人工知能応用とデータ駆動科学

東京大学・大学院新領域創成科学研究科
岡田 真人

内容

1. 自己紹介
2. 人工知能を明日から利用するには?
3. 顔応答細胞のポピュレーションダイナミクス
4. 新学術領域研究「スパースモデリングの深化と高次元データ駆動科学の創成」
5. ベイズ計測
6. SPring-8全ビームラインベイズ化計画
7. まとめ

自己紹介 (理論物理学者)

- 大阪市立大学理学部物理学科 (1981 - 1985)
  - アモルファスシリンコンの成長と構造解析
- 大阪大学大学院理学研究科(金森研) (1985 – 1987)
  - 希土類元素の光励起スペクトルの理論
- 三菱電機 (1987 - 1989)
  - 化合物半導体(半導体レーザー)の結晶成長
- 大阪大学大学院基礎工学研究科生物工学(福島研) (1989 - 1996)
  - ニューラルネットワーク(DCNN)
- JST ERATO 川人学習動態脳プロジェクト (1996 - 2001)
  - 計算論的神経科学
- 理化学研究所 脳科学総合研究センター 甘利チーム (2001 - 04/06)
  - 情報統計力学
  - ベイズ推論, 機械学習, データ駆動型科学
- 東京大学・大学院新領域創成科学研究科 複雑理工学専攻 (2004/07 –)
- NIMS 情報統合型 物質・材料開発イニシアティブ (2015 –2020)
- NIMS 統合型材料開発・情報基盤部門(MaDIS) (2017/4 –

内容

1. 自己紹介
2. 人工知能を明日から利用するには?
3. 顔応答細胞のポピュレーションダイナミクス
4. 新学術領域研究「スパースモデリングの深化と高次元データ駆動科学の創成」
5. ベイズ計測
6. SPring-8全ビームラインベイズ化計画
7. まとめ

人工知能を明日から利用するには?

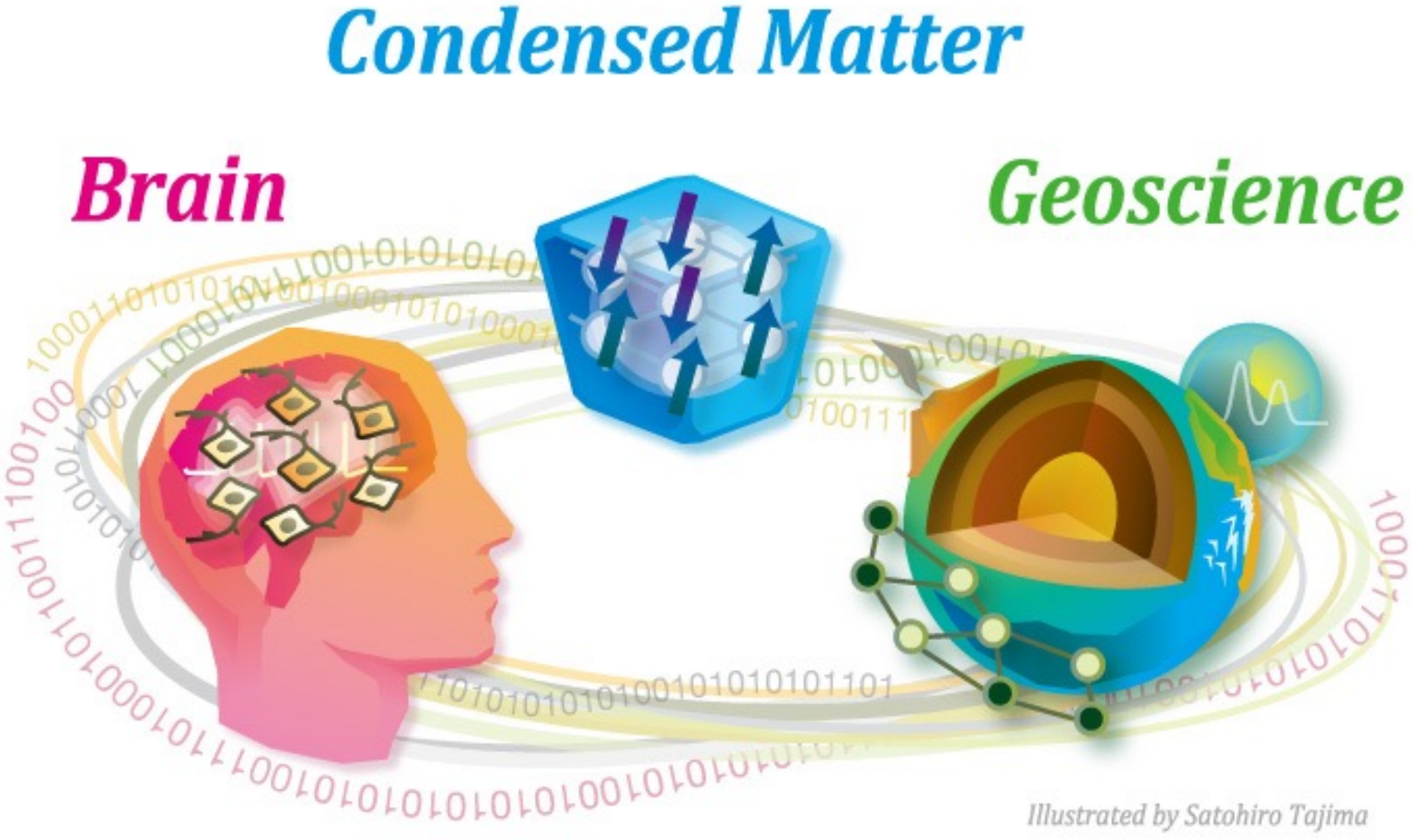
- 人工知能の応用を自分の問題として捉えて実行しようとした場合，具体的に何をやればよいか困ってしまうことが多いと思います．

- 人工知能の勉強をしてから利用するという考えは捨てる
- まず自分の研究をモダンな形で実行する枠組みはないかと考える
- データ駆動科学
- スパースモデリングとベイズ推論

データ駆動科学とは

- 機械学習などの人工知能を使い，各学問分野の問題を解いていくというアプローチ
- 実験/計測/計算データの背後にある潜在的構造の抽出に関して，データが対象とする学問に依存しない普遍的な学問体系
- 同じアルゴリズムがスケールや対象を超えて，有用であることが多いという経験的事実を背景として，その理由を問い，背後にある普遍性から，データ解析自体を学問的対象とする枠組み.

データ駆動科学


Condensed Matter
Brain
Geoscience
Illustrated by Satohiro Tajima

内容

1. 自己紹介
2. 人工知能を明日から利用するには?
3. 顔応答細胞のポピュレーションダイナミクス
4. 新学術領域研究「スパースモデリングの深化と高次元データ駆動科学の創成」
5. ベイズ計測
6. SPring-8全ビームラインベイズ化計画
7. まとめ

