メッセージ伝播法入門
第１回講義資料
導入

広島市立大学
令和５年９月25日～９月26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

本講義の内容

1. 第１回から第３回

   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.

   S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Aug. 2011, pp. 2168–2172.

2. 第４回から第６回

   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.

本講義の心得

本講義の目的

学部の教養科目として習った微分積分学、線形代数、確率論等の基礎数学が、最先端の研究にどう活用されているかを学ぶ。

難易度　最高難度？（基礎数学の連続に耐える体力が必要）

論文は、教養レベルの数学に精通している前提で書かれている。

- 微分積分学：積分（高校レベル）、偏微分
- 線形代数：固有分解、特異値分解、ノルム
- 確率論：大数の法則、中心極限定義

難しいと思ったら

数学的議論の詳細ではなく、議論のアイデアを把握せよ。

各回の講義は独立なので、敗者復活できる。

数理モデル

線形観測モデル

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}, \qquad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_M)$$

$\boldsymbol{y} = (y_1, \dots, y_M)^{\mathrm{T}} \in \mathbb{R}^M$: M次元観測ベクトル

$\boldsymbol{x} = (x_1, \dots, x_N)^{\mathrm{T}} \in \mathbb{R}^N$: N次元信号ベクトル

$\boldsymbol{w} = (w_1, \dots, w_M)^{\mathrm{T}} \in \mathbb{R}^M$: M次元ノイズベクトル

$\boldsymbol{A} = [A_{mn}] \in \mathbb{R}^{M \times N}$: M行N列の観測行列

目標

$\boldsymbol{y}$、$\boldsymbol{A}$、未知の確率変数の統計的性質に関する情報を使って、最小限の計算量かつ最良な性能で信号ベクトル$\boldsymbol{x}$を推定せよ。

本講義を通じての仮定

信号ベクトル

$\boldsymbol{x}$は平均0分散1の独立同一分布に従う(i.i.d.)成分からなる。

観測行列

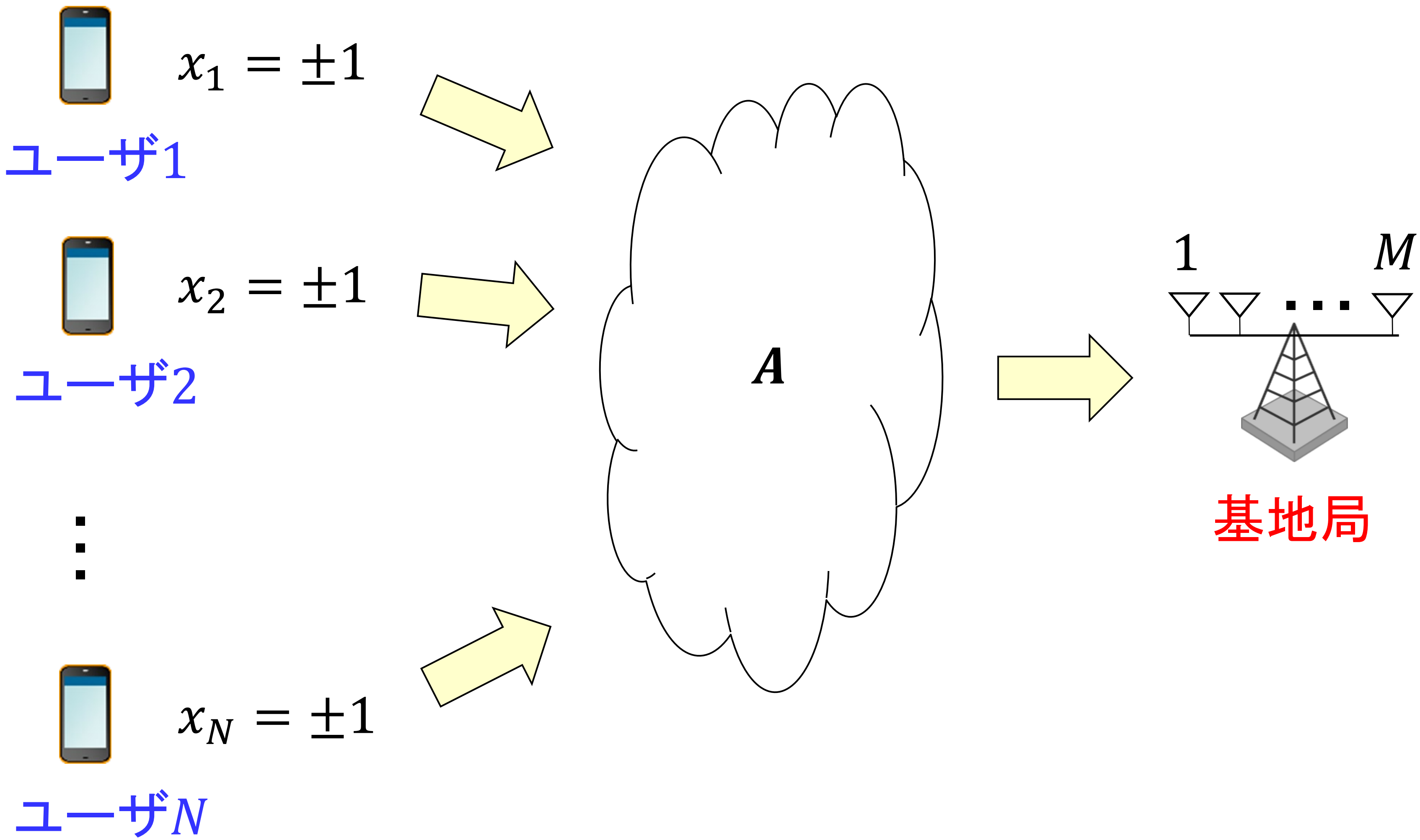
$\boldsymbol{A}$は平均0分散M^{-1}のガウス分布に従う
独立な成分からなる。

