メッセージ伝播法入門
後半第3回講義資料
近似的メッセージ伝播法

広島市立大学
令和４年９月26日～９月28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

TOYOHASHI
UNIVERSITY OF TECHNOLOGY

ベイズ最適AMP

更新式

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

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

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

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

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

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

信号の事前分布が既知である必要がある。

AMP[3-1]

更新式

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

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

初期値

$$\boldsymbol{x}^0 = \boldsymbol{0}, \qquad \boldsymbol{z}^{-1} = \boldsymbol{0}.$$

Denoiser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.

オンサーガ(Onsager)項

オンサーガ項

大システム極限において、誤差の漸近ガウス性を成立させるための補正項

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

L. Onsager (1903—1976)

オンサーガ相反関係(不可逆過程の熱力学の基礎)を発見した貢献により、
1968年にノーベル化学賞を受賞

名前の由来

項の名前は、液体論に関わるオンサーガの1936年の論文でξ_t提案された
方法論に由来する。

物理学者は、彼の方法論に従って導出される反跳場(reaction field)を
オンサーガ反跳場と呼ぶ。

オンサーガのキャビティ(Cavity)アプローチ

キャビティ場(Cavity field)

n番目の信号を取り除いた$M \times (N-1)$キャビティ観測システムにおいて、干渉除去後に残っている干渉雑音

AWGNとみなせることを仮定する。

反跳場(Reaction field)

キャビティ観測システムにn番目の信号を追加したときに、

追加した信号がキャビティ場に与える無視できない影響

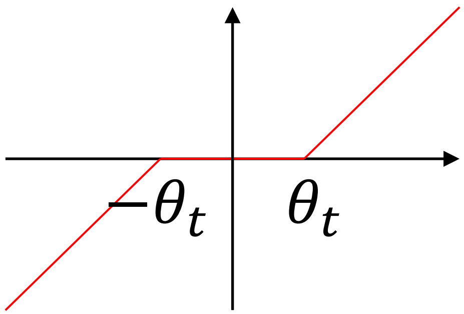
$[\boldsymbol{A}^{\mathrm{T}}(\delta^{-1}\xi_{t-1}\boldsymbol{z}^{t-1})]_n$のことだと思えばよい。

真の残留干渉$[\hat{\boldsymbol{x}}^t + \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}^t)]_n - x_n$に反跳場による影響の補正を加えた$[\hat{\boldsymbol{x}}^t + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z}^t]_n - x_n$は、$x_n$と独立なキャビティ場とみなせるはず。

Denoiserの例

Soft thresholding (ST)

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



ミニマックス最適性[3-2]

任意の$p, q > 0$に対して、 $$\lim_{a \downarrow 0} \frac{r_{\mathrm{ST}}(a)}{r_{\mathrm{opt}}(a)} = 1\,.$$

