

大阪市立大学電子・物理工学特別講義

今日からできる スパースモデリング

京都大学大学院情報学研究科システム科学専攻

大関 真之

QUANTUM
ANNEALING



MACHINE
LEARNING

Sparse Modeling



京都大学
KYOTO UNIVERSITY

ボルツマン機械学習編

- ▶ 多数の数値列、データからの分布の推定
 - ▶ 生成モデルの存在の仮定

$$P(\mathbf{x}|\mathbf{u}) = \frac{1}{Z(\mathbf{u})} \exp \{ -E(\mathbf{x}|\mathbf{u}) \}$$

- ▶ 例：Ising模型

$$E(\mathbf{x}|\mathbf{u}) = \sum_{i=1}^N \sum_{j \in \partial i} J_{ij} x_i x_j + \sum_{i=1}^N h_i x_i$$



ボルツマン機械学習の基礎

▶ 最尤法

- ▶ データに対するモデルの尤もらしさ（対数尤度関数の経験平均）

$$L(\mathbf{u}) = \frac{1}{D} \sum_{d=1}^D \log P(\mathbf{x} = \mathbf{x}^{(d)} | \mathbf{u})$$



▶ 最尤法

- ▶ データに対するモデルの尤もらしさ（対数尤度関数の経験平均）

$$L(\mathbf{u}) = \frac{1}{D} \sum_{d=1}^D \log P(\mathbf{x} = \mathbf{x}^{(d)} | \mathbf{u})$$

- ▶ 統計力学のエントロピー最大原理と等価

$$L(\mathbf{u}) = -\frac{1}{D} \sum_{d=1}^D E(\mathbf{x} = \mathbf{x}^{(d)} | \mathbf{u}) - \log Z(\mathbf{u})$$



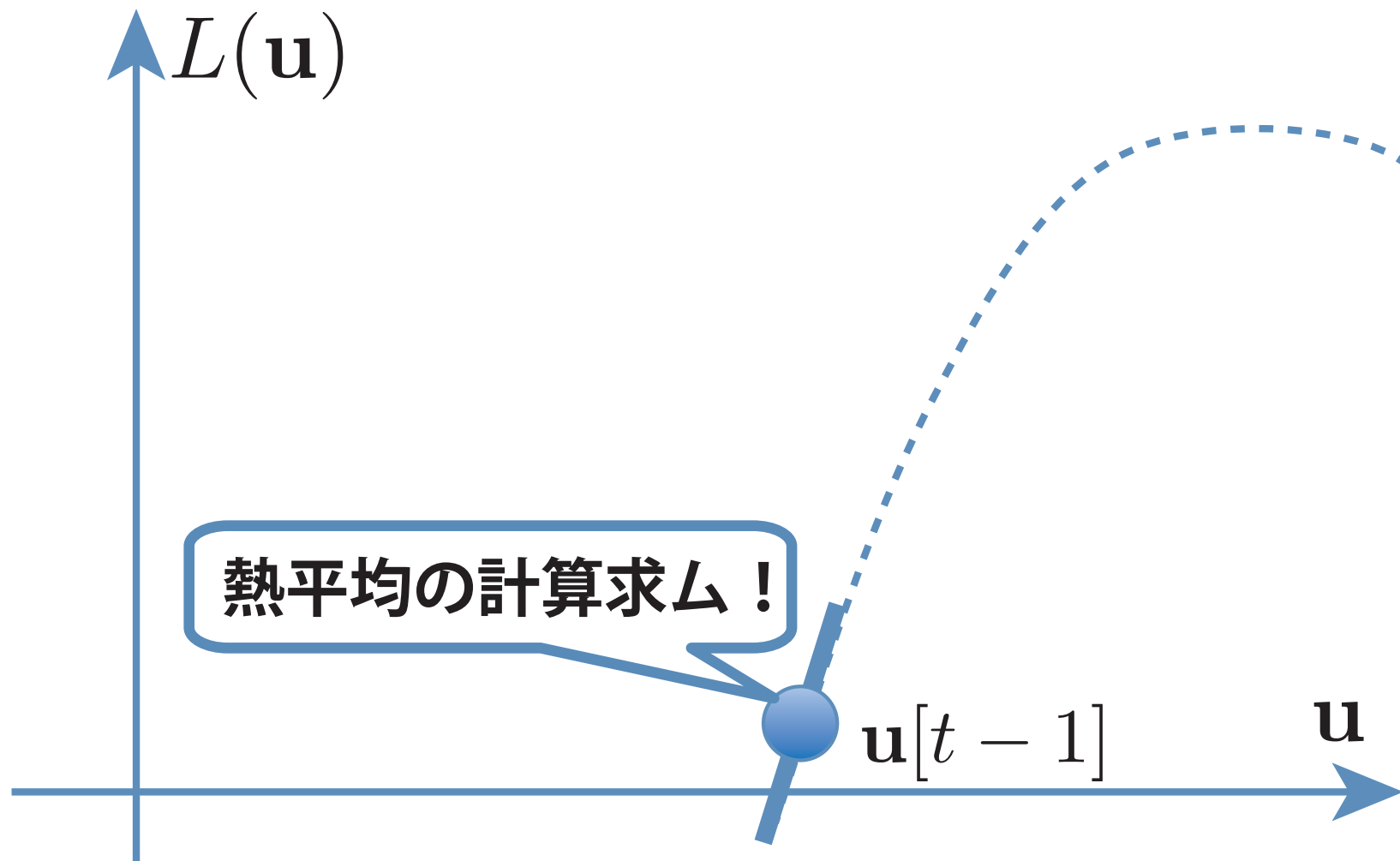
- ▶ 単純な勾配法の適用
 - ▶ 勾配方向に更新して最大値を求める

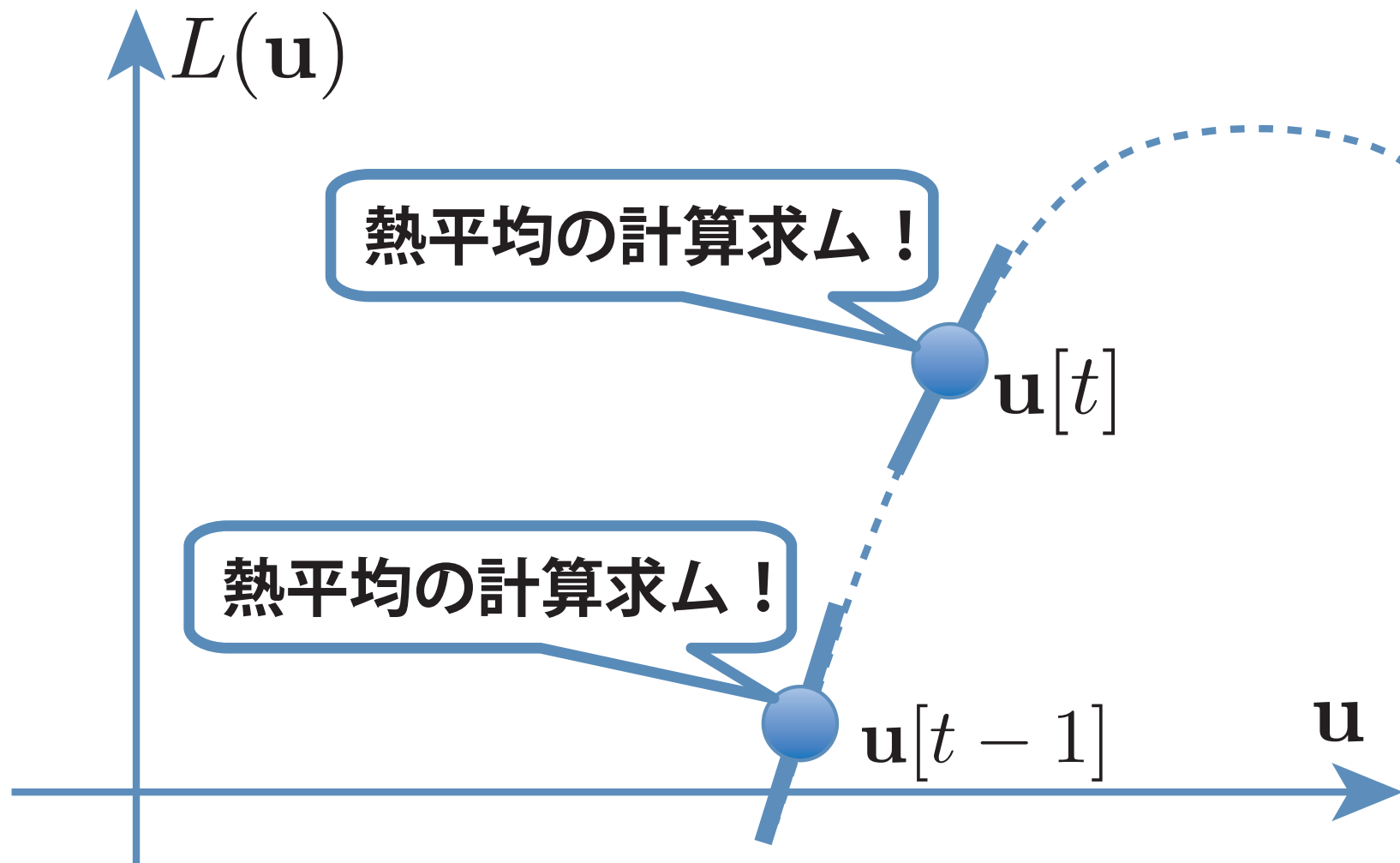
$$\mathbf{u}[t + 1] = \mathbf{u}[t] + \eta \frac{\partial L(\mathbf{u})}{\partial \mathbf{u}}$$