電力の規格化

$$\frac{1}{N}\mathbb{E}[\|\boldsymbol{A}\boldsymbol{x}\|^2] = \frac{1}{N}\mathbb{E}\left[\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\right)\right] = \frac{1}{N}\mathbb{E}\left[\mathrm{Tr}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)\right] = 1.$$

TOYOHASHI
UNIVERSITY OF TECHNOLOGY

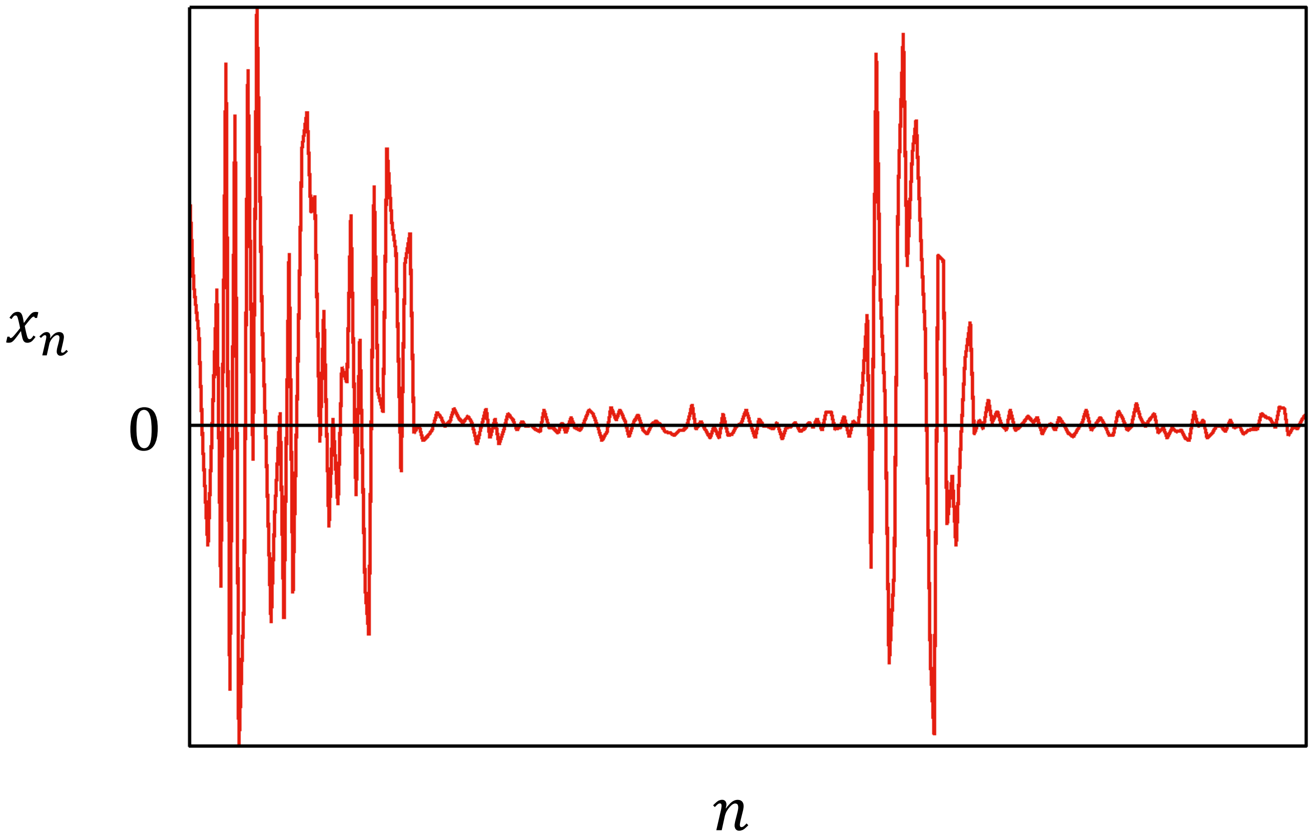
Multiple-input multiple-output (MIMO)



圧縮センシング

K-スパース信号

信号ベクトル$\boldsymbol{x}$の非零要素数は高々Kである。



スパース重ね合わせ符号[1-1]

セクション数LブロックサイズBの符号語

$$\boldsymbol{c} = \boldsymbol{A}\boldsymbol{x} = [\boldsymbol{A}[1] \cdots \boldsymbol{A}[L]] \begin{bmatrix} \boldsymbol{x}[1] \\ \vdots \\ \boldsymbol{x}[L] \end{bmatrix}$$

$x[l] \in \mathbb{R}^B$:非零要素が1のB次元1‑スパース情報ベクトル

符号化率

$$R = \frac{L}{M} \log B$$

[1-1] A. Joseph and A. R. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58,no. 5, pp. 2541-2557, May 2012.

情報理論的スパース性[1-2]

ノイズなしの場合

