する。

ミニマックス最適性

- モーメント制約を満たす分布の中で**最悪な信号分布**を想定
- パラメータ θ を**最適値**に設定
- ST関数は L_q 誤差を**最小化**

最適解の唯一性は主張していない。

パラメータ θ_t の最適化[2-3,3-3,3-1]

1. 状態発展法による性能解析
2. 閾値 ρ_{SE} の定義
3. 閾値基準によるパラメータ θ_t の最適化

[3-3] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

状態発展法[3-4]

状態発展方程式

$$\bar{v}^t = \sigma^2 + \frac{1}{\delta} \Psi(\bar{v}^{t-1}, \lambda_{t-1}), \quad \bar{v}^0 = 1.$$

$$\Psi(v, \lambda) = \mathbb{E}[\{x_1 - \eta_{\text{ST}}(x_1 + \sqrt{v}\omega_1; \lambda\sqrt{v})\}^2].$$

閾値を $\theta_t = \lambda_t \sqrt{\bar{v}^t}$ と表現しなおした。

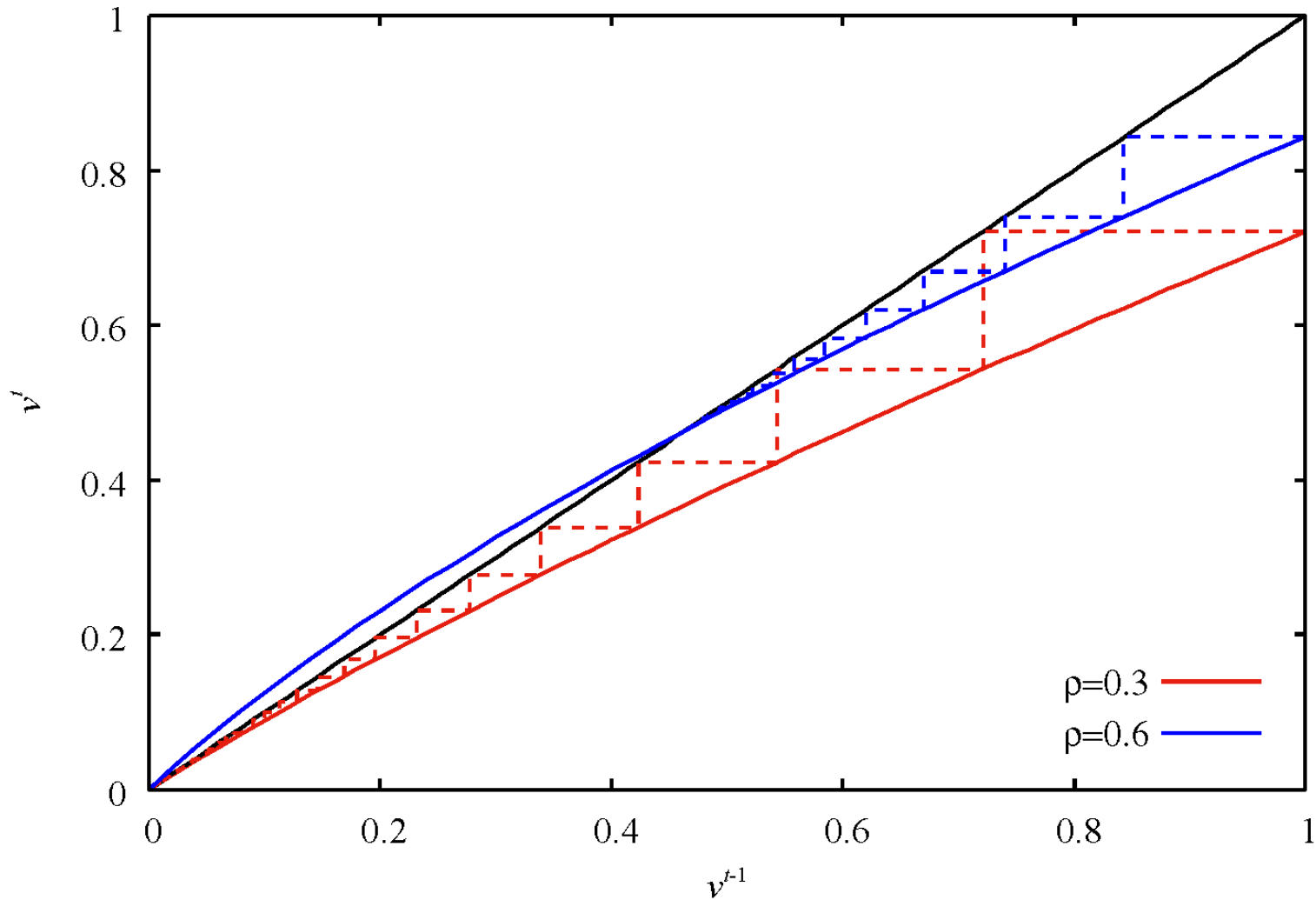
誤差の漸近評価

$$\lim_{M=\delta N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}^t\|^2] = \Psi(\bar{v}^{t-1}, \lambda^{t-1}).$$