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

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

[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

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

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

$M_{p,\rho}$がρに依存しない信号分布では、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

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

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

閾値を$\theta_t = \lambda\sqrt{\bar{v}^t}$とおいた。

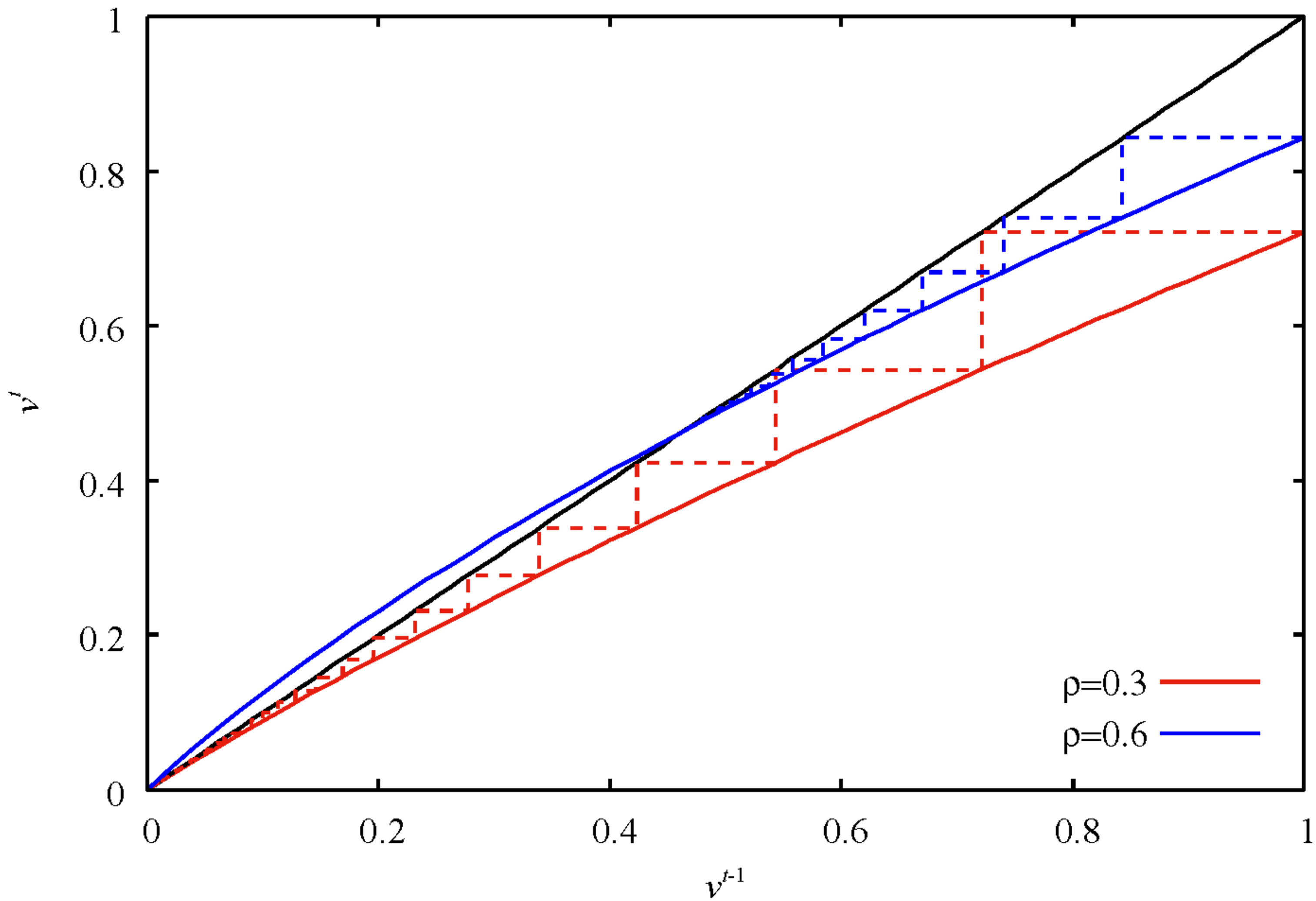
誤差の漸近評価

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

状態発展法の解説は、後半第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

閾値[2-4, 3-1]

定義

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

ただし、

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

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

$$\rho_{\mathrm{SE}} = \sup_{\lambda>0} \frac{\delta - 2[(1+\lambda^2)P_{\mathrm{G}}(-\lambda) - \lambda p_{\mathrm{G}}(\lambda)]}{(1+\lambda^2) - 2[(1+\lambda^2)P_{\mathrm{G}}(-\lambda) - \lambda p_{\mathrm{G}}(\lambda)]}.$$

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

$$\lambda = \lambda_{\mathrm{opt}} \equiv \operatorname*{argsup}_{\lambda>0} \frac{\delta - 2[(1+\lambda^2)P_{\mathrm{G}}(-\lambda) - \lambda p_{\mathrm{G}}(\lambda)]}{(1+\lambda^2) - 2[(1+\lambda^2)P_{\mathrm{G}}(-\lambda) - \lambda p_{\mathrm{G}}(\lambda)]}.$$

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$\bar{v} = \delta^{-1}\Psi(\bar{v}, \lambda)$は解$\bar{v} = 0$を持つ。

性質２

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

性質１と性質２から、

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

上記の$\rho_{\mathrm{SE}}(\lambda, p(x_1), \delta)$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。