共同研究者

松本有央（産業総合研究所）

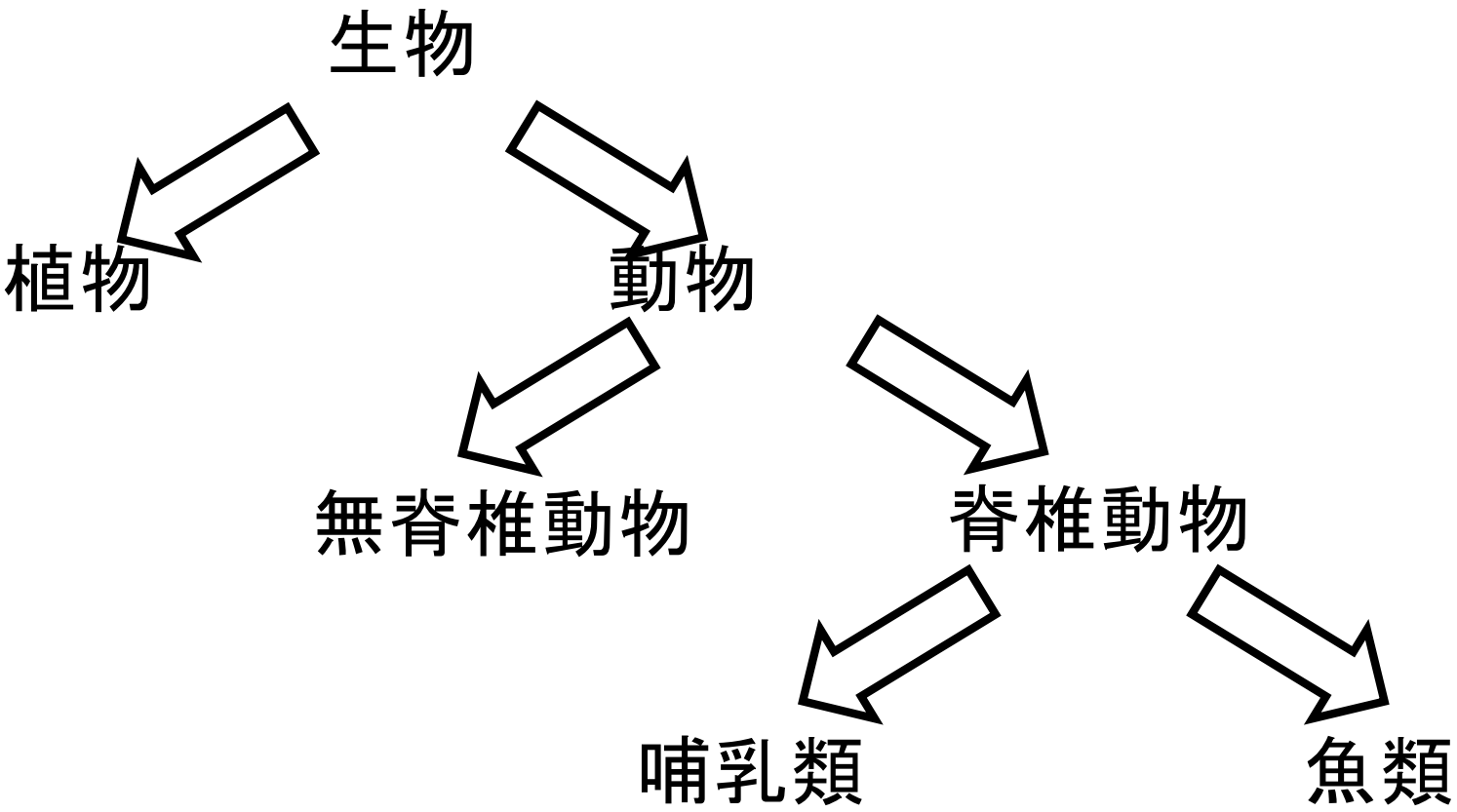
菅生康子（産業総合研究所）

山根茂（前橋工科大学）

河野 憲二（京都大学）

動機（1/2）

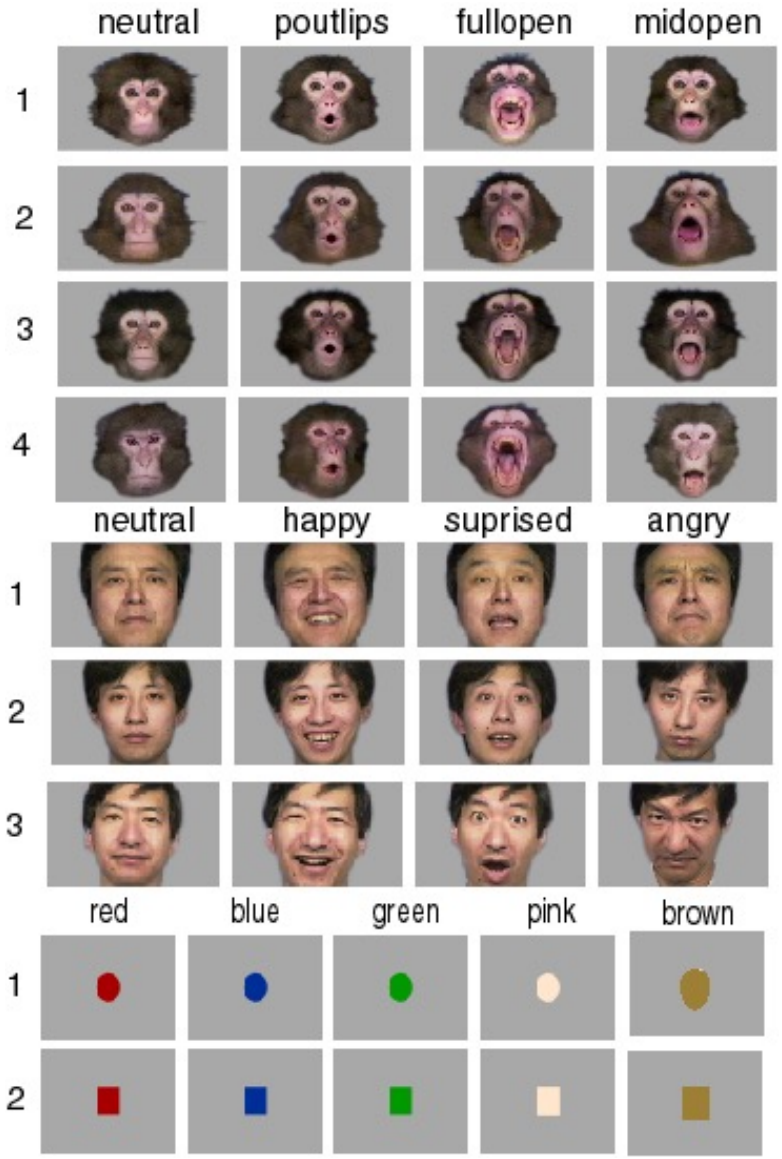
- 我々は世界を階層的に捉える傾向がある

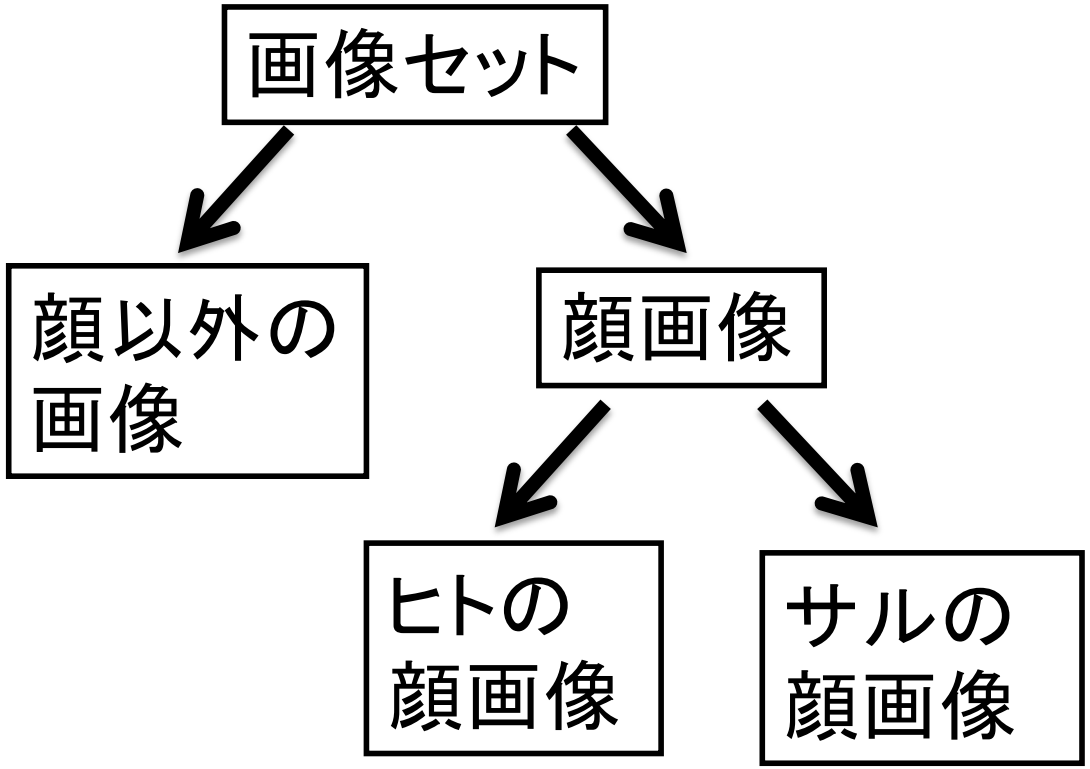


動機 (2/2)

- 我々は世界を階層的に捉える傾向がある
- なぜ？
  - 世界がそもそも階層的であるから.
  - 記憶容量の観点から，我々の脳が世界を階層的にとらえているだけ.
- 脳で階層的な刺激がどのように表現されているか？
- 階層的な関係性を表現する神経メカニズムを探る.

階層的な画像セット

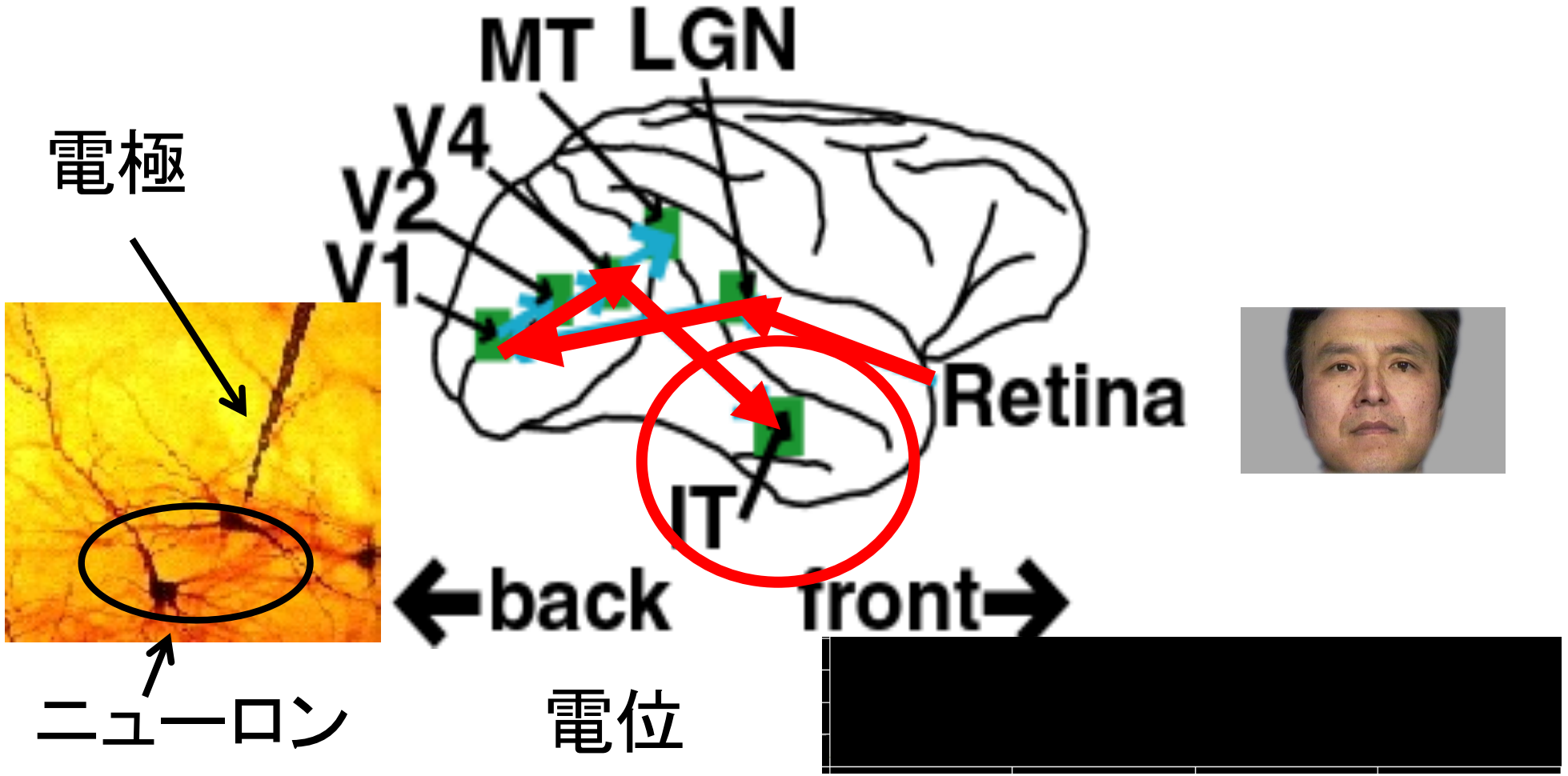




個体別
表情

(Sugase, Yaname, Ueno and Kawano, *Nature*, 1999)

電気生理学実験

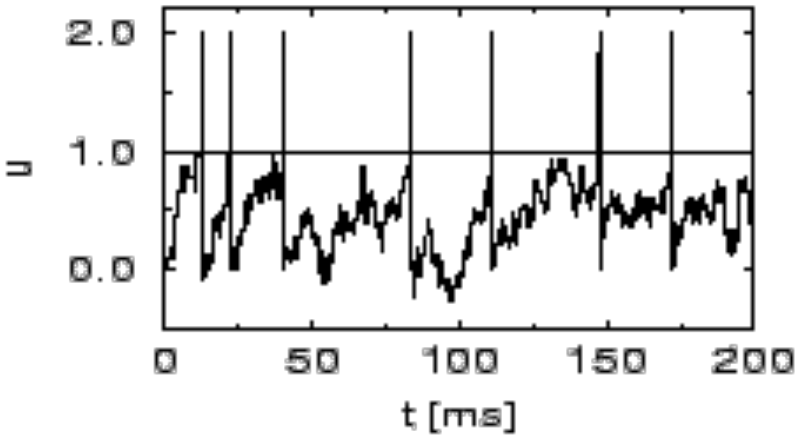


IT: Inferior Temporal Cortex (側頭葉)