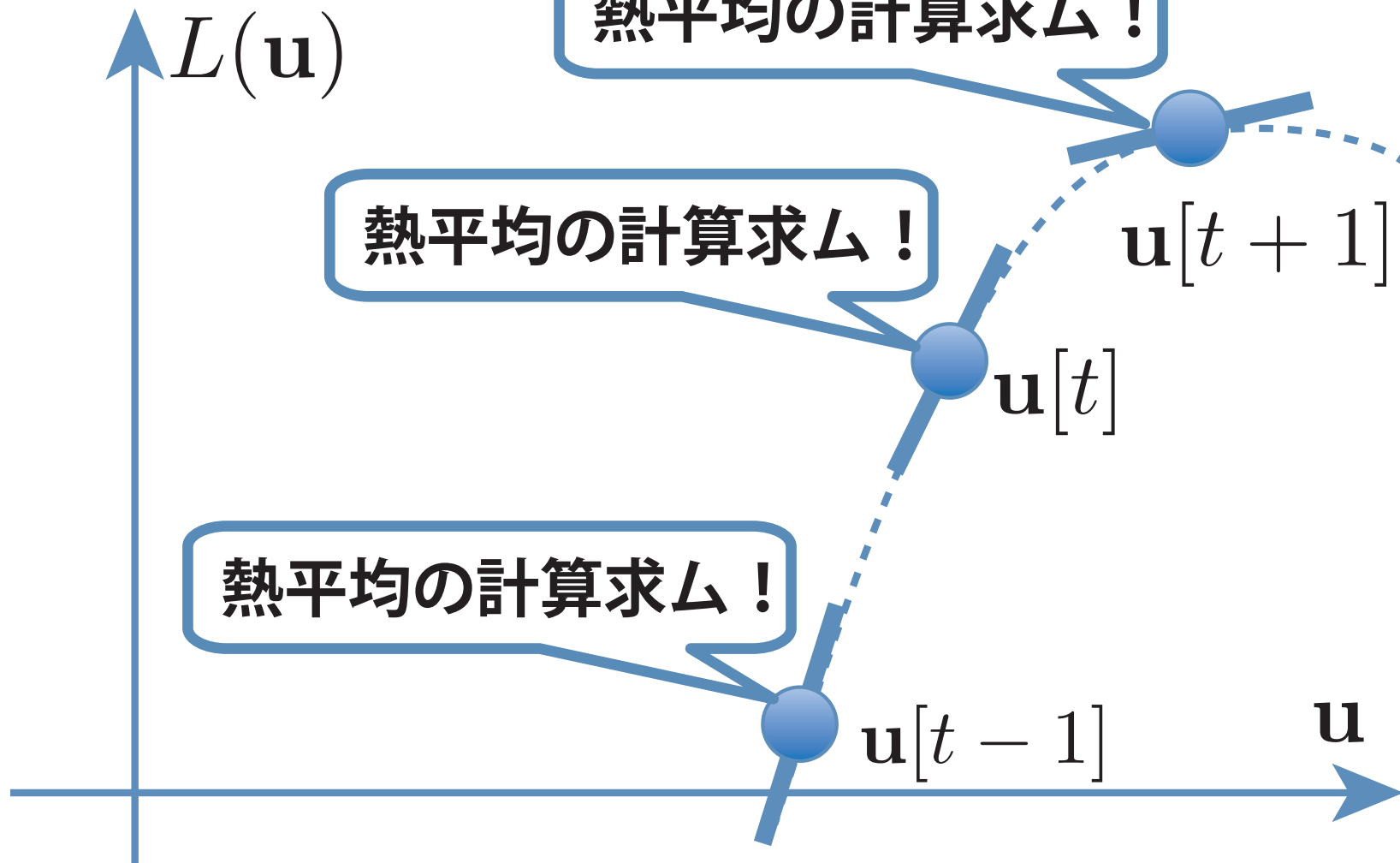
- ▶ η : 学習係数、更新幅を設定している
- ▶ じゃあ勾配を計算してみよう！

$$\frac{\partial L(\mathbf{u})}{\partial \mathbf{u}} = -\frac{1}{D} \sum_{d=1}^D \frac{\partial E(\mathbf{x} = \mathbf{x}^{(d)} | \mathbf{u})}{\partial \mathbf{u}} + \left\langle \frac{\partial E(\mathbf{x} | \mathbf{u})}{\partial \mathbf{u}} \right\rangle_{\mathbf{u}}$$









$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial \mathbf{u}} \right\rangle_{\mathbf{u}}$$

ボルツマン機械学習の危機

$$\log Z(\mathbf{u})$$

- ▶ イジング模型の場合
 - ▶ モーメントマッチング

$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial J_{ij}} \right\rangle_{\mathbf{u}} = \langle x_i x_j \rangle_{\mathbf{u}}$$