ノイズがない場合を考え、信号成分$\{x_n\}$は特異成分を持たない独立同一分布(i.i.d.)に従うと仮定する。

圧縮率$\delta = M/N$を固定して次元MとNを無限大にした大システム極限において、任意の$\epsilon \in (0,1)$に対して信号再構成に失敗する確率がϵ以下となるような観測行列$\boldsymbol{A}$と再構成方法が存在するための必要十分条件は、圧縮率δがレニー情報次元$d(x_1)$以上であることである。

[1-2] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721-3748, Aug. 2010.

レニー情報次元[1-3]

以下、$H(\lfloor X \rfloor) < \infty$を満たすとする。

確率変数Xのレニー情報次元d

レニーエントロピー

$$H(n^{-1}\lfloor nX \rfloor) = d \log n + h + o(1) \ \text{ as } n \to \infty.$$

離散確率変数Xの場合

$$H(n^{-1}\lfloor nX \rfloor) = 0 \log n + H(X) + o(1).$$

絶対連続な確率変数Xの場合

$$H(n^{-1}\lfloor nX \rfloor) = 1 \log n + h(X) + o(1).$$

混合分布Pに従う確率変数Xの場合

$$H(n^{-1}\lfloor nX \rfloor) = \rho \log n + O(1).$$

$P = (1-\rho)P_{\mathrm{d}} + \rho P_{\mathrm{c}}$ (P_{d}:離散分布、P_{c}:絶対連続分布)

[1-3] A. Rényi, “On the dimension and entropy of probability distributions,” *Acta Mathematica Hungarica*, vol. 10, no. 1-2, Mar. 1959.