側頭葉はパターン認識の責任領野

スパイクから発火率へ

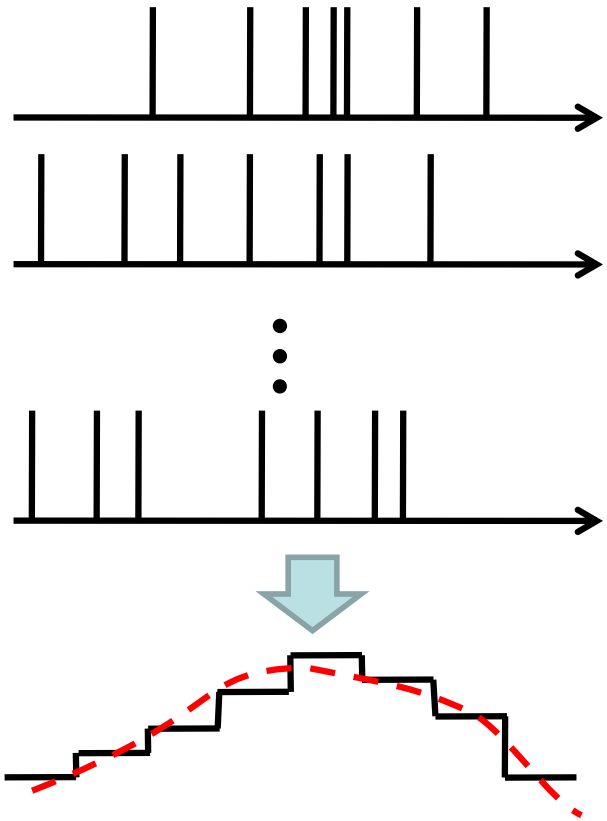
スパイク



発火状態:　1

非発火状態:　0

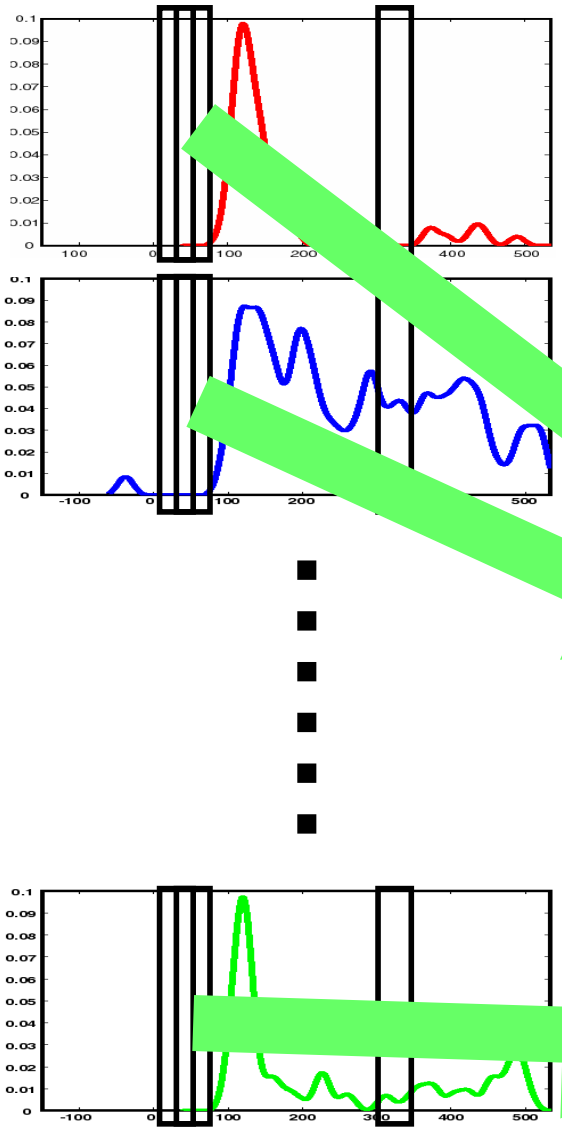
ランダムスパイク系列



発火率

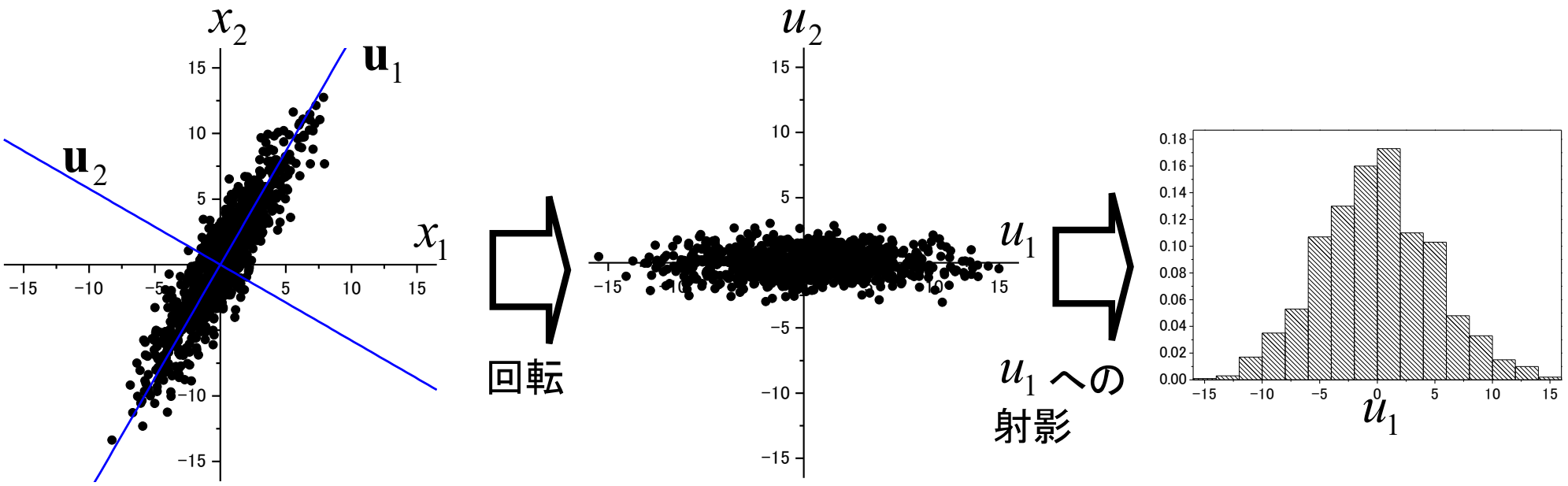
神経集団ベクトル

- 50 msecの時間窓
- 一つの刺激:
  45次元ベクトル
- 38個のベクトル

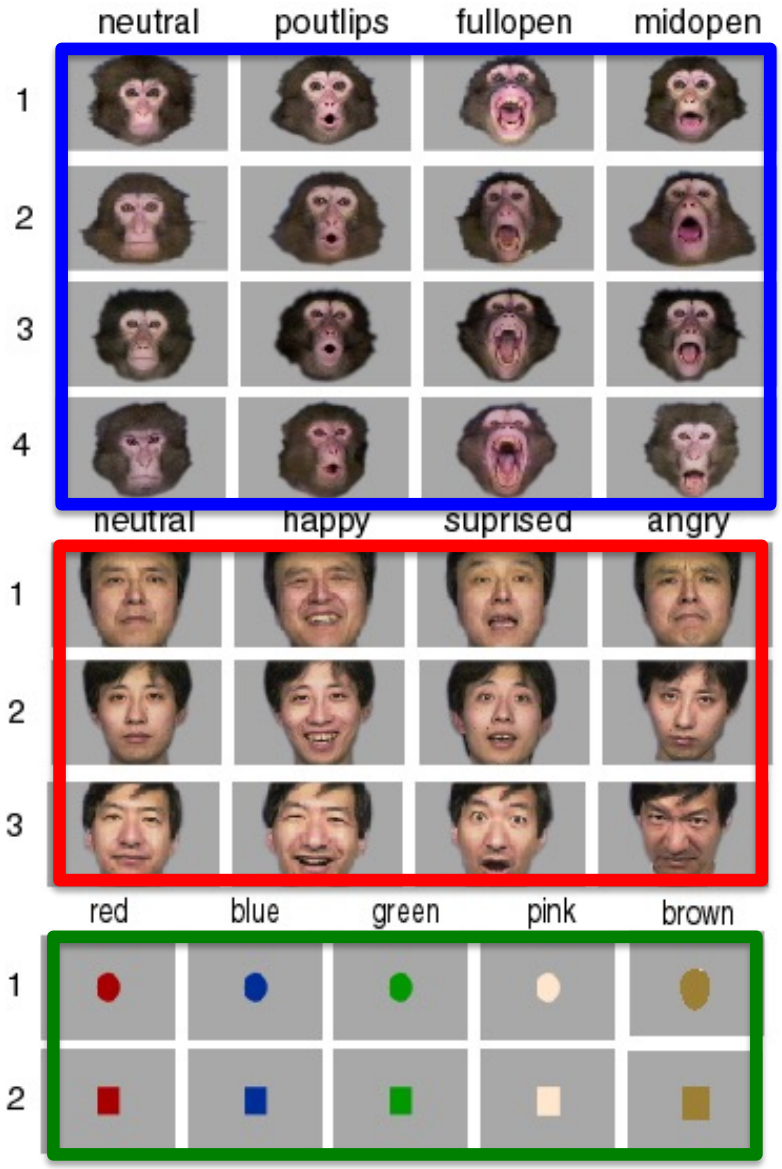


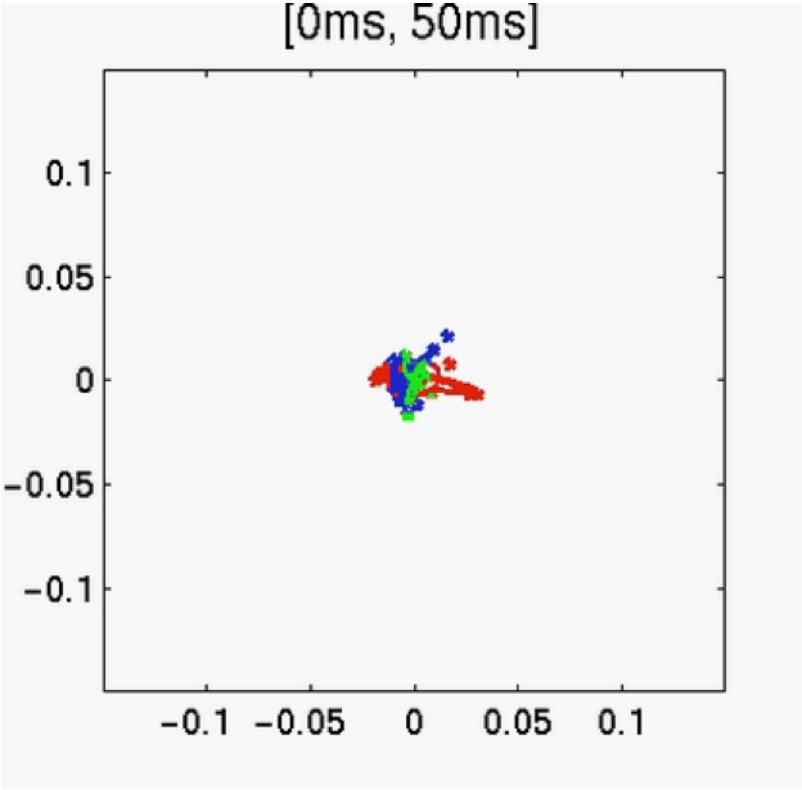
主成分分析
(Principal Component Analysis, PCA))

方法：射影先での分散が最大になる低次元の空間を探す



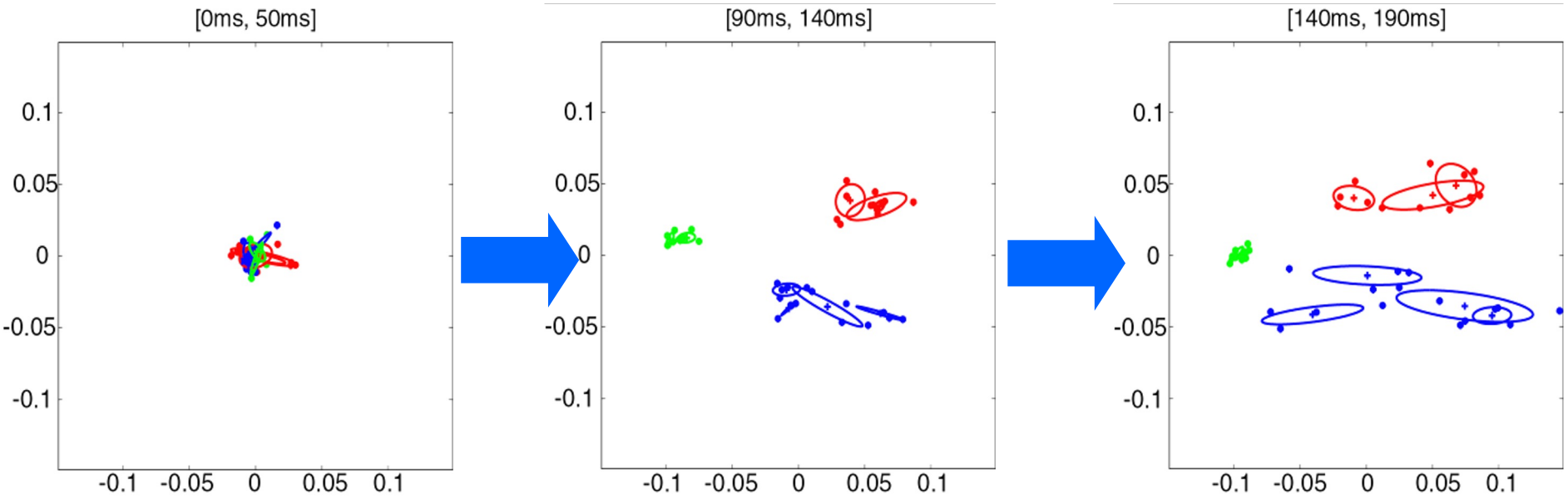
側頭葉の神経ダイナミクス





(Sugase, Yaname, Ueno and Kawano, 1999)
(Matsumoto, Okada, Sugase-Miyamoto, Yamane and Kawano, 2005)