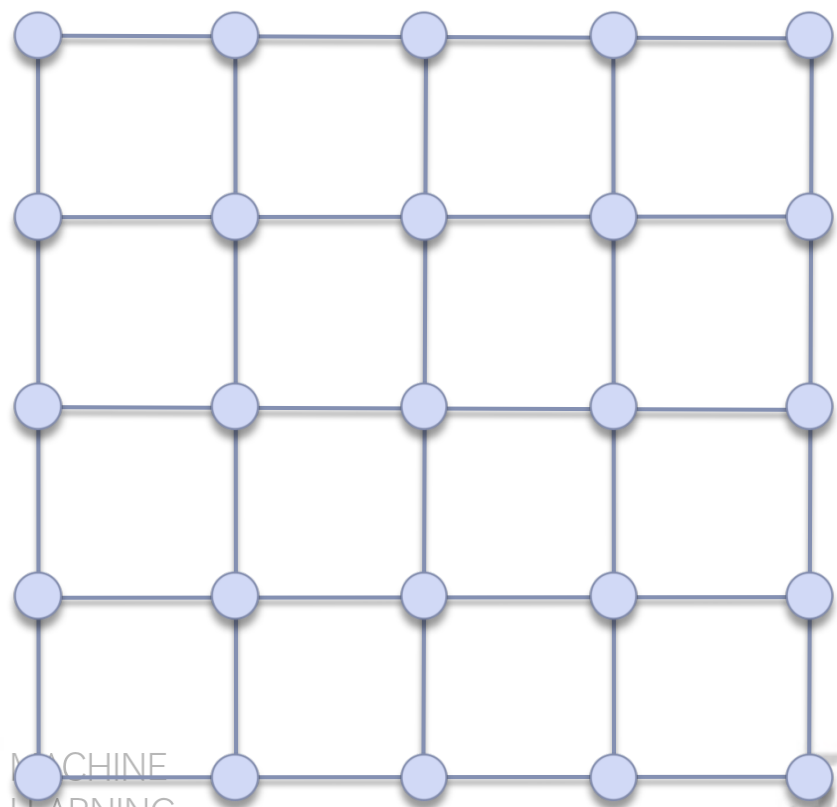
$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial h_i} \right\rangle_{\mathbf{u}} = \langle x_i \rangle_{\mathbf{u}}$$



最尤法に向けて

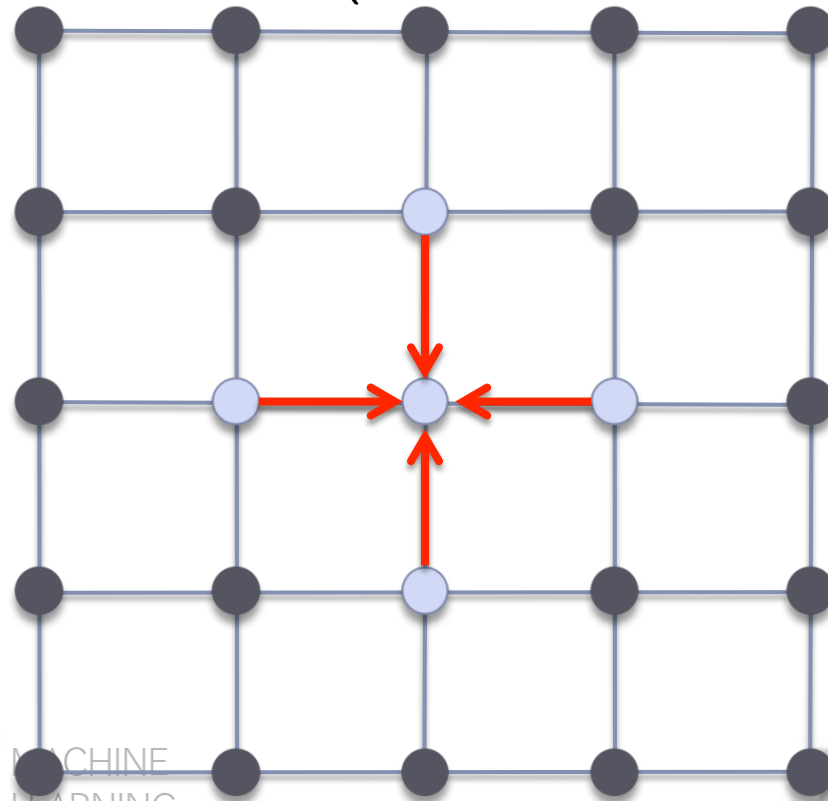
▶ 熱平均の計算

$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial h_i} \right\rangle_{\mathbf{u}} = \langle x_i \rangle_{\mathbf{u}}$$



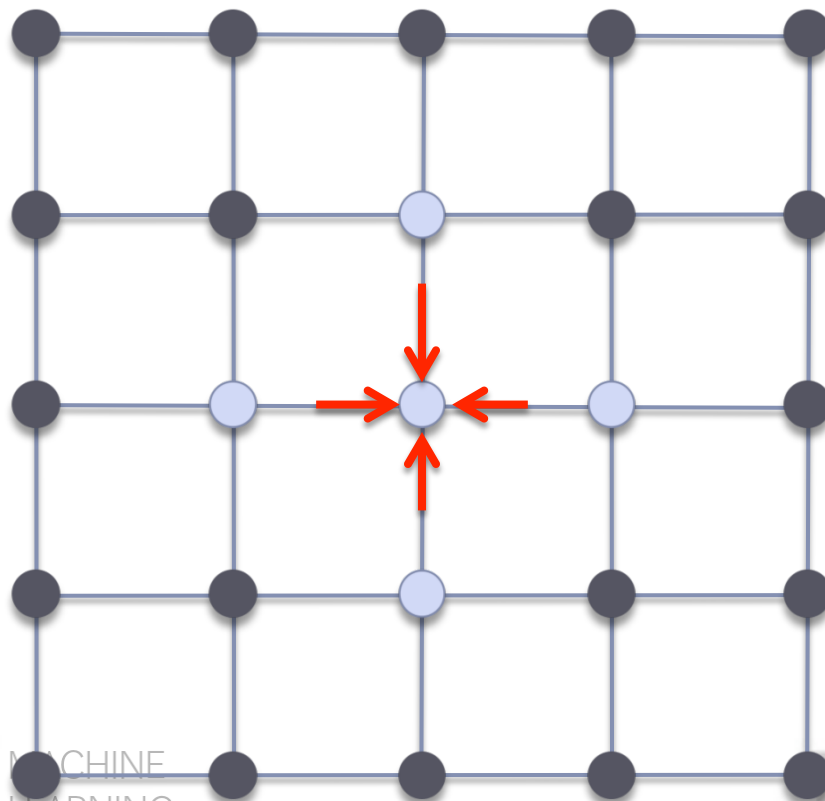
▶ 熱平均の近似

$$m_i = \tanh \left(\sum_{j \in \partial i} J_{ij} m_j + h_i \right)$$



▶ 熱平均の近似

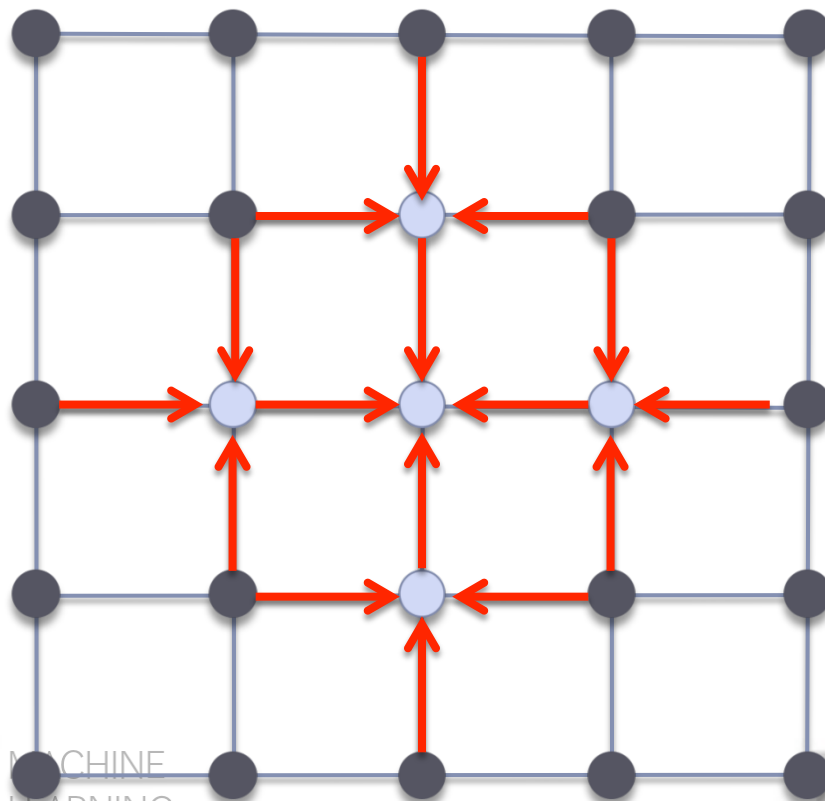
$$m_i = \tanh \left(\sum_{j \in \partial i} J_{ij} m_j + h_i + \sum_{j \in \partial i} J_{ij}^2 (1 - m_j)^2 m_i \right)$$