詳細な解析[1-4]

ノイズがない場合を考え、信号成分$\{x_n\}$は± 1を独立に等確率で取るものと仮定する。

圧縮率$\delta = M/N$が以下の不等式を満たすならば、

$$\frac{M}{N} > \frac{2}{\log_3 N}$$

次元MとNを無限大した極限において、ビット誤り確率が0に収束するような観測行列$\boldsymbol{A}$と信号再構成方法が存在する。

注意

$\boldsymbol{A}$をi.i.d.ガウス行列に取れば良い。

[1-4] M. A. Sedaghat, R. R. Müller, and F. Marvasti, “On Optimum Asymptotic Multiuser Efficiency of Randomly Spread CDMA,” *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6635-6642, Dec. 2015.

ノイズがある場合

ノイズがある場合の性能限界の評価は困難である。

レプリカ法[1-5, 1-6]

- 統計力学で開発された計算手法である。
- 他の厳密な手法では評価が困難な問題を解ける。
- いくつかの未証明な仮定を使う。

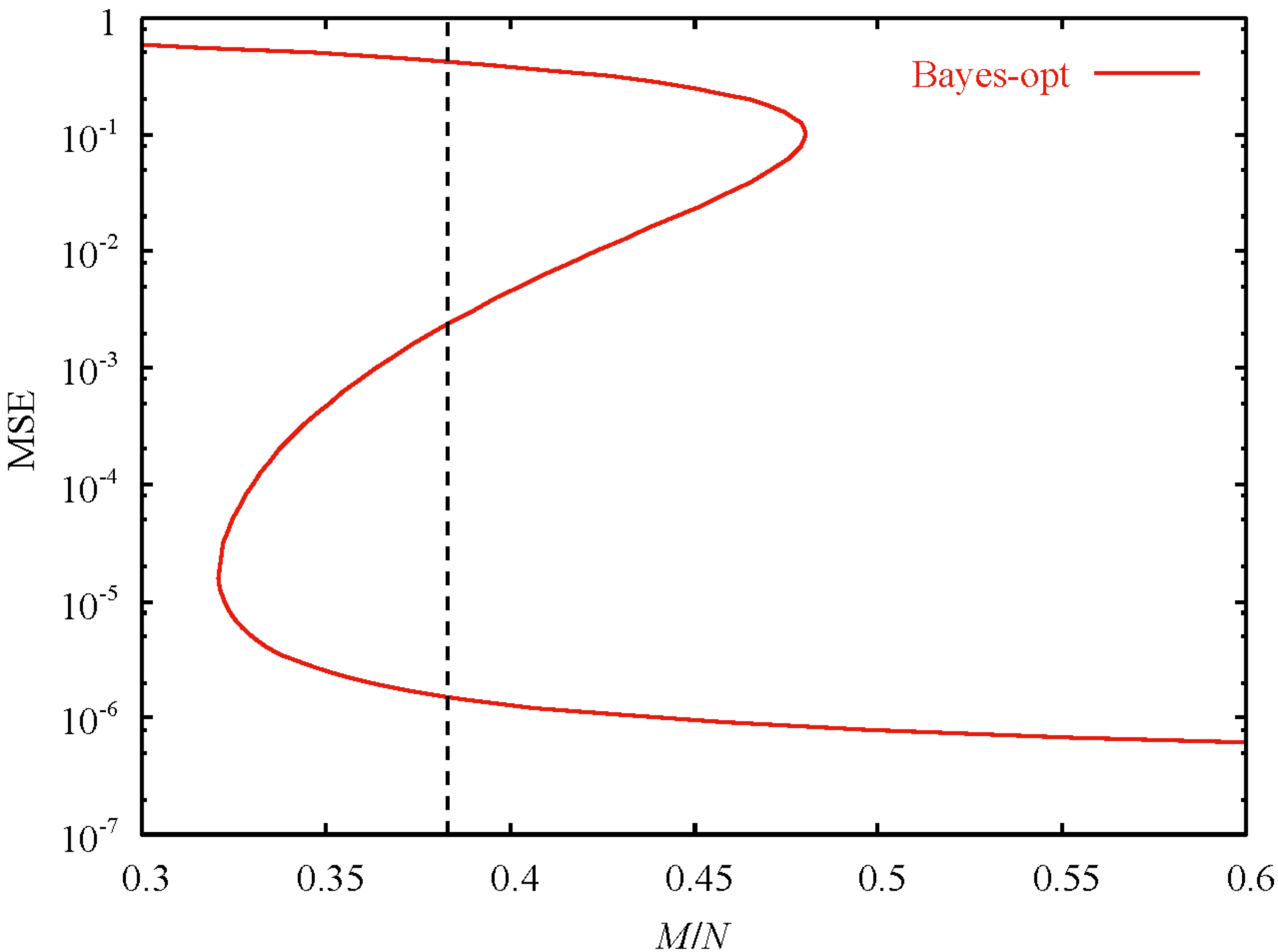
仮定の一例

モーメント系列$\{\mathbb{E}[X^n]\}_{n=1}^{\infty}$から、期待値$\mathbb{E}[X^\alpha]$$(\alpha < 1)$を外挿できる。

[1-5] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[1-6] D. Guo and S. Verdú, “Randomly spread CDMA: Asymptotics via statistical physics,” *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.

ノイズがある場合[1-5, 1-6]



信号密度$\rho = 0.3$のBernoulli–Gauss(BG)分布、SNR: $1/\sigma^2 = 60$ dB.

厳密性に関するコメント

文献[1−7, 1−8]の結果から、前ページの結果は正しい。

相互情報量$I(x; y|A)$に関する厳密解

[1-7] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for random linear estimation with Gaussian matrices is exact," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2252-2283, Apr. 2019.

[1-8] J. Barbier, N. Macris, M. Dia and F. Krzakala, "Mutual information and optimality of approximate message-passing in random linear estimation," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4270-4303, Jul. 2020.

一般化線形観測モデルに関する厳密解（第3回資料参照）