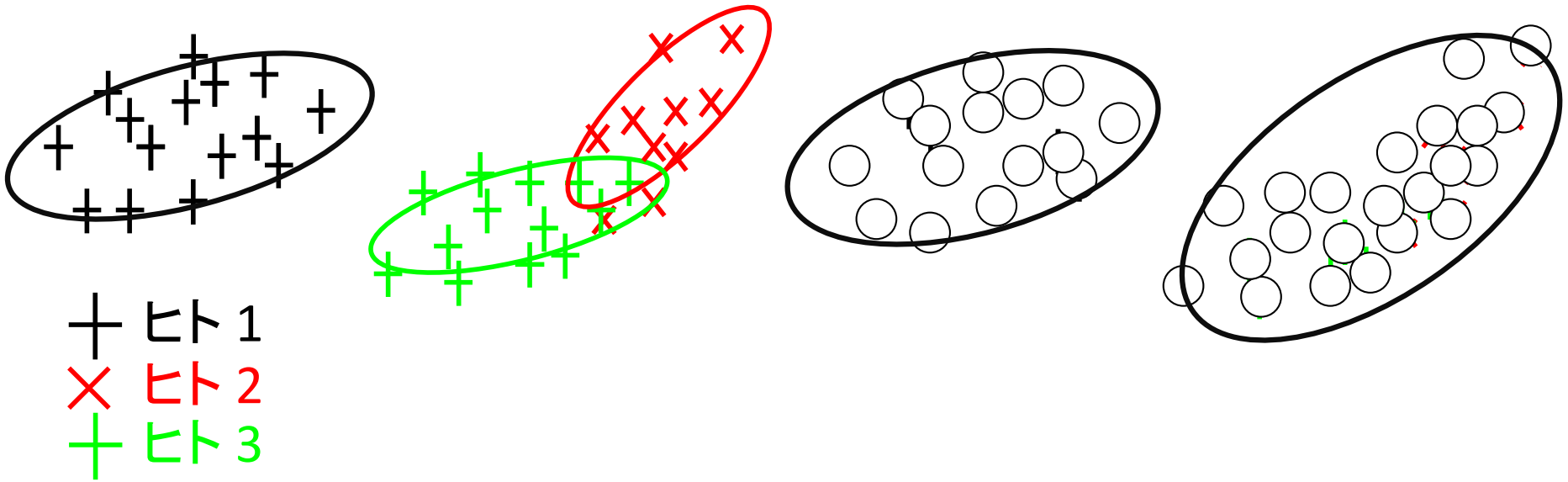
神経集団ダイナミクス



- [90ms, 140ms]でグローバルな分類が起こる
  （サル対ヒト対図形）
- [140ms, 190ms]でファインが分類が起こる
  （サル表情, ヒトの個体識別, 図形の形）

視覚刺激の中の階層的な関係性が神経集団のダイナミクスに情報表現されている

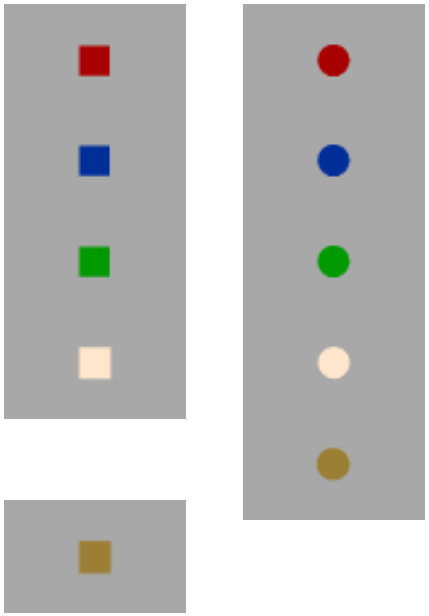
ベイズ的クラスタリング

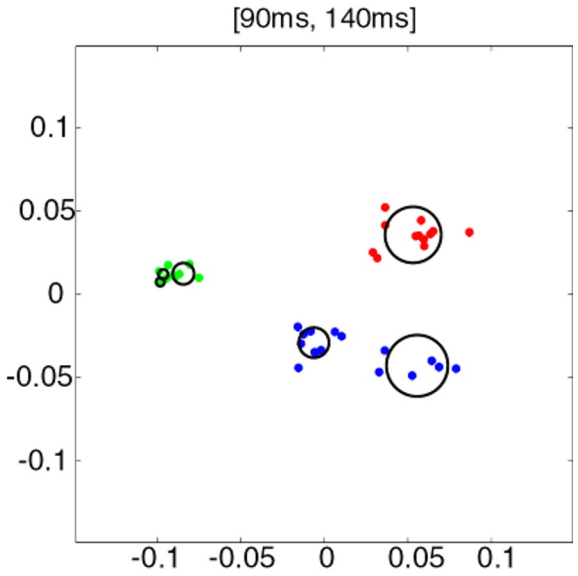


教師ありクラスタリング

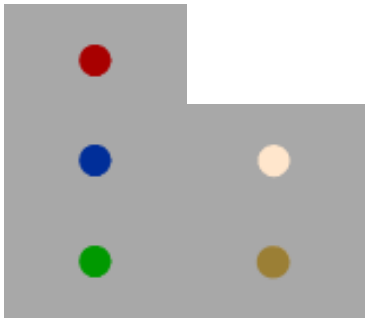
教師なしクラスタリング
混合正規分布解析
クラスターの数の自動決定

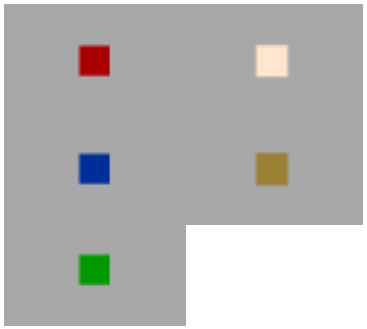
[90, 140ms]

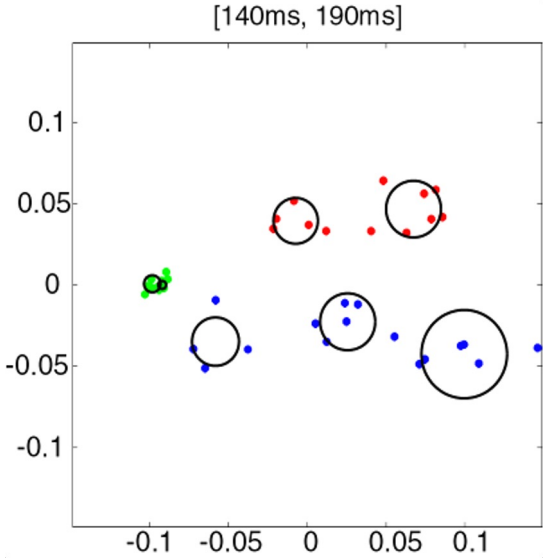


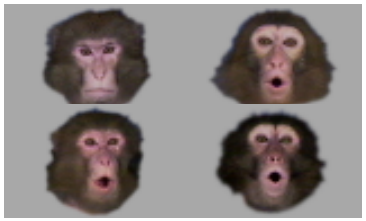


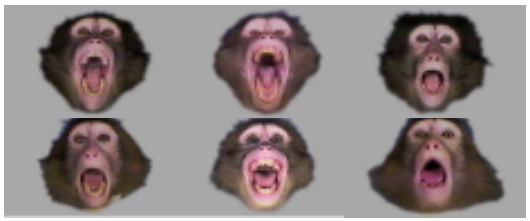
[140, 190ms]











共同研究者

松本有央（産業総合研究所）

菅生康子（産業総合研究所）

山根茂（前橋工科大学）

河野 憲二（京都大学）

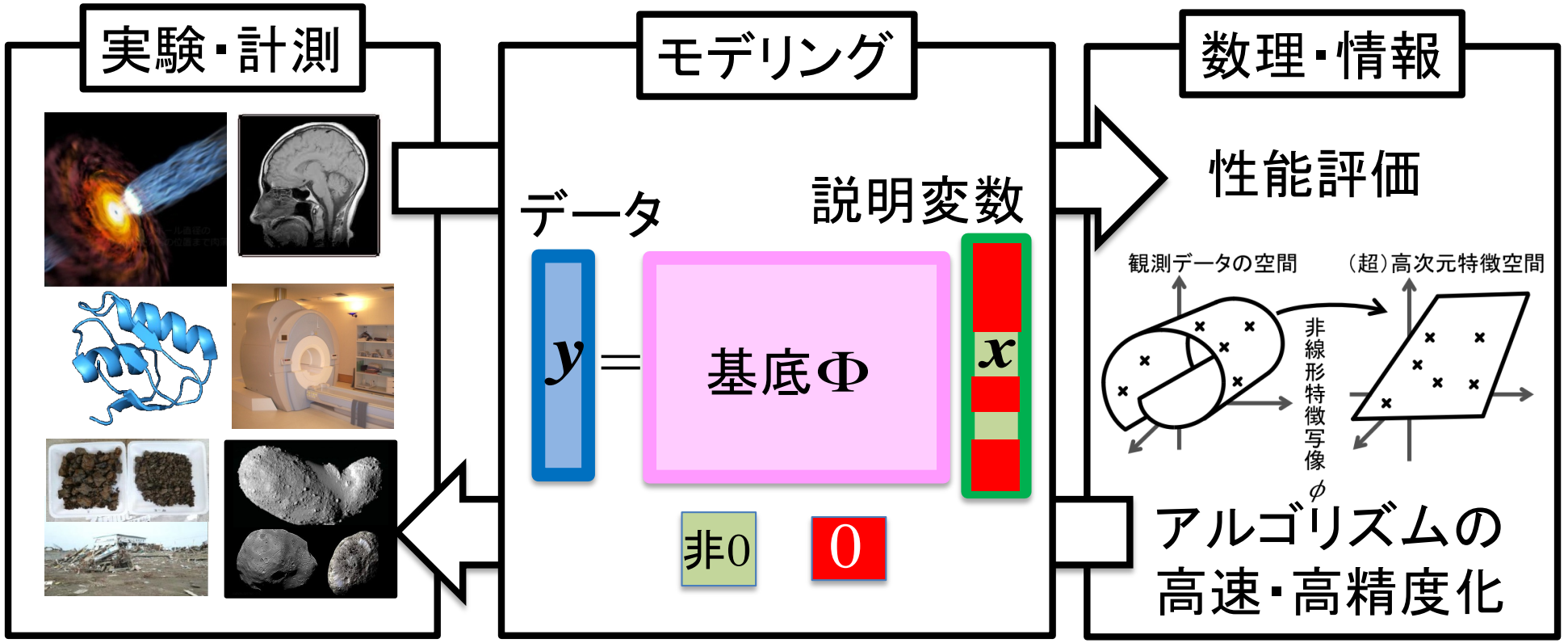
内容

1. 自己紹介
2. 人工知能を明日から利用するには?
3. 顔応答細胞のポピュレーションダイナミクス
4. 新学術領域研究「スパースモデリングの深化と高次元データ駆動科学の創成」
5. ベイズ計測
6. SPring-8全ビームラインベイズ化計画
7. まとめ

新学術領域研究　平成25～29年度
スパースモデリングの深化と高次元データ駆動科学の創成

個人的な狙い

世界を系統的に記述したい
その方法論と枠組みを創りたい
ヒトが世界を認識するとは?