▶ 熱平均の近似

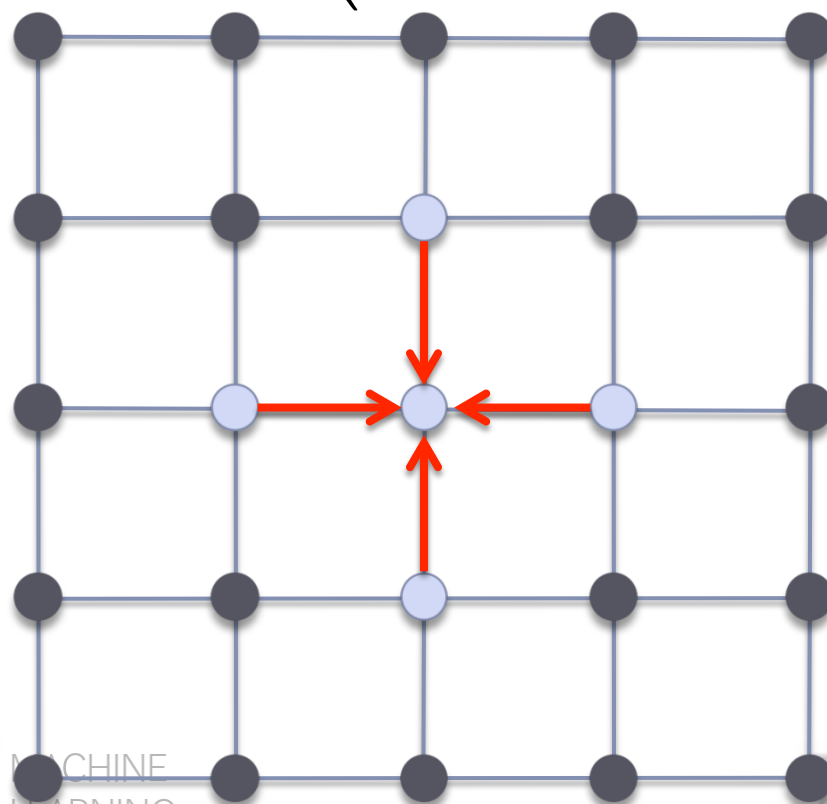
$$P(x_k) \propto f_k(x_k) \prod_{a \in \partial k} \tilde{M}_{a \rightarrow k}(x_k)$$

$$P(\mathbf{x}_{\partial a}) \propto f_a(\mathbf{x}_{\partial a}) \prod_{k \in \partial a} M_{k \rightarrow a}(x_k)$$



- ▶ 熱平均のストイックな近似

$$m_i = \tanh \left(\sum_{j \in \partial i} J_{ij} m_j + h_i \right)$$

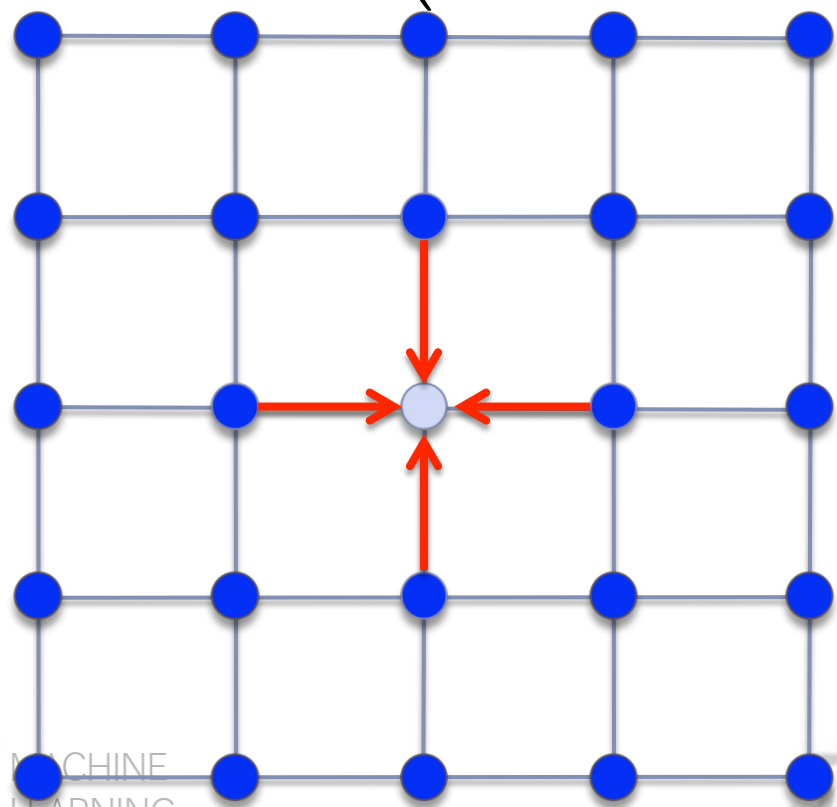


目標はデータに近い分布を探すこと

$$P_0(\mathbf{x}) = \frac{1}{D} \sum_{s=1}^D \delta(\mathbf{x} - \mathbf{x}^{(s)})$$

▶ 熱平均のデータによる近似

$$\langle x_i \rangle_{\mathbf{u}} \approx \tanh \left(\sum_{j \in \partial i} J_{ij} x_j^{(d)} + h_i \right)$$



あるものは使えばいいという思想が大事

$$P_0(\mathbf{x}) = \frac{1}{D} \sum_{s=1}^D \delta(\mathbf{x} - \mathbf{x}^{(s)})$$

あるものは使えばいいという思想が大事

$$P_0(\mathbf{x}) = \frac{1}{D} \sum_{s=1}^D \delta(\mathbf{x} - \mathbf{x}^{(s)})$$

データ駆動型

- ▶ 熱平均の計算が命
 - ▶ 勾配法の適用のために何度も熱平均の計算が必要

$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial J_{ij}} \right\rangle_{\mathbf{u}} = \langle x_i x_j \rangle_{\mathbf{u}}$$

$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial h_i} \right\rangle_{\mathbf{u}} = \langle x_i \rangle_{\mathbf{u}}$$

- ▶ 平均場近似は高速で良いが…近似は近似
- ▶ 疑似最尤法はデータが大量に必要…ビッグデータ時代以前は見向きされず



- ▶ 熱平均の計算が命
 - ▶ 勾配法の適用のために何度も熱平均の計算が必要

$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial J_{ij}} \right\rangle_{\mathbf{u}} = \langle x_i x_j \rangle_{\mathbf{u}}$$

$$\left\langle \frac{\partial E(\mathbf{x}|\mathbf{u})}{\partial h_i} \right\rangle_{\mathbf{u}} = \langle x_i \rangle_{\mathbf{u}}$$

- ▶ 平均場近似は高速で良いが…近似は近似
- ▶ 疑似最尤法はデータが大量に必要…ビッグデータ時代以前は見向きされず
- ▶ 低速だけど厳密な方法を追求！





マルコフ連鎖モンテカルロ法

N.Metropolis, et al: J.Chem. Phys. 21,(1953) 1087.

- ▶ 熱平均の計算法として標準的
- ▶ 離散時間マスター方程式

$$P_{[t+1]}(\mathbf{y}) = \sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x}) P_{[t]}(\mathbf{x})$$

- ▶ 十分な時間経過後に定常分布に到達





マルコフ連鎖モンテカルロ法

N.Metropolis, et al: J.Chem. Phys. 21,(1953) 1087.