状態発展法の解説は、第4～6回講義資料を参照

[3-4] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

数値的評価



信号密度 ρ のBG事前分布、 $\sigma^2 = 0$ 、 $\delta = 0.65$ 、最適な λ_t

閾値[2-4, 3-1]

定義

$$\rho_{\text{SE}}(\delta) = \sup_{\{\lambda_t > 0\}} \inf_{p(x_1): \Pr(x_1=0)=1-\rho} \rho_{\text{SE}}(\{\lambda_t\}, p(x_1), \delta),$$

ただし、

$$\rho_{\text{SE}}(\{\lambda_t\}, p(x_1), \delta) = \sup \left\{ \rho > 0 : \lim_{t \rightarrow \infty} \lim_{\sigma^2 \rightarrow 0} \bar{v}^t = 0 \right\}.$$

閾値の意義

閾値 ρ_{SE} は、ノイズの分散 σ^2 が0に収束する場合に、**最悪の信号分布**を想定したときに大システム極限における平均二乗誤差が0に収束するような最大の信号密度 ρ に等しい。

Soft Thresholdingの閾値

閾値

$$\rho_{SE} = \sup_{\lambda > 0} \frac{\delta - 2[(1 + \lambda^2)P_G(-\lambda) - \lambda p_G(\lambda)]}{(1 + \lambda^2) - 2[(1 + \lambda^2)P_G(-\lambda) - \lambda p_G(\lambda)]}.$$

ただし、 p_G と P_G はそれぞれ標準ガウス分布の確率密度関数と累積分布関数を表す。

最適なパラメータ

$$\lambda_t = \lambda_{opt} \equiv \operatorname{argsup}_{\lambda > 0} \frac{\delta - 2[(1 + \lambda^2)P_G(-\lambda) - \lambda p_G(\lambda)]}{(1 + \lambda^2) - 2[(1 + \lambda^2)P_G(-\lambda) - \lambda p_G(\lambda)]}.$$

閾値が最大になるように λ_t を設定

閾値の評価の概要[3-1]

性質1

不動点方程式 $\bar{v} = \delta^{-1}\Psi(\bar{v}, \lambda_t)$ は解 $\bar{v} = 0$ を持つ。

性質2

関数 $\Psi(\bar{v}, \lambda_t)$ は \bar{v} に関して上に凸である。

性質1と性質2から、

$$\begin{aligned}\rho_{\text{SE}}(\{\lambda_t\}, p(x_1), \delta) &= \sup \left\{ \rho > 0 : \lim_{t \rightarrow \infty} \lim_{\sigma^2 \rightarrow 0} \bar{v}^t = 0 \right\} \\ &= \sup \left\{ \rho > 0 : \frac{1}{\delta} \lim_{v \downarrow 0} \frac{\partial}{\partial v} \Psi(v; \lambda_t) < 1 \right\}.\end{aligned}$$

上記の $\rho_{\text{SE}}(\{\lambda_t\}, p(x_1), \delta)$ は事前分布の詳細に依存しない。

準備1: 超関数の意味での微分

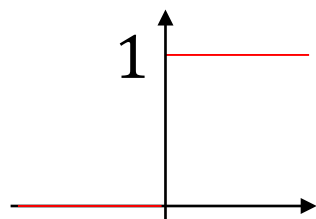
任意の急減少関数 ϕ に対して以下を満たす関数 Df が存在するとき、 Df を f の超関数の意味での微分と呼び、 f' と書く。

$$\int Df(x)\phi(x)dx = - \int f(x)\phi'(x)dx.$$

意義

$$\int f'(x)\phi(x)dx = - \int f(x)\phi'(x)dx \quad \text{部分積分}$$

例



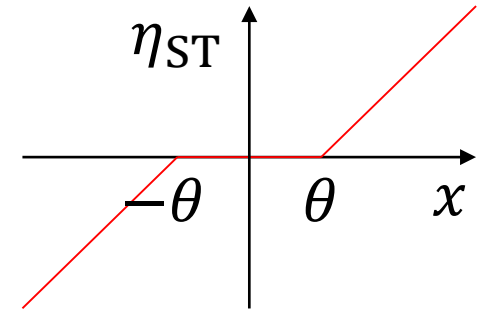
$$f(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x \leq 0. \end{cases} \quad f'(x) = \delta(x).$$

$$\therefore - \int f(x)\phi'(x)dx = - \int_0^{\infty} \phi'(x)dx = \phi(0) = \int \delta(x)\phi(x)dx.$$

最後の等号はデルタ関数の定義である。

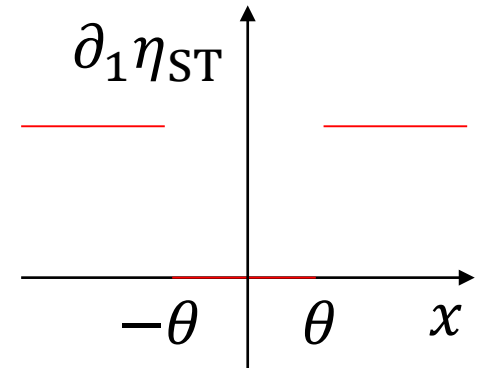
ST関数の微分

$$\eta_t(x) = \eta_{ST}(x; \theta) = \begin{cases} x - \theta & \text{for } x > \theta \\ 0 & \text{for } |x| \leq \theta \\ x + \theta & \text{for } x < -\theta. \end{cases}$$