領域の目的と戦略

目的：高次元データ駆動科学の創成

大量の高次元データから 仮説（モデル）を系統的に導く方法論を「生物」，「地学」分野に確立し，それを実践するための研究体制のコアを我が国に形成する．

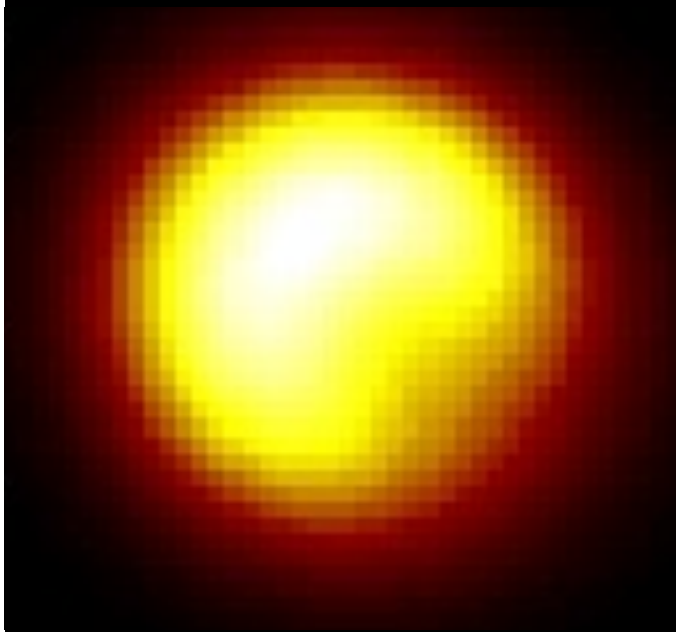
3つの戦略

1. スパースモデリングに重点投資
   今後5年で飛躍的発展が確実視される枠組み
2. 分野の壁を取り去り，知識伝播を飛躍的に加速
   分野をまたぐモデルの構造的類似性を明確化
3. 実験家と理論家との有機的協働
   仮説の提案／検証ループを効率的に稼働させる体制

SpMによるブラックホールの直接撮像
本間@国立天文台

計測シミュレーション結果

一般相対性理論による予測

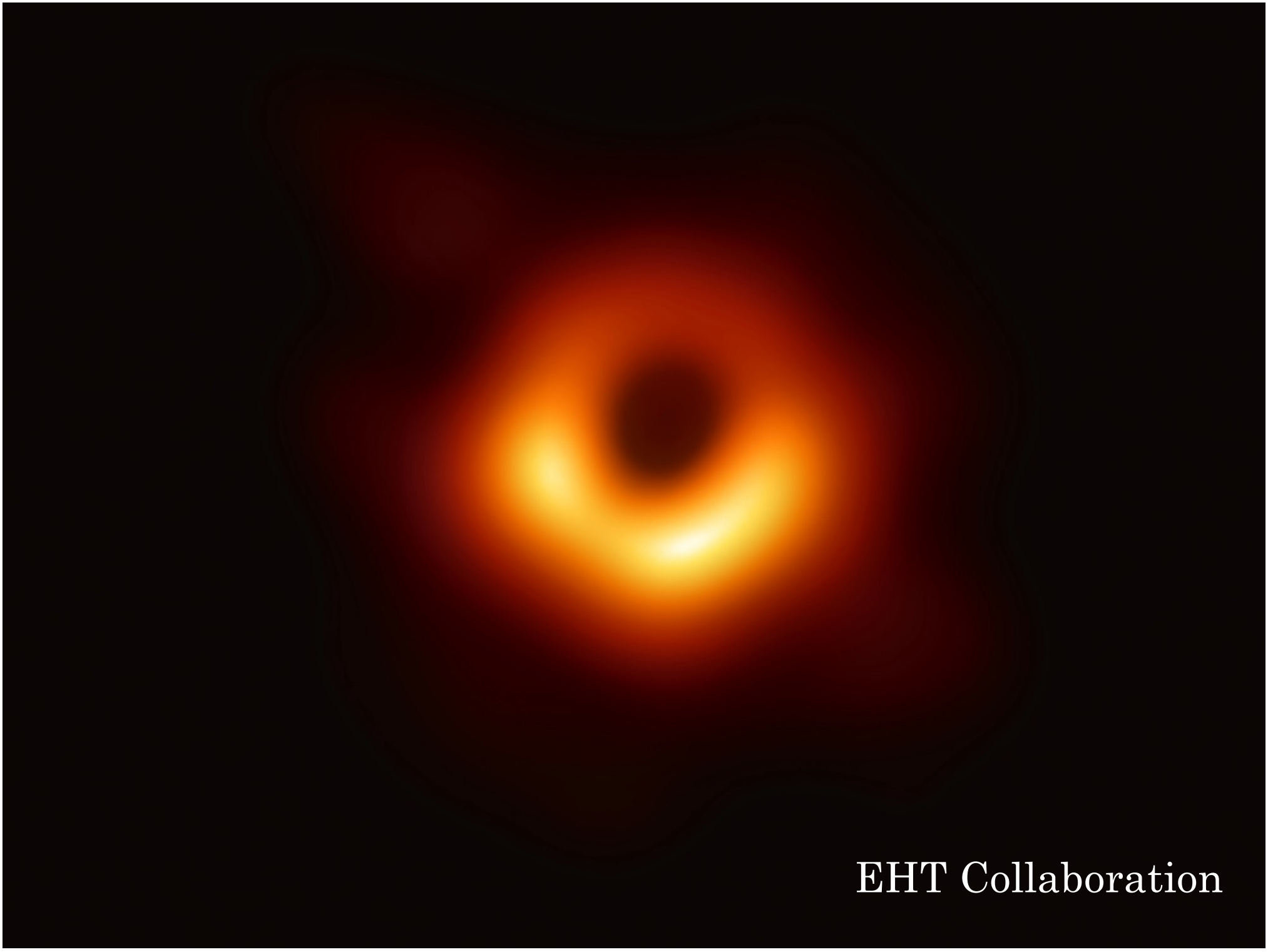


従来法

スパースモデリング

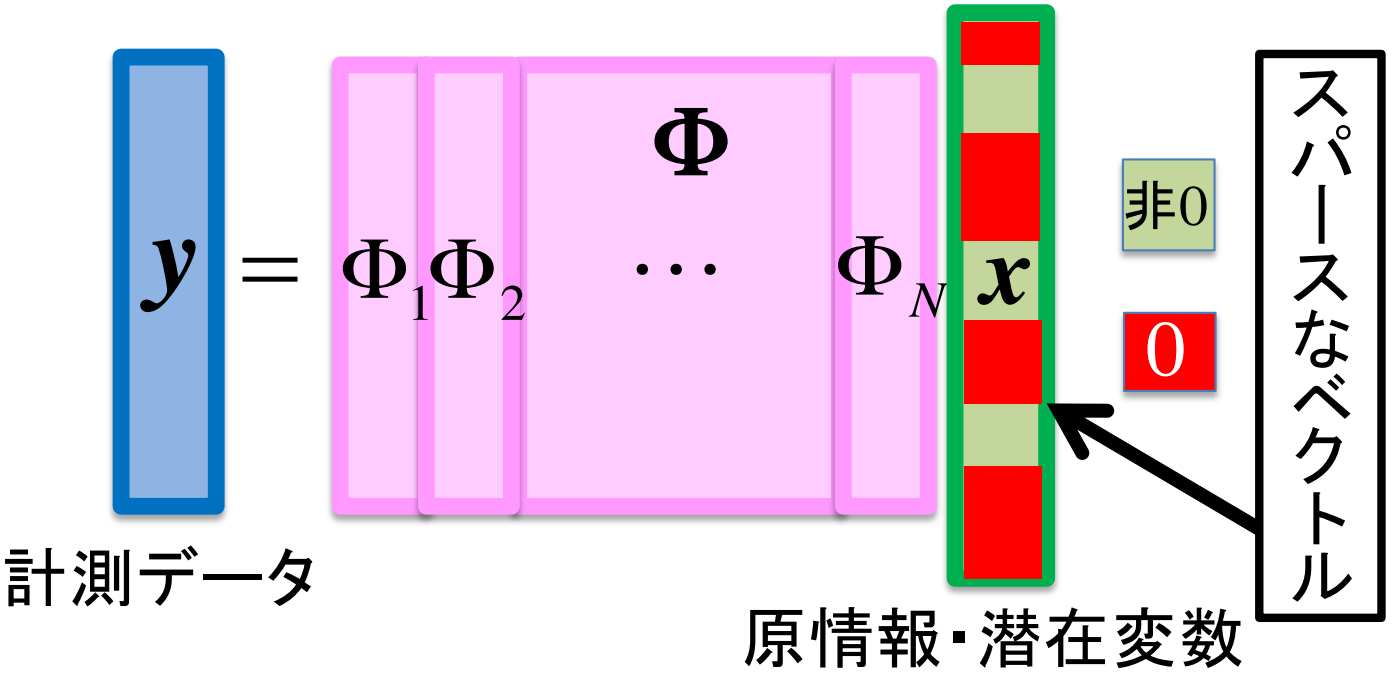
苫小牧高専
高橋准教授提供

光子軌道の回転速度により高速度極限($v/c \sim 1$)での一般相対性理論の検証


EHT Collaboration

スパースモデリング
スパース性でサンプリング定理を超える

線形計測：ブラックホール，MRI，NMR・・・



変数の個数が，式の数より多い　⇒　解が求まらない

スパースなベクトル

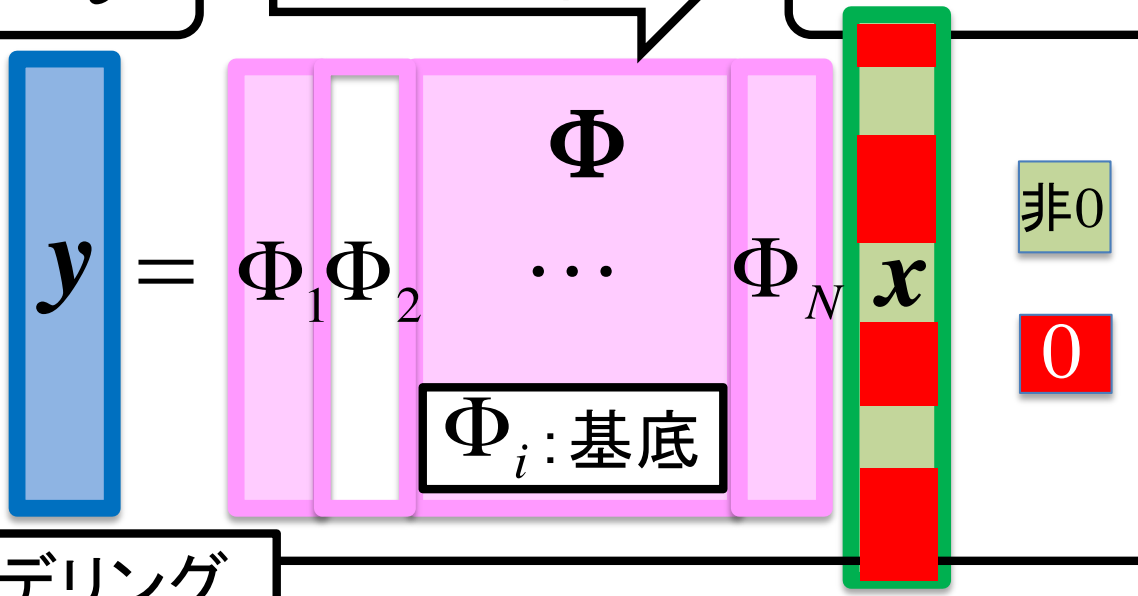
変数の要素に0が多いベクトル．

スパースモデリング
スパース性でサンプリング定理を超える

計測データ $\boldsymbol{y}$

スパース化

原情報・潜在変数 $\boldsymbol{x}$



スパースモデリング

潜在変数がスパース(0が多い)