- ▶ 熱平均の計算法として標準的
- ▶ 離散時間マスター方程式

$$P_{[t+1]}(\mathbf{y}) = \sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x}) P_{[t]}(\mathbf{x})$$

- ▶ 十分な時間経過後に定常分布に到達
- ▶ 定常分布をギブスボルツマン分布にしよう (釣り合い条件)

$$\sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x}) P_{ss}(\mathbf{x}|\mathbf{u}) = P_{ss}(\mathbf{y}|\mathbf{u})$$





マルコフ連鎖モンテカルロ法

N.Metropolis, et al: J.Chem. Phys. 21,(1953) 1087.

- ▶ 熱平均の計算法として標準的
 - ▶ 離散時間マスター方程式

$$P_{[t+1]}(\mathbf{y}) = \sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x}) P_{[t]}(\mathbf{x})$$

- ▶ 十分な時間経過後に定常分布に到達
- ▶ 定常分布をギブスボルツマン分布にしよう (釣り合い条件)

$$\sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x}) P_{ss}(\mathbf{x}|\mathbf{u}) = P_{ss}(\mathbf{y}|\mathbf{u})$$

- ▶ 初期緩和を捨てて、時系列による平均で期待値を計算

$$\langle f(\mathbf{x}) \rangle_{\mathbf{u}} \approx \frac{1}{T} \sum_{s=t_w+1}^{T+t_w} f(\mathbf{x}[s])$$





マルコフ連鎖モンテカルロ法

N.Metropolis, et al: J.Chem. Phys. 21,(1953) 1087.

- ▶ 詳細釣り合い条件

$$\frac{W_{\mathbf{u}}(\mathbf{y}|\mathbf{x})}{W_{\mathbf{u}}(\mathbf{x}|\mathbf{y})} = \frac{P_{\text{SS}}(\mathbf{y}|\mathbf{u})}{P_{\text{SS}}(\mathbf{x}|\mathbf{u})}$$





マルコフ連鎖モンテカルロ法

N.Metropolis, et al: J.Chem. Phys. 21,(1953) 1087.

- ▶ 詳細釣り合い条件

$$\frac{W_{\mathbf{u}}(\mathbf{y}|\mathbf{x})}{W_{\mathbf{u}}(\mathbf{x}|\mathbf{y})} = \frac{P_{\text{SS}}(\mathbf{y}|\mathbf{u})}{P_{\text{SS}}(\mathbf{x}|\mathbf{u})}$$

- ▶ 釣り合い条件を満たす十分条件 $\sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x})P_{\text{SS}}(\mathbf{x}|\mathbf{u}) = P_{\text{SS}}(\mathbf{y}|\mathbf{u})$





マルコフ連鎖モンテカルロ法

N.Metropolis, et al: J.Chem. Phys. 21,(1953) 1087.

- ▶ 詳細釣り合い条件

$$\frac{W_{\mathbf{u}}(\mathbf{y}|\mathbf{x})}{W_{\mathbf{u}}(\mathbf{x}|\mathbf{y})} = \frac{P_{SS}(\mathbf{y}|\mathbf{u})}{P_{SS}(\mathbf{x}|\mathbf{u})}$$

- ▶ 釣り合い条件を満たす十分条件 $\sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x})P_{SS}(\mathbf{x}|\mathbf{u}) = P_{SS}(\mathbf{y}|\mathbf{u})$

- ▶ 強い制限、しかし遷移確率を計算しやすい

- ▶ 例：メトロポリス法

$$W_{\mathbf{u}}(\bar{\mathbf{x}}|\mathbf{x}) = \min \left(1, \frac{P_{SS}(\bar{\mathbf{x}}|\mathbf{u})}{P_{SS}(\mathbf{x}|\mathbf{u})} \right) = \min \left(1, \exp \left(-2 \sum_{j \in \partial i} J_{ij} x_i x_j - 2h_i x_i \right) \right)$$

- ▶ 例：熱浴法（グラウバーダイナミクス）

$$W_{\mathbf{u}}(\bar{\mathbf{x}}|\mathbf{x}) = \frac{P_{SS}(\bar{\mathbf{x}})}{P_{SS}(\mathbf{x}|\mathbf{u}) + P_{SS}(\bar{\mathbf{x}}|\mathbf{u})} = \frac{1}{2} \left\{ 1 - \tanh \left(\sum_{j \in \partial i} J_{ij} x_i x_j + h_i x_i \right) \right\}$$





マルコフ連鎖モンテカルロ法

N.Metropolis, et al: J.Chem. Phys. 21,(1953) 1087.

1. $\mathbf{x}[0]$ を初期化する. 例えば一様分布からサンプリングする.
2. スピンを反転するサイト i を選ぶ.
3. 反転に必要なエネルギー $\Delta E(\mathbf{x}[s]|\mathbf{u})$ を計算する.

$$\Delta E(\mathbf{x}[s]|\mathbf{u}) = -2 \sum_{j \in \partial i} J_{ij} x_i[s] x_j[s] - 2h_i x_i[s]$$

4. エネルギーが下がるときはスピン反転を受け入れる. そうでない場合は一様乱数 r が以下の不等式を満たすときに反転を受け入れる.

$$r \leq \exp\left(-\frac{2\Delta E(\mathbf{x}[s]|\mathbf{u})}{T}\right)$$

5. 終了基準を迎えるまで、ステップ 2-4 を繰り返す.

統計力学の大失態再び

あるものは使えばいいという思想が大事

$$P_0(\mathbf{x}) = \frac{1}{D} \sum_{s=1}^D \delta(\mathbf{x} - \mathbf{x}^{(s)})$$

あるものは使えばいいという思想が大事

$$P_0(\mathbf{x}) = \frac{1}{D} \sum_{s=1}^D \delta(\mathbf{x} - \mathbf{x}^{(s)})$$

データ駆動型



コントラスティヴ・ダイヴァージェンス法

M. Welling, and G. Hinton: *Artificial Neural Networks*, 2414, (2002) 351.

1. $\mathbf{x}[0]$ を初期化する. データの経験分布からサンプリングする. $\mathbf{x}[0] = \mathbf{x}^{(d)}$
2. スピンを反転するサイト i を選ぶ.
3. 反転に必要なエネルギー $\Delta E(\mathbf{x}[s]|\mathbf{u})$ を計算する.

$$\Delta E(\mathbf{x}[s]|\mathbf{u}) = -2 \sum_{j \in \partial i} J_{ij} x_i[s] x_j[s] - 2h_i x_i[s]$$

4. エネルギーが下がる時はスピン反転を受け入れる. そうでない場合は一様乱数 r が以下の不等式を満たすときに反転を受け入れる.

$$r \leq \exp\left(-\frac{2\Delta E(\mathbf{x}[s]|\mathbf{u})}{T}\right)$$

5. 終了基準を迎えるまで、ステップ 2-4 を繰り返す.