[1-9] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, “Optimal errors and phase transitions in high-dimensional generalized linear models,” *Proc. Nat. Acad. Sci.*, vol. 116, no. 12, pp. 5451–5460, Mar. 2019.

ベイズ最適な推定

$\hat{\boldsymbol{x}} \in \mathbb{R}^N$を信号ベクトル$\boldsymbol{x}$の推定量とする。

最適性の基準・・・平均二乗誤差（MSE）

$$\hat{\boldsymbol{x}}_{\mathrm{opt}} = \underset{\hat{\boldsymbol{x}} \in \mathbb{R}^N}{\operatorname{argmin}} \mathbb{E}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2 | \boldsymbol{A}]$$

事後平均推定量$\hat{\boldsymbol{x}}_{\mathrm{PME}} = \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}]$は最適である。

証明

$$\begin{aligned}
\mathbb{E}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2] &= \mathbb{E}_{\boldsymbol{y}}\left[\mathbb{E}_{\boldsymbol{x}}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}_{\mathrm{PME}} + \hat{\boldsymbol{x}}_{\mathrm{PME}} - \hat{\boldsymbol{x}}\|^2 | \boldsymbol{y}]\right] \\
&= \mathbb{E}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}_{\mathrm{PME}}\|^2] + \mathbb{E}[\|\hat{\boldsymbol{x}}_{\mathrm{PME}} - \hat{\boldsymbol{x}}\|^2] \\
&\quad + 2\mathbb{E}_{\boldsymbol{y}}\left[\mathbb{E}_{\boldsymbol{x}}[(\boldsymbol{x} - \hat{\boldsymbol{x}}_{\mathrm{PME}})^{\mathrm{T}}(\hat{\boldsymbol{x}}_{\mathrm{PME}} - \hat{\boldsymbol{x}})]\right] \\
&= \mathbb{E}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}_{\mathrm{PME}}\|^2] + \mathbb{E}[\|\hat{\boldsymbol{x}}_{\mathrm{PME}} - \hat{\boldsymbol{x}}\|^2] \geq \mathbb{E}[\|\boldsymbol{x} - \hat{\boldsymbol{x}}_{\mathrm{PME}}\|^2]
\end{aligned}$$