一階微分

$$\partial_1 \eta_{ST} \equiv \frac{\partial \eta_{ST}}{\partial x} = \begin{cases} 1 & \text{for } x > \theta \\ 0 & \text{for } |x| \leq \theta \\ -1 & \text{for } x < -\theta. \end{cases}$$



$$\partial_2 \eta_{ST} \equiv \frac{\partial \eta_{ST}}{\partial \theta} = \begin{cases} -1 & \text{for } x > \theta \\ 0 & \text{for } |x| \leq \theta \\ 1 & \text{for } x < -\theta. \end{cases}$$

二階微分

$$\partial_1^2 \eta_{ST} = \delta(x - \theta) - \delta(x + \theta).$$

準備2: Steinの補題

$Z \sim \mathcal{N}(0, \sigma^2)$ とすると、

$$\mathbb{E}[Zf(Z)] = \sigma^2 \mathbb{E}[f'(Z)].$$

証明

$$p(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}},$$

$$\begin{aligned} \mathbb{E}[Zf(Z)] &= \int zf(z)p(z)dz = -\sigma^2 \int f(z)p'(z)dz \\ &= \sigma^2 \int f'(z)p(z)dz = \sigma^2 \mathbb{E}[f'(Z)]. \quad \blacksquare \end{aligned}$$

閾値の評価

$$\Psi(v, \lambda) = \mathbb{E}[\{\eta(x_1 + \sqrt{v}\omega_1; \lambda\sqrt{v}) - x_1\}^2].$$

準備

$$\frac{\partial}{\partial v} \eta = \frac{\omega_1}{2\sqrt{v}} \partial_1 \eta + \frac{\lambda}{2\sqrt{v}} \partial_2 \eta, \quad \frac{\partial \eta}{\partial \omega_1} = \sqrt{v} \partial_1 \eta.$$

導関数の評価

$$\begin{aligned} \frac{\partial \Psi}{\partial v} &= \mathbb{E} \left[2(\eta - x_1) \frac{\partial \eta}{\partial v} \right] \\ &= \frac{1}{\sqrt{v}} \mathbb{E}[\omega_1(\eta - x_1) \partial_1 \eta] + \frac{\lambda}{\sqrt{v}} \mathbb{E}[(\eta - x_1) \partial_2 \eta] \\ &= \mathbb{E}[(\partial_1 \eta)^2] + \mathbb{E}[(\eta - x_1) \partial_1^2 \eta] + \frac{\lambda}{\sqrt{v}} \mathbb{E}[(\eta - x_1) \partial_2 \eta]. \end{aligned}$$

最後の等号はSteinの補題から従う。

閾値の評価

スライド13に示した微分結果を使うと、

$$\begin{aligned}\frac{\partial \Psi}{\partial v} &= (1 + \lambda^2) \mathbb{E} \left[P_G \left(\frac{x_1 - \lambda \sqrt{v}}{\sqrt{v}} \right) + P_G \left(-\frac{x_1 + \lambda \sqrt{v}}{\sqrt{v}} \right) \right] \\ &\quad - \mathbb{E} \left[\left(\lambda + \frac{x_1}{\sqrt{v}} \right) p_G \left(\frac{\lambda \sqrt{v} - x_1}{\sqrt{v}} \right) \right] \\ &\quad + \mathbb{E} \left[\left(\frac{x_1}{\sqrt{v}} - \lambda \right) p_G \left(\frac{x_1 + \lambda \sqrt{v}}{\sqrt{v}} \right) \right].\end{aligned}$$

閾値の評価

$$\begin{aligned}\lim_{v \downarrow 0} \frac{\partial \Psi}{\partial v} &= (1 + \lambda^2) \{ \Pr(x_1 > 0) + \Pr(x_1 < 0) \} \\ &\quad + 2 \{ (1 + \lambda^2) P_G(-\lambda) - \lambda p_G(\lambda) \} \Pr(x_1 = 0) \\ &= (1 + \lambda^2) \{ \rho + 2 P_G(-\lambda) (1 - \rho) \} - 2 \lambda p_G(\lambda) (1 - \rho).\end{aligned}$$

最後の等号で、 $\Pr(x_1 = 0) = 1 - \rho$ 、 $\Pr(x_1 \neq 0) = \rho$ を使った。

閾値の条件

$$\frac{1}{\delta} \lim_{v \downarrow 0} \frac{\partial}{\partial v} \Psi(v; \lambda_t) < 1$$

$$\iff \rho < \frac{\delta - 2[(1 + \lambda^2) P_G(-\lambda) - \lambda p_G(\lambda)]}{(1 + \lambda^2) - 2[(1 + \lambda^2) P_G(-\lambda) - \lambda p_G(\lambda)]}. \quad \blacksquare$$