コントラストティブ・ダイバージェンス法

M. Welling, and G. Hinton: *Artificial Neural Networks*, 2414, (2002) 351.

- ▶ 脅威の性能とお手軽感！

$$\langle f(\mathbf{x}) \rangle_{\mathbf{u}} \approx \frac{1}{D} \sum_{d=1}^D f(\mathbf{x}^{(d)}[k])$$

- ▶ データの経験分布からの時系列で平均
- ▶ k (=1) ステップで良い！
- ▶ 比較：マルコフ連鎖モンテカルロ法

$$\langle f(\mathbf{x}) \rangle_{\mathbf{u}} \approx \frac{1}{T} \sum_{s=t_w+1}^{T+t_w} f(\mathbf{x}[s])$$





詳細釣り合いを破っても良い

- ▶ 非平衡統計力学の応用
 - ▶ 詳細釣り合いは十分条件

$$\frac{W_{\mathbf{u}}(\mathbf{y}|\mathbf{x})}{W_{\mathbf{u}}(\mathbf{x}|\mathbf{y})} = \frac{P_{\text{SS}}(\mathbf{y}|\mathbf{u})}{P_{\text{SS}}(\mathbf{x}|\mathbf{u})}$$

- ▶ 詳細釣り合いの破れた（非平衡系）での確率過程
 - ▶ Suwa-Todo法（遷移確率の最適化） **H. Suwa and S. Todo (2010)**
 - ▶ ひねり詳細釣り合い（ふたつの系での離散的マスター方程式） **K. S. Turitsyn, et al (2011)**
 - ▶ 固有値による解析 **A. Ichiki and M. Ohzeki (2013)**
 - ▶ 量子最速降下曲線による解析 **K. Takahashi and M. Ohzeki (2015?)**
 - ▶ 大関・一木法（ふたつの系でのランジュバン方程式） **M. Ohzeki and A. Ichiki (2015?)**





詳細釣り合いを破っても良い

- ▶ 非平衡統計力学の応用
 - ▶ 詳細釣り合いは十分条件

$$\frac{W_{\mathbf{u}}(\mathbf{y}|\mathbf{x})}{W_{\mathbf{u}}(\mathbf{x}|\mathbf{y})} = \frac{P_{\text{SS}}(\mathbf{y}|\mathbf{u})}{P_{\text{SS}}(\mathbf{x}|\mathbf{u})}$$

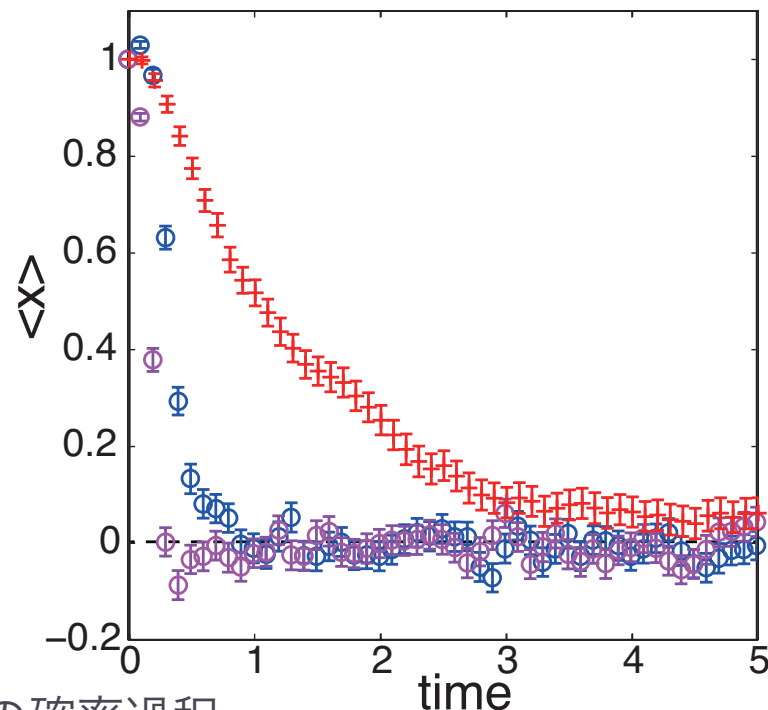
- ▶ 詳細釣り合いの破れた（非平衡系）での確率過程
 - ▶ Suwa-Todo法（遷移確率の最適化） **H. Suwa and S. Todo (2010)**
 - ▶ ひねり詳細釣り合い（ふたつの系での離散的マスター方程式） **K. S. Turitsyn, et al (2011)**
 - ▶ 固有値による解析 **A. Ichiki and M. Ohzeki (2013)**
 - ▶ 量子最速降下曲線による解析 **K. Takahashi and M. Ohzeki (2015?)**
 - ▶ 大関・一木法（ふたつの系でのランジュバン方程式） **M. Ohzeki and A. Ichiki (2015?)**



詳細釣り合いを破っても良い

- ▶ 非平衡統計力学の応用
 - ▶ 詳細釣り合いは十分条件

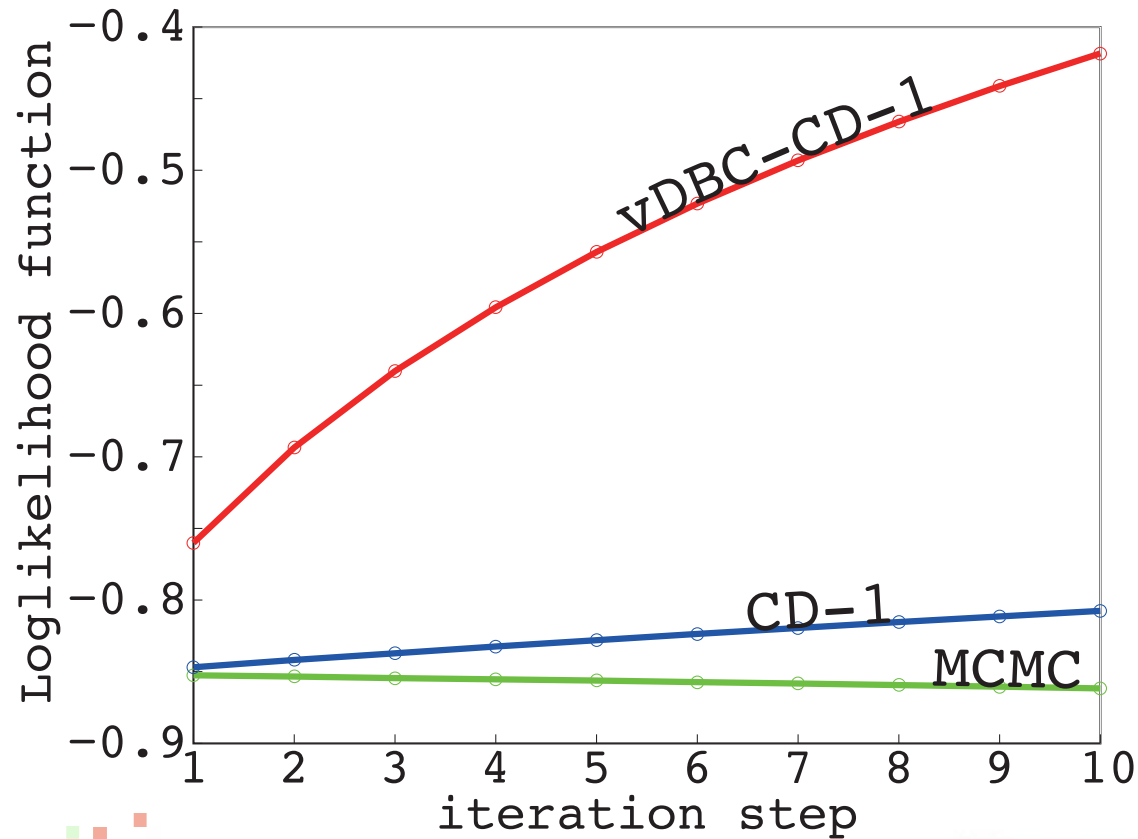
$$\frac{W_{\mathbf{u}}(\mathbf{y}|\mathbf{x})}{W_{\mathbf{u}}(\mathbf{x}|\mathbf{y})} = \frac{P_{SS}(\mathbf{y}|\mathbf{u})}{P_{SS}(\mathbf{x}|\mathbf{u})}$$



- ▶ 詳細釣り合いの破れた（非平衡系）での確率過程
 - ▶ Suwa-Todo法（遷移確率の最適化） **H. Suwa and S. Todo (2010)**
 - ▶ ひねり詳細釣り合い（ふたつの系での離散的マスター方程式） **K. S. Turitsyn, et al (2011)**
 - ▶ 固有値による解析 **A. Ichiki and M. Ohzeki (2013)**
 - ▶ 量子最速降下曲線による解析 **K. Takahashi and M. Ohzeki (2015?)**
 - ▶ 大関・一木法（ふたつの系でのランジュバン方程式） **M. Ohzeki and A. Ichiki (2015?)**