0の場所を推定しながら，方程式を解く

付録スライド7参照

$$E(\boldsymbol{x}) = \left\| \boldsymbol{y} - \sum_i \Phi_i x_i \right\|^2 + \lambda \sum_i |x_i|$$

Φ_i：基底

データの再構成　スパースな変数

圧縮センシング

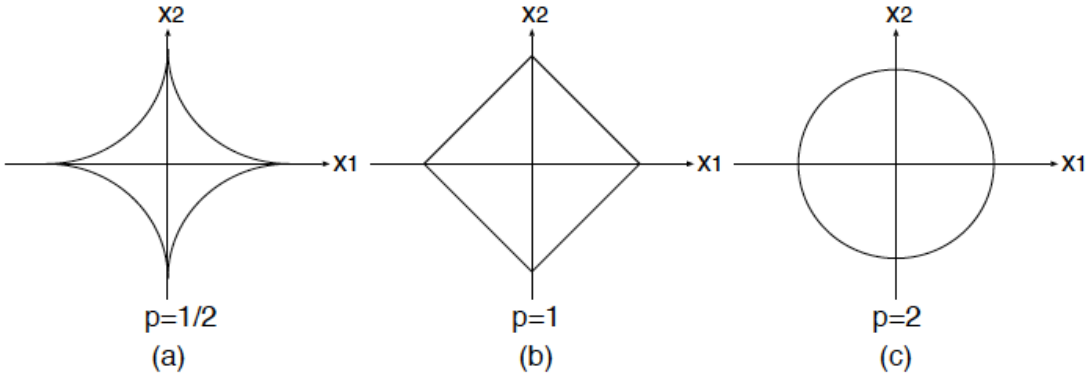


図 3: 単位 ℓ_p 球．$N = 2$, $|x_1|^p + |x_2|^p = 1$（単位 $\ell_{1/2}$ 球は慨形）．

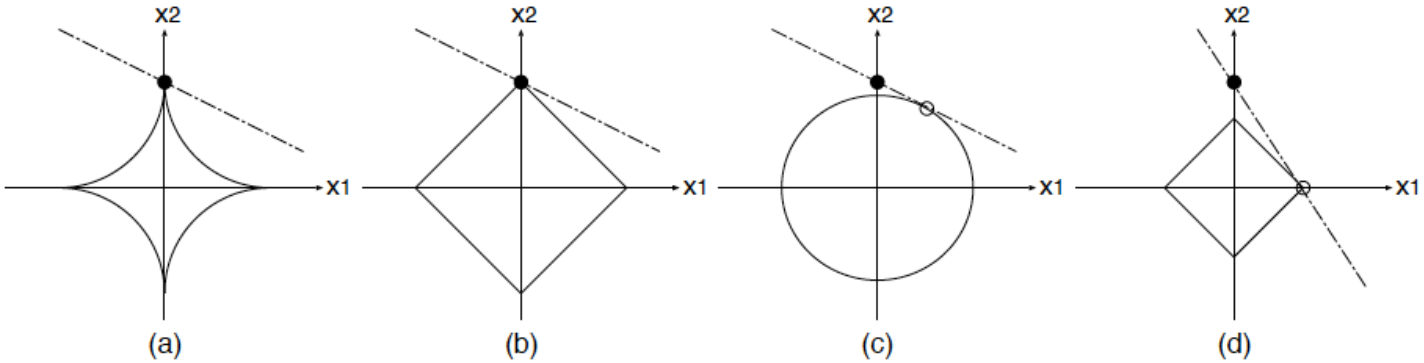


図 4: ℓ_p 再構成．$K = 1$, $M = 1$, $N = 2$ の場合．

(三村和史，2012)

領域の目的と戦略

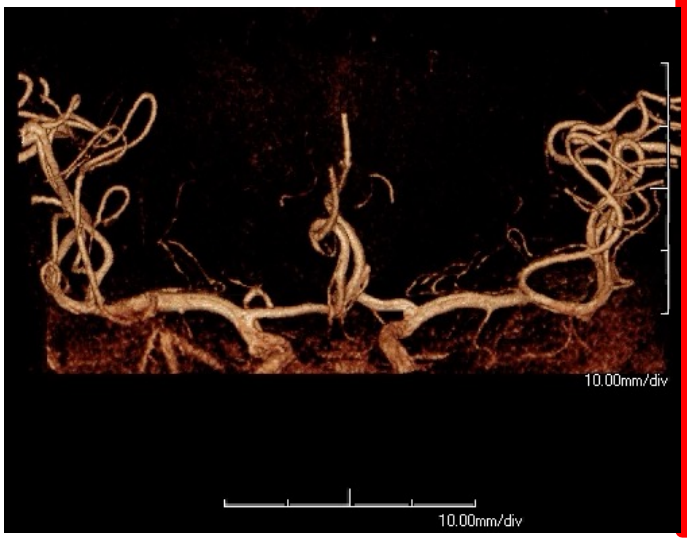
目的：高次元データ駆動科学の創成

大量の高次元データから 仮説（モデル）を系統的に導く方法論を「生物」，「地学」分野に確立し，それを実践するための研究体制のコアを我が国に形成する．

3つの戦略

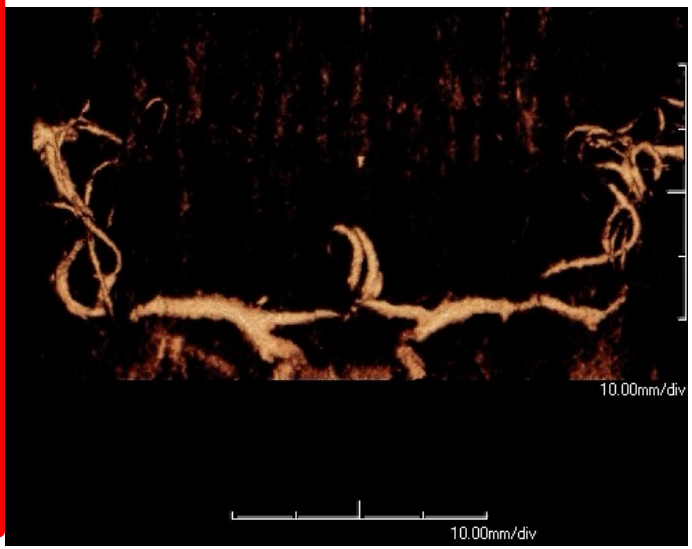
1. スパースモデリングに重点投資
   今後5年で飛躍的発展が確実視される枠組み
2. 分野の壁を取り去り，知識伝播を飛躍的に加速
   分野をまたぐモデルの構造的類似性を明確化
3. 実験家と理論家との有機的協働
   仮説の提案／検証ループを効率的に稼働させる体制

スパースモデリングによるMRIの高速撮像
富樫@京大医学部

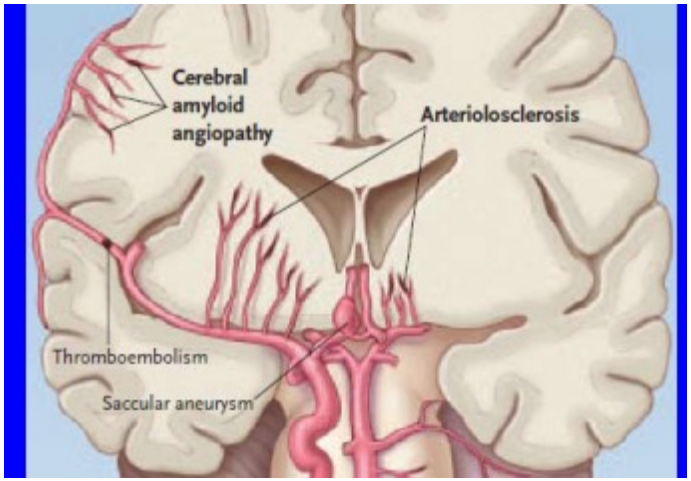


元画像

スパースモデリング



従来法



データ獲得時間を5分の1
⇒予防医療に革新，患者負担の軽減

MRIが実用化されて30年
装置の改良でデータ獲得時間5分の1

内容

1. 自己紹介
2. 人工知能を明日から利用するには?
3. 顔応答細胞のポピュレーションダイナミクス
4. 新学術領域研究「スパースモデリングの深化と高次元データ駆動科学の創成」
5. ベイズ計測とスペクトル分解
6. SPring-8全ビームラインベイズ化計画
7. まとめ

共同研究者

永田賢二
NIMS

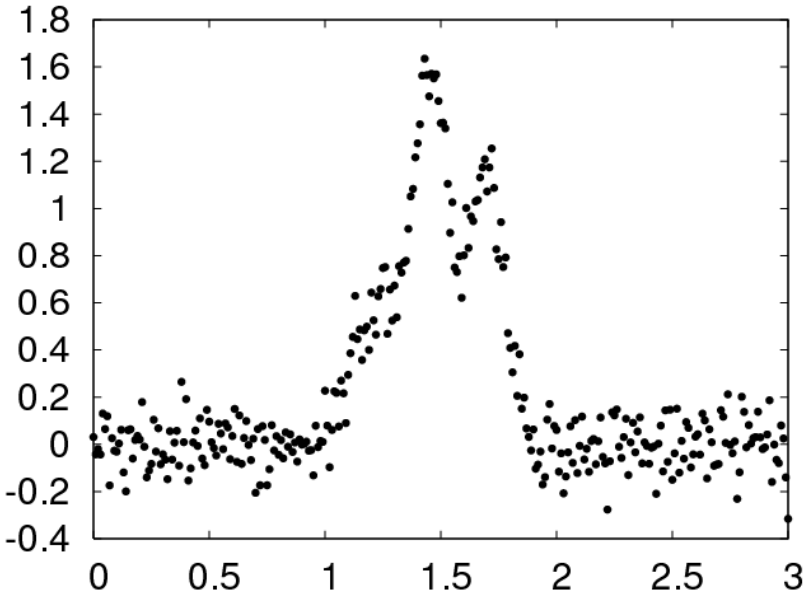
杉田精司
東京大学

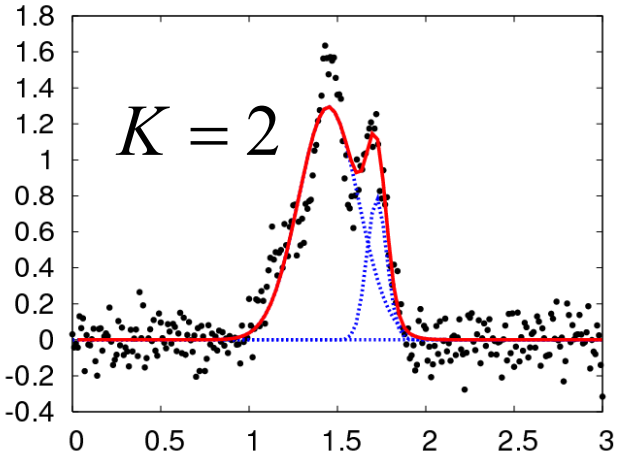
アンケート

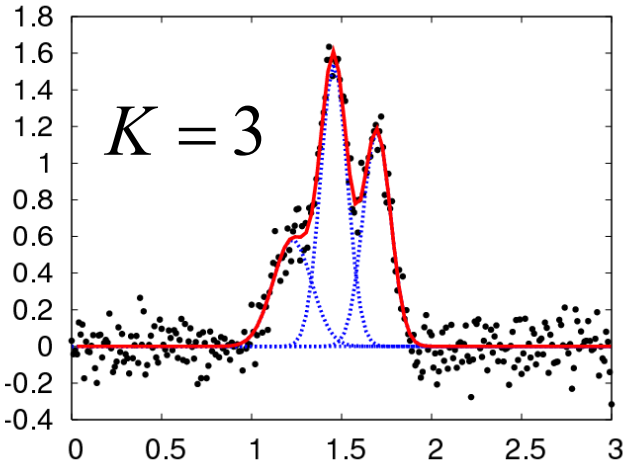
- スペクトルや画像データからフィッティングを行なっている
- そのフィッティングの際に、パラメータを手打ちで決めている。最急降下法などを使っているが、うまくいかない。
- フィッティング用のモデルが複数あって、事前にどれを使うかを決めておかないといけない。
- S/Nが悪いデータや欠損データをなんとかした。
- 複数計測の統合を行いたい。