$\boldsymbol{A}$に関する条件付けを省略した。 ■

事後平均推定量

事後平均推定量の定義

$$\hat{\boldsymbol{x}}_{\mathrm{PME}} = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}) d\boldsymbol{x}$$

ベイズの公式

$$p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{A}) = \frac{p(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}) p(\boldsymbol{x})}{p(\boldsymbol{y}|\boldsymbol{A})}, \quad p(\boldsymbol{y}|\boldsymbol{A}) = \int p(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}$$

事後平均推定量

$$\hat{\boldsymbol{x}}_{\mathrm{PME}} = \frac{\int \boldsymbol{x} p(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}}{\int p(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x}}$$

事後平均推定量の性質

一般に、事後平均推定量はその誤差と無相関である。

$$\mathbb{E}[\hat{\boldsymbol{x}}_{\mathrm{PME}}(\hat{\boldsymbol{x}}_{\mathrm{PME}} - \boldsymbol{x})^{\mathrm{T}} | \boldsymbol{A}] = \boldsymbol{O}$$

証明

$\hat{\boldsymbol{x}}_{\mathrm{PME}}$は$\boldsymbol{y}$と$\boldsymbol{A}$の決定論的な関数なので、

$$\mathbb{E}[\hat{\boldsymbol{x}}_{\mathrm{PME}}\boldsymbol{x}^{\mathrm{T}} | \boldsymbol{A}] = \mathbb{E}\left[\hat{\boldsymbol{x}}_{\mathrm{PME}}\mathbb{E}[\boldsymbol{x}^{\mathrm{T}} | \boldsymbol{y}, \boldsymbol{A}] | \boldsymbol{A}\right] = \mathbb{E}\left[\hat{\boldsymbol{x}}_{\mathrm{PME}}\hat{\boldsymbol{x}}_{\mathrm{PME}}^{\mathrm{T}} | \boldsymbol{A}\right]$$

この式から

$$\mathbb{E}[\hat{\boldsymbol{x}}_{\mathrm{PME}}(\hat{\boldsymbol{x}}_{\mathrm{PME}} - \boldsymbol{x})^{\mathrm{T}} | \boldsymbol{A}]$$

$$= \mathbb{E}\left[\hat{\boldsymbol{x}}_{\mathrm{PME}}\hat{\boldsymbol{x}}_{\mathrm{PME}}^{\mathrm{T}} \middle| \boldsymbol{A}\right] - \mathbb{E}\left[\hat{\boldsymbol{x}}_{\mathrm{PME}}\hat{\boldsymbol{x}}_{\mathrm{PME}}^{\mathrm{T}} \middle| \boldsymbol{A}\right] = \boldsymbol{O} \qquad \blacksquare$$