- ▶ コントラストティブ・ダイバージェンスの加速
 - ▶ 詳細釣り合いの破れたMCMCの採用による劇的な加速



統計力学 + データ駆動型

あるものは使えばいいという思想が大事

$$P_0(\mathbf{x}) = \frac{1}{D} \sum_{s=1}^D \delta(\mathbf{x} - \mathbf{x}^{(s)})$$



最小確率流法

J. Sohl-Dickstein et al: Phys. Rev. Lett. 107, (2011) 220601

- ▶ 熱平均の計算法として標準的
 - ▶ 連続時間マスター方程式

$$\frac{dP_t(\mathbf{y})}{dt} = \sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x}) P_t(\mathbf{x})$$





最小確率流法

J. Sohl-Dickstein et al: Phys. Rev. Lett. 107, (2011) 220601

- ▶ 熱平均の計算法として標準的
 - ▶ 連続時間マスター方程式

$$\frac{dP_t(\mathbf{y})}{dt} = \sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x}) P_t(\mathbf{x})$$

- ▶ CD法のように初期分布をデータの経験分布に設定

$$P_0(\mathbf{x}) = \frac{1}{D} \sum_{d=1}^D \delta(\mathbf{x} - \mathbf{x}^{(d)})$$





最小確率流法

J. Sohl-Dickstein et al: Phys. Rev. Lett. 107, (2011) 220601

- ▶ 熱平均の計算法として標準的
 - ▶ 連続時間マスター方程式

$$\frac{dP_t(\mathbf{y})}{dt} = \sum_{\mathbf{x}} W_{\mathbf{u}}(\mathbf{y}|\mathbf{x}) P_t(\mathbf{x})$$

- ▶ CD法のように初期分布をデータの経験分布に設定

$$P_0(\mathbf{x}) = \frac{1}{D} \sum_{d=1}^D \delta(\mathbf{x} - \mathbf{x}^{(d)})$$

- ▶ とりあえず解析しやすいように詳細釣り合いを課す

$$W_{\mathbf{u}}(\bar{\mathbf{x}}|\mathbf{x}) = \exp \left\{ -\frac{1}{2} (E(\bar{\mathbf{x}}|\mathbf{u}) - E_j(\mathbf{x}|\mathbf{u})) \right\}$$



統計力学

緩和は平衡状態に近ければ早い

統計力学

緩和は平衡状態に近ければ早い

データ駆動型

データの経験分布は何かの平衡状態



最小確率流法

J. Sohl-Dickstein et al: Phys. Rev. Lett. 107, (2011) 220601

- ▶ 正しいパラメータの分布とデータの経験分布は近いはず
 - ▶ データの経験分布から漏れる確率流を最小化

$$D_{\text{KL}}(P_0(\mathbf{x})|P_t(\mathbf{x})) \approx \frac{dt}{D} \sum_{s=1}^D \sum_{\mathbf{y} \in \partial \mathbf{x}^{(d)}} \exp \left\{ -\frac{1}{2} \left(E(\mathbf{y}|\mathbf{u}) - E(\mathbf{x}^{(d)}|\mathbf{u}) \right)^2 \right\}$$

- ▶ 対数尤度関数の代わりにのコスト関数を導入

$$L_{\text{MPF}}(\mathbf{u}) = \frac{1}{D} \sum_{d=1}^D \sum_{i=1}^N \exp \left\{ - \left(h_i + \sum_{j \in \partial i} J_{ij} x_j^{(d)} \right) x_i^{(d)} \right\}$$

- サンプルング不要！
- 疑似最尤法よりも高速な収束性
- 精度は疑似最尤法と同程度





ボルツマン機械学習 = データ駆動型統計力学

- ▶ 様々な手法が存在 + 計測技術の向上
 - ▶ 統計力学由来の既存手法
 - ▶ 平均場近似
 - ▶ 信念伝搬法
 - ▶ マルコフ連鎖モンテカルロ法
 - ▶ 統計的機械学習
 - ▶ 疑似最尤法 = 平均場近似 + データの利用
 - ▶ コントラストティブ・ダイバージェンス
 - = マルコフ連鎖モンテカルロ法 + データの利用
 - ▶ 最小確率流法
 - = 確率過程 + データの利用

QUANTUM
ANNEALING



MACHINE
LEARNING

Sparse Modeling



L1ノルムによる変数選択

M. Ohzeki: J. Phys. Soc. Jpn. 84, 054801 (2015)

- ▶ 基本は勾配法ベースの最尤法
 - ▶ コスト関数は様々に選べる
 - ▶ ボルツマン機械学習はデータが少ないと精度がでない
 - ビッグデータ時代だから大丈夫？
 - L1ノルムを使った正則化で解を選択

$$\min_{\mathbf{u}} \left\{ -L(\mathbf{u}) + \lambda \sum_{i < j} \|J_{ij}\|_1 \right\}$$

- ▶ メジャライザー最小化で実行可能！

$$f(\mathbf{u}) = \|\mathbf{u}\|_1 \quad g(\mathbf{u}) = -L(\mathbf{u})$$

$$\mathbf{u}[t + 1] = \arg \min_{\mathbf{u}} \left\{ f(\mathbf{u}) + (\nabla g(\mathbf{u}))^T (\mathbf{u} - \mathbf{u}[t]) + \frac{L}{2} \|\mathbf{u} - \mathbf{u}[t]\|_2^2 \right\}$$



L1ノルムによる変数選択

M. Ohzeki: J. Phys. Soc. Jpn. 84, 054801 (2015)

▶ LASSO型最小化問題のときは…

▶ リプシッツ定数が計算可能

$$L = \frac{\|A^T A\|_2}{\lambda}$$

▶ 非自明なときは適応的に計算

1. 初期化する. 適当な大きさのリプシッツ定数 $L[0]$ を用意する. 係数 $\alpha > 1$ を用意する.
2. 仮のリプシッツ定数 $L[t]$ で暫定的にメジャライザー最小解 $\mathbf{u}[t]$ を求める.

$$\mathbf{u}[t] = \arg \min_{\mathbf{u}} \{f(\mathbf{u}) + q_L(\mathbf{u}, \mathbf{u}[t-1])\}$$

3. 利用したリプシッツ定数によるメジャライザーが $g(\mathbf{u}[t])$ を上回ることを確認する.

$$g(\mathbf{u}[t]) \leq q_L(\mathbf{u}[t], \mathbf{u}[t-1])$$

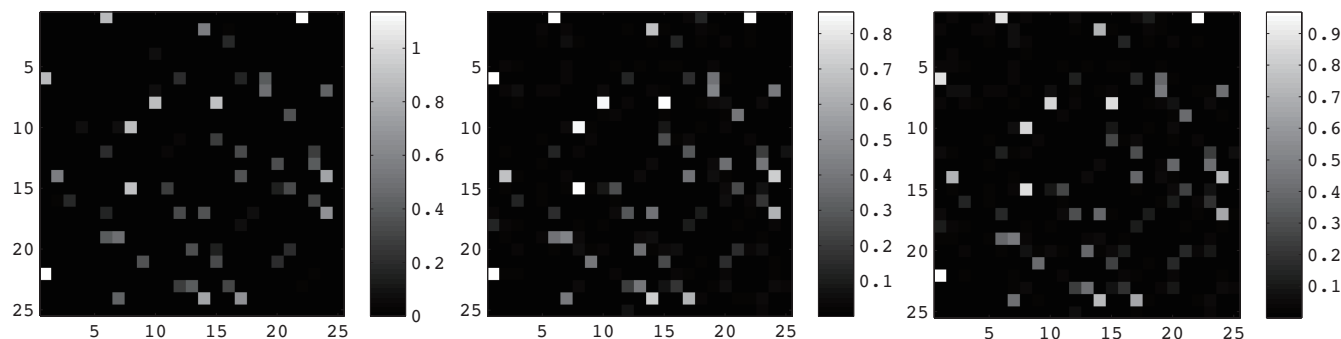
もし成立していなかったら、非負の整数値 i を増やして $L[t] = \alpha^i L[t-1]$ として更新してステップ2に戻る.