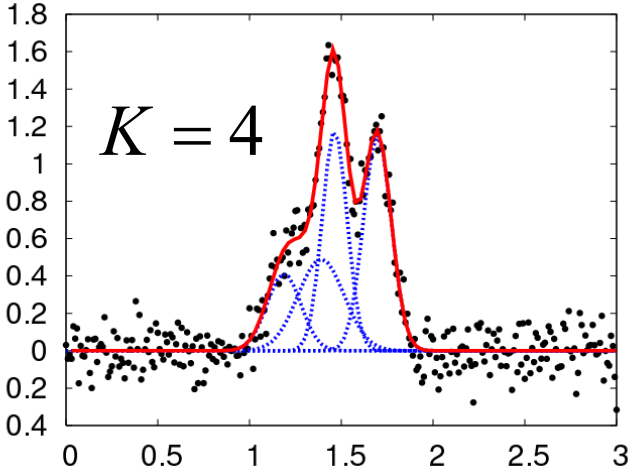
- そのような方は、一度ベイズ計測をお試しください。

スペクトル分解









Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

スペクトル分解の定式化

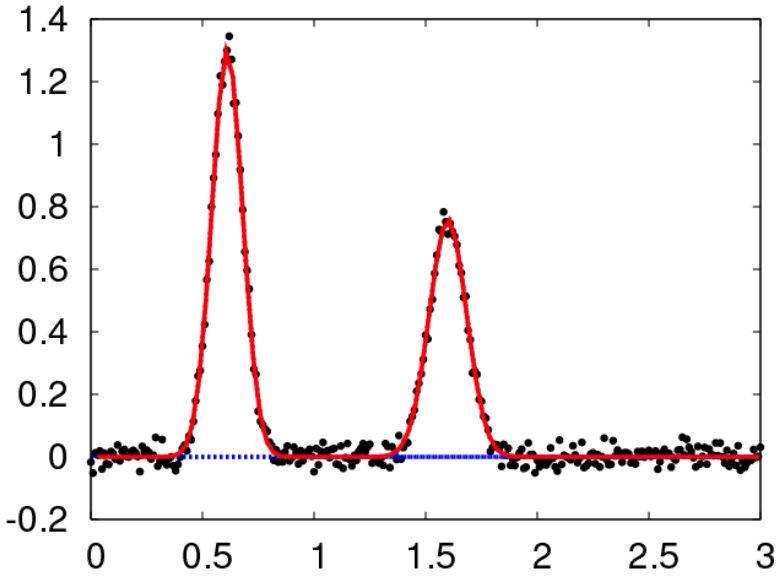
ガウス関数(基底関数)の足し合わせにより，スペクトルデータを近似

観測データ：$D = \{x_i, y_i\}_{i=1}^{n}$

x_i：入力　y_i：出力

$$f(x;\theta) = \sum_{k=1}^{K} a_k \exp\left(-\frac{b_k\left(x-\mu_k\right)^2}{2}\right)$$

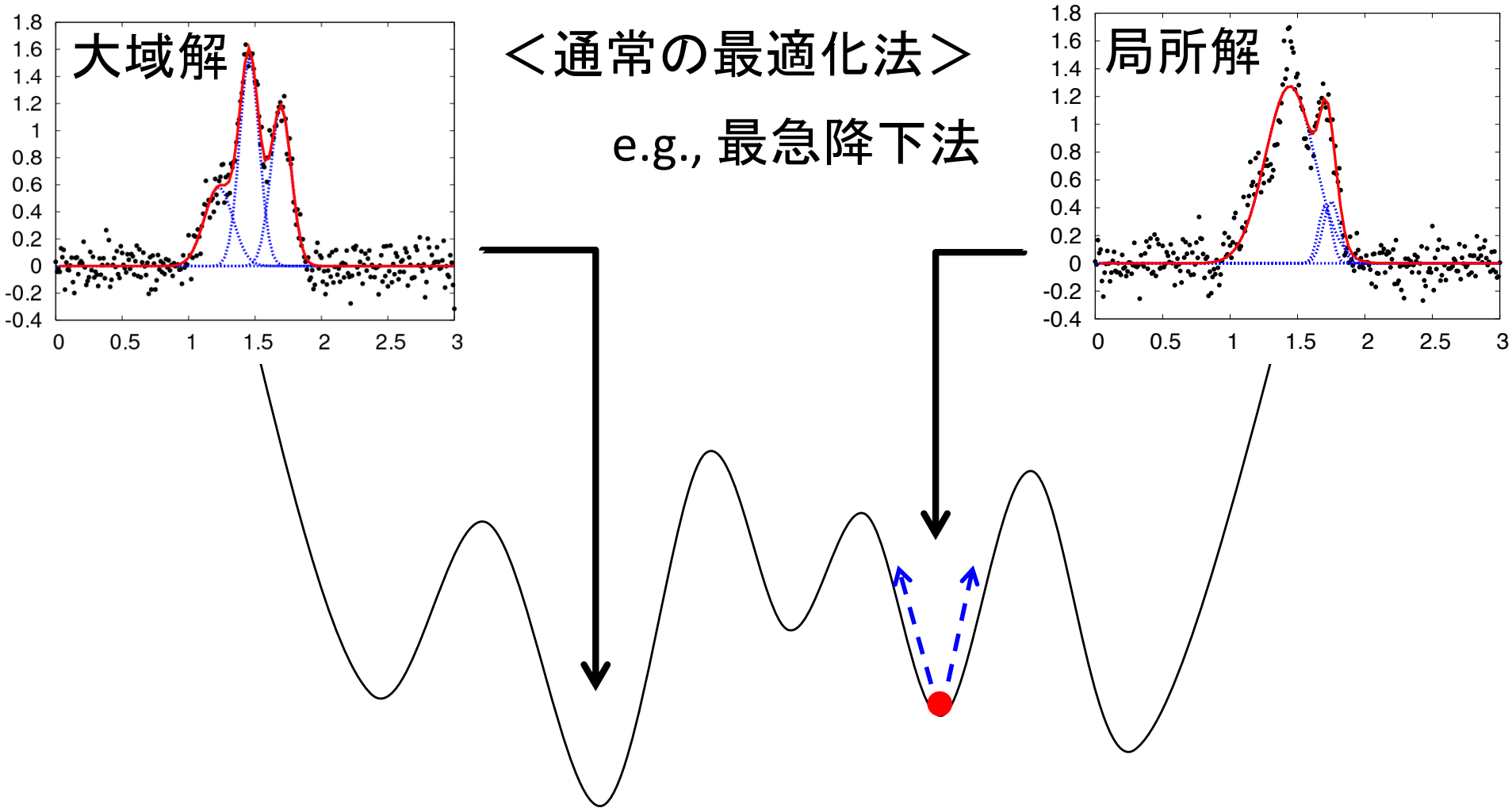
$$\theta = \{a_k, b_k, \mu_k\} \quad k = 1, \cdots, K$$



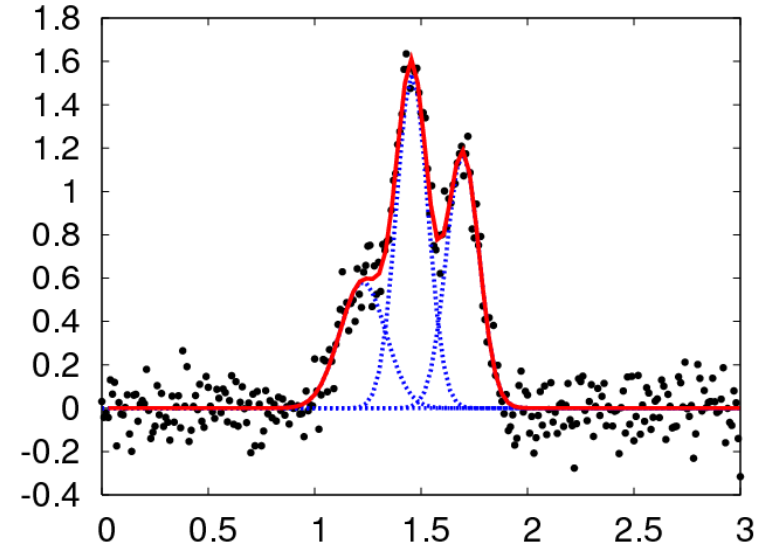
二乗誤差を最小にするようにパラメータをフィット（最小二乗法）

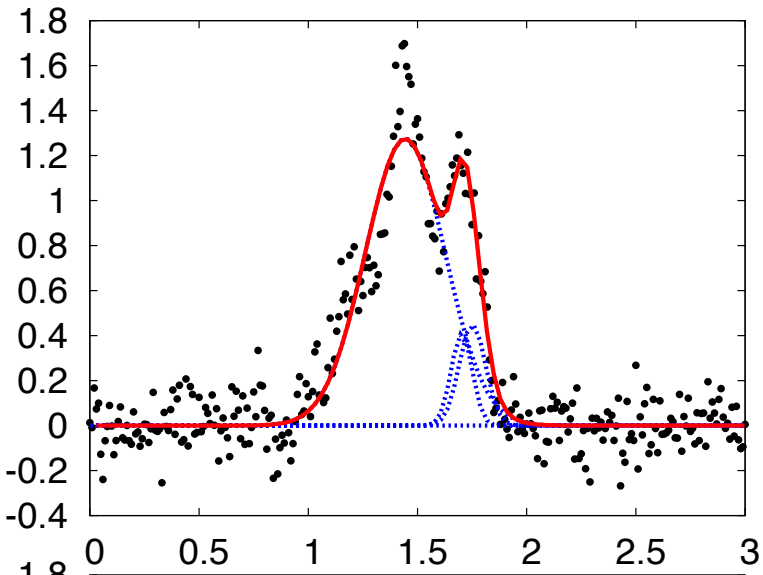
$$E(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f\left(x_i;\theta\right)\right)^2$$