事後平均推定量の問題点

計算量

$\boldsymbol{x}$が連続の場合・・・N重積分が必要

$\boldsymbol{x} \in \{\pm 1\}^N$の場合・・・少なくとも$2^N$回の足し算が必要

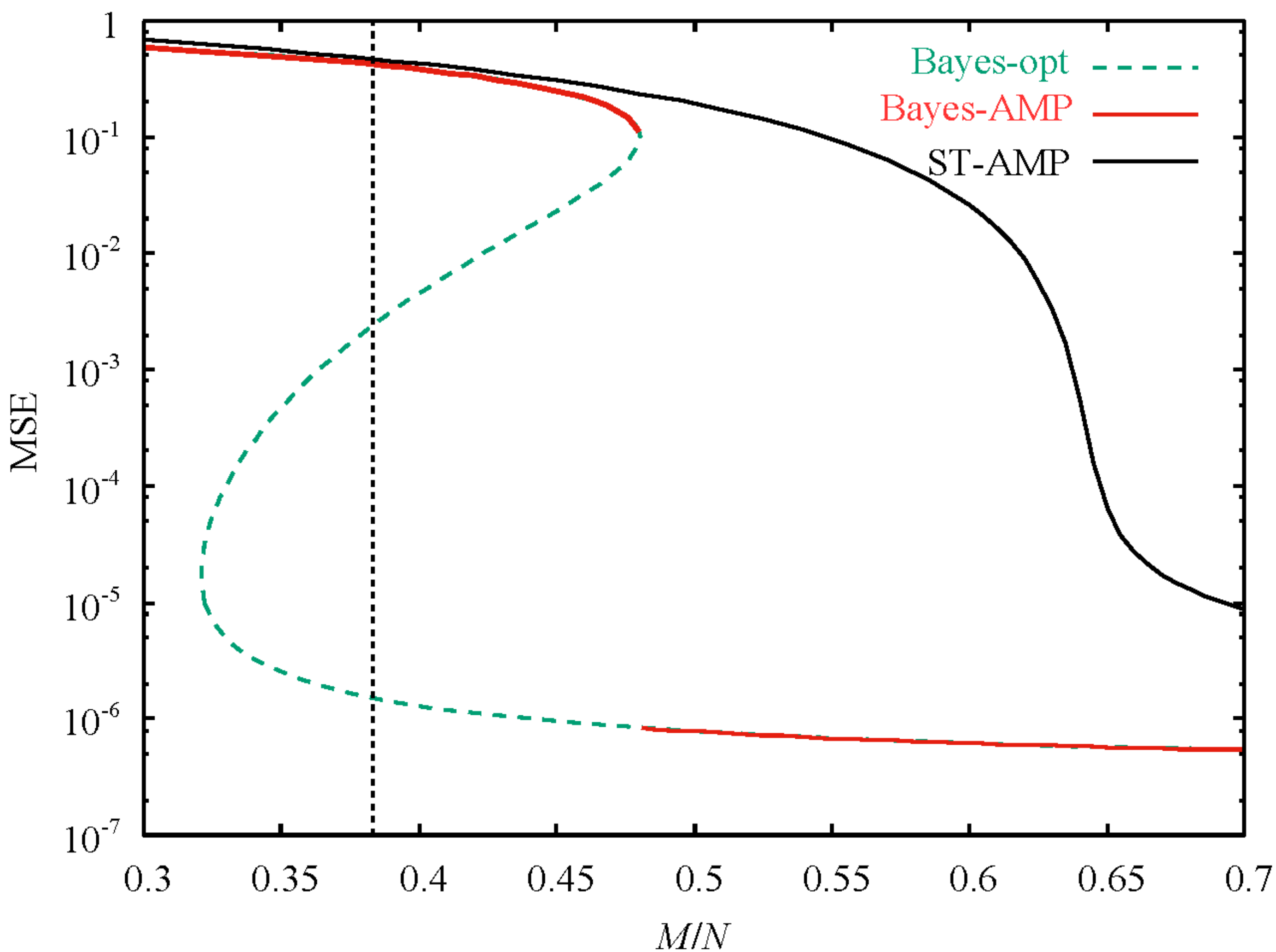
最小限の計算量

$\mathcal{O}(IMN)$の計算量で評価できる$\boldsymbol{x}$の推定量$\hat{\boldsymbol{x}}$は、
最小限の計算量で評価可能と言う。

$I : M$やNに直接依存しないパラメータ（反復回数）

本講義で扱うAMPは、最小限の計算量である。

数値的比較



信号密度$\rho = 0.3$のBG事前分布、SNR: $1/\sigma^2 = 60$ dB.