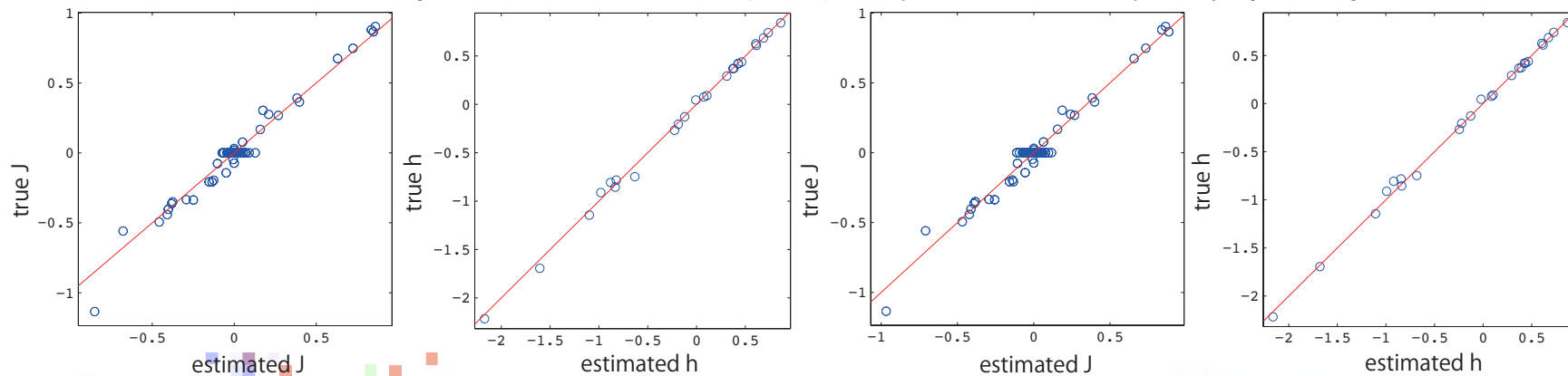
L1ノルムによる変数選択

M. Ohzeki: J. Phys. Soc. Jpn. 84, 054801 (2015)

- ▶ スパース相関推定：実験結果
 - ▶ $N=25$ 、 $D=5000$
 - ▶ 相互作用に1割程度の非零成分（ガウスランダム）



(左：オリジナル、中：疑似最尤法、右：最小確率流法)

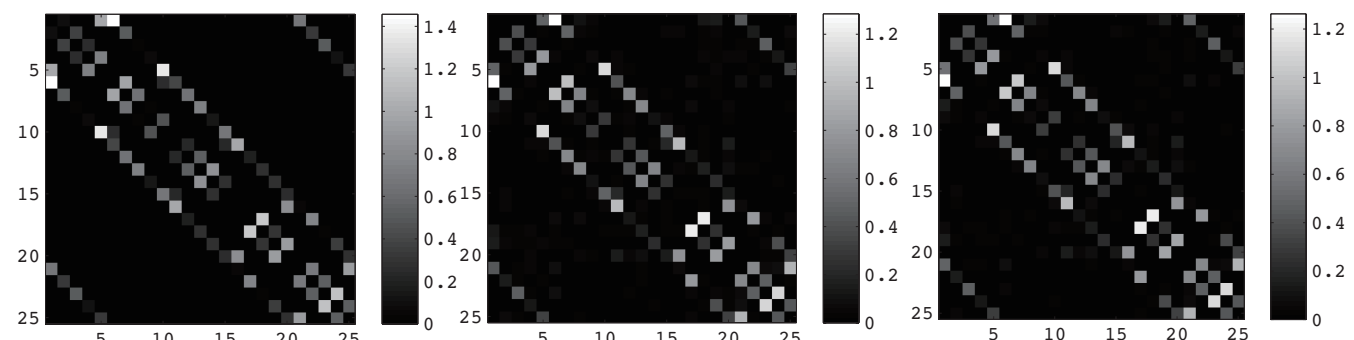




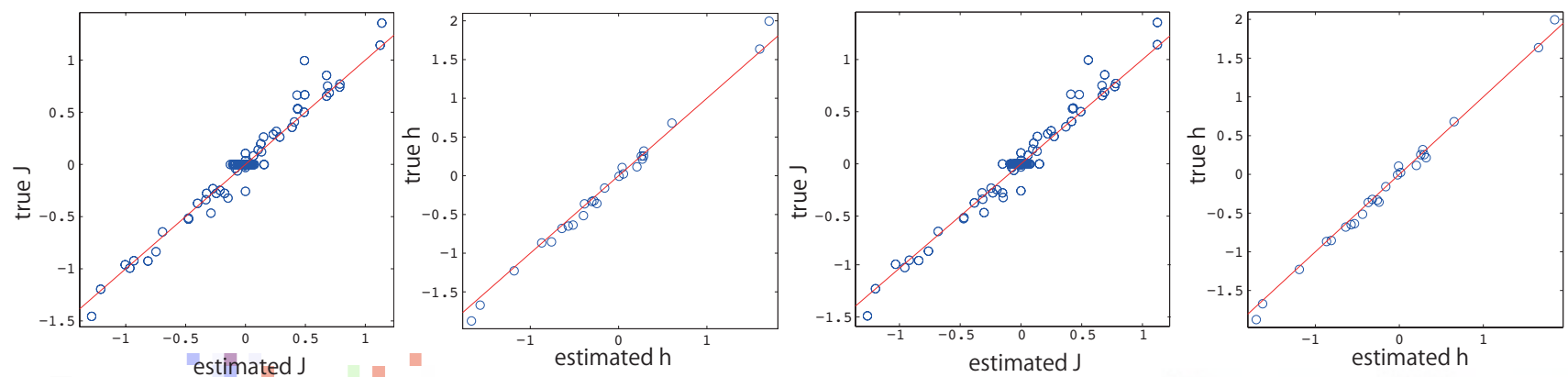
L1ノルムによる変数選択

M. Ohzeki: J. Phys. Soc. Jpn. 84, 054801 (2015)

- ▶ スパース相関推定：実験結果
 - ▶ N=25、D=5000
 - ▶ 相互作用を2次元最近接に限定、1割程度の非零成分（ガウスランダム）



(左：オリジナル、中：疑似最尤法、右：最小確率流法)



QUANTUM ANNEALING



MACHINE LEARNING

Sparse Modeling



まとめ

- ▶ 圧縮センシング
 - ▶ L1ノルムの解選択能力＋スパース解の精度良い推定能力
 - ▶ 観測数の削減、観測量不足のデータからの入力推定
 - ▶ 豊富なアルゴリズム
 - ▶ BP型最小化問題：内点法、Bregman反復法（拡張ラグランジュ法）
 - ▶ LASSO型最小化問題：NESTA、FISTA、ADMM
 - ▶ 情報統計力学的アプローチ
 - ▶ 積分する詐欺＝意外に計算は怖くない
 - ▶ 最適化問題の性能評価が可能
- ▶ ボルツマン機械学習
 - ▶ 統計力学＋データ駆動
 - ▶ 平均場近似＋データ＝疑似最尤法
 - ▶ マルコフ連鎖モンテカルロ法＋データ＝CD法、最小確率流法
 - ▶ L1ノルムの解選択能力を付加して、本質・構造抽出