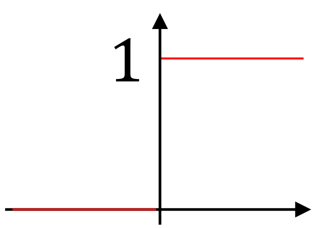
$$\int Df(x)\phi(x)dx = -\int f(x)\phi'(x)dx\,.$$

意義

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

部分積分

例



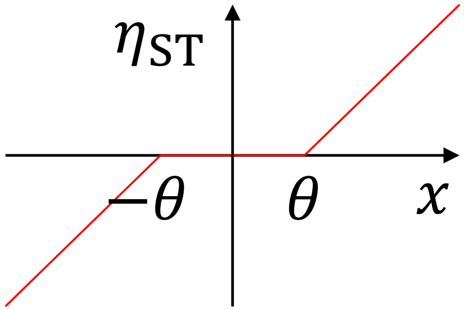
$$f(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x \leq 0. \end{cases} \qquad f'(x) = \delta(x).$$

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

最後の等号はデルタ関数の定義である。

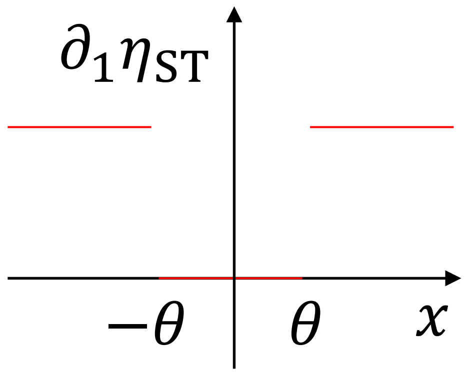
ST関数の微分

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



一階微分

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



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

二階微分

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

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

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

$$\mathbb{E}[Zf(Z)] = \int zf(z)p(z)dz \ = -\sigma^2 \int f(z)p'(z)dz$$

$$= \sigma^2 \int f'(z)p(z)dz = \sigma^2 \mathbb{E}[f'(Z)]. \qquad \blacksquare$$

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

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。

TOYOHASHI UNIVERSITY OF TECHNOLOGY

閾値の評価

スライド１５に示した微分結果を使うと、

$$
\frac{\partial \Psi}{\partial v} = (1+\lambda^2)\mathbb{E}\left[P_{\mathrm{G}}\left(\frac{x_1 - \lambda\sqrt{v}}{\sqrt{v}}\right) + P_{\mathrm{G}}\left(-\frac{x_1 + \lambda\sqrt{v}}{\sqrt{v}}\right)\right]
$$

$$
-\mathbb{E}\left[\left(\lambda + \frac{x_1}{\sqrt{v}}\right) p_{\mathrm{G}}\left(\frac{\lambda\sqrt{v} - x_1}{\sqrt{v}}\right)\right]
$$

$$
+\mathbb{E}\left[\left(\frac{x_1}{\sqrt{v}} - \lambda\right) p_{\mathrm{G}}\left(\frac{x_1 + \lambda\sqrt{v}}{\sqrt{v}}\right)\right].
$$

閾値の評価

$$\lim_{v \downarrow 0} \frac{\partial \Psi}{\partial v} = (1+\lambda^2)\{\Pr(x_1 > 0) + \Pr(x_1 < 0)\}$$
$$+2\{(1+\lambda^2)P_{\mathrm{G}}(-\lambda) - \lambda p_{\mathrm{G}}(\lambda)\}\Pr(x_1 = 0)$$
$$= (1+\lambda^2)\{\rho + 2P_{\mathrm{G}}(-\lambda)(1-\rho)\} - 2\lambda p_{\mathrm{G}}(\lambda)(1-\rho).$$

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

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

$$\Longleftrightarrow \quad \rho < \frac{\delta - 2[(1+\lambda^2)P_{\mathrm{G}}(-\lambda) - \lambda p_{\mathrm{G}}(\lambda)]}{(1+\lambda^2) - 2[(1+\lambda^2)P_{\mathrm{G}}(-\lambda) - \lambda p_{\mathrm{G}}(\lambda)]}. \quad \blacksquare$$