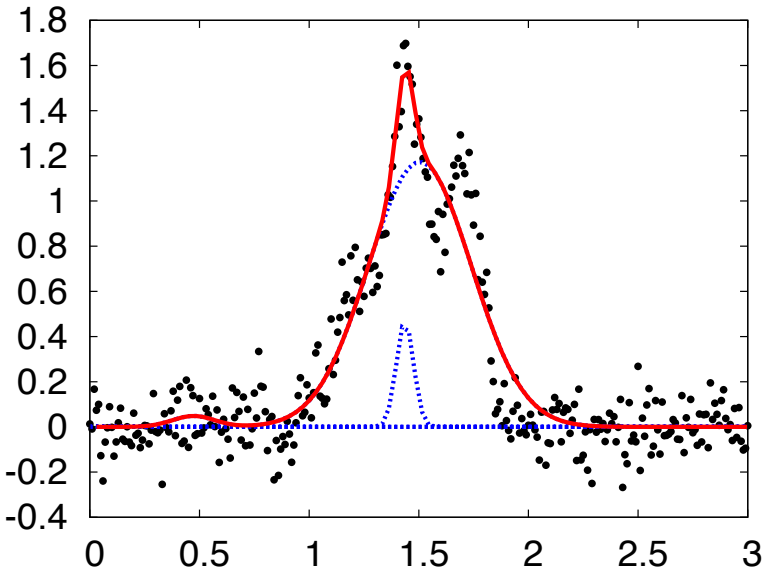
誤差関数は局所解を持つ

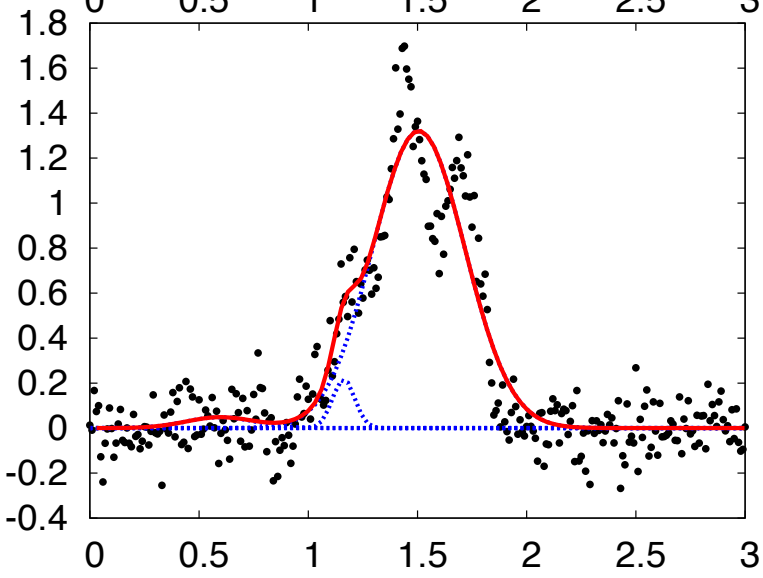


ローカルミニマム


1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
0.5
1
1.5
2
2.5
3


1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
0.5
1
1.5
2
2.5
3


1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
0.5
1
1.5
2
2.5
3


1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
0.5
1
1.5
2
2.5
3

ベイズ計測

順アプローチ

計測データ ⇒ 実験の結果 ⇕ 比較 理論の結果

モデル ⇒ 理論の結果

$p(Y|\theta,K)$ 解析計算，数値計算

逆アプローチ

対象とする物理系 ⇒ 系の物理モデル $p(\theta,K)$ ⇒ 観測過程 計測機器の特性 $p(\theta,K|Y)$ ⇒ 計測データ

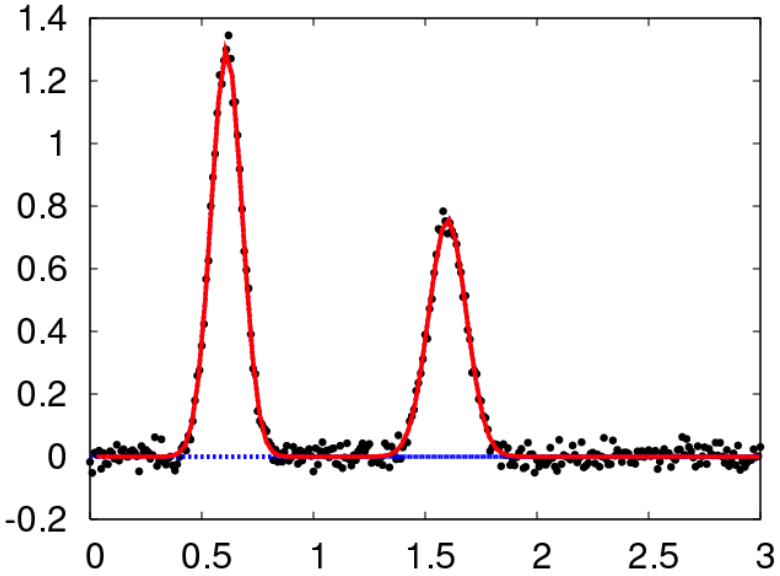
全てをモデル化し
ベイズの定理で因果をさかのぼる

確率的定式化

出力は，入力からの応答とノイズの足し合わせにより生成

⇒出力は，確率変数である．

$$y_i = f\left(x_i;\theta\right) + \varepsilon$$



ノイズが正規分布であるとすると，

$$p\left(y_i \mid \theta\right) \propto \exp\left(-\frac{1}{2}\left(y_i - f\left(x_i;\theta\right)\right)^2\right)$$

それぞれの出力 y_i が，独立であるとすると，

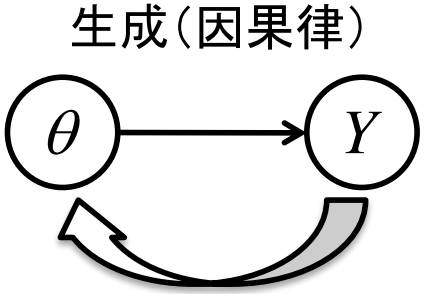
$$p\left(Y \mid \theta\right) = \prod_{i=1}^{n} p\left(y_i \mid \theta\right) \propto \exp\left(-nE\left(\theta\right)\right) \qquad Y = \left\{y_1, \cdots, y_n\right\}$$

$$E\left(\theta\right) = \frac{1}{2n}\sum_{i=1}^{n}\left(y_i - f\left(x_i;\theta\right)\right)^2$$

ボルツマン分布

ベイズ推論：因果律を組み込んでデータ解析

$$p(Y,\theta) = p(Y \mid \theta)p(\theta) = p(\theta \mid Y)p(Y)$$



＜ベイズの定理＞

$$p(\theta \mid Y) = \frac{p(Y \mid \theta)p(\theta)}{p(Y)} \propto \exp\left(-nE(\theta)\right)p(\theta)$$

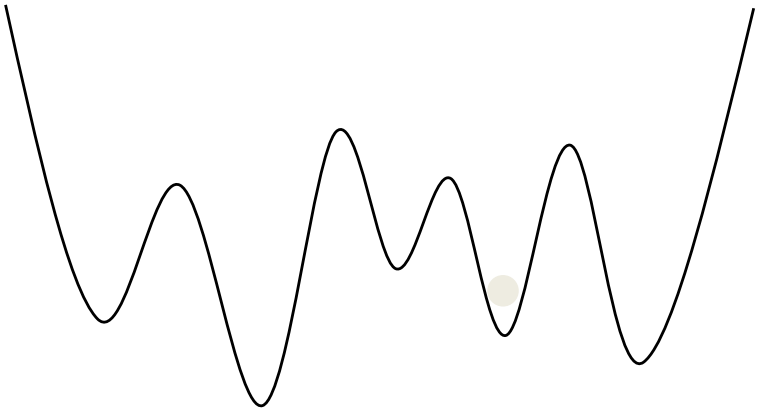
$p(\theta \mid Y)$ ：事後確率。データが与えられたもとでの，パラメータの確率.

$p(\theta)$ ：事前確率。あらかじめ設定しておく必要がある。
これまで蓄積されてきた科学的知見

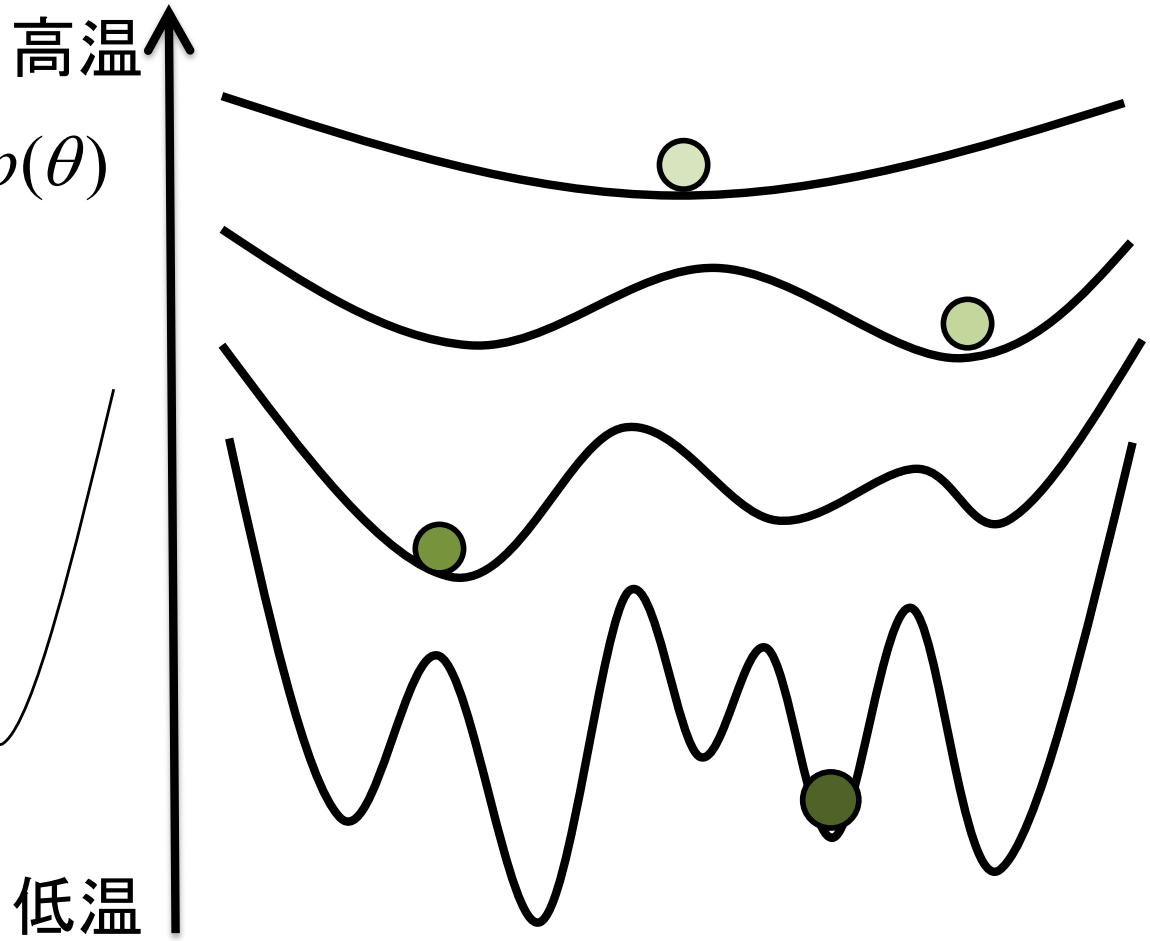
レプリカ交換モンテカルロ法

メトロポリス法

$$p_\beta(\theta) \propto \exp\left(-\frac{n}{\sigma^2}\beta E(\theta)\right) p(\theta)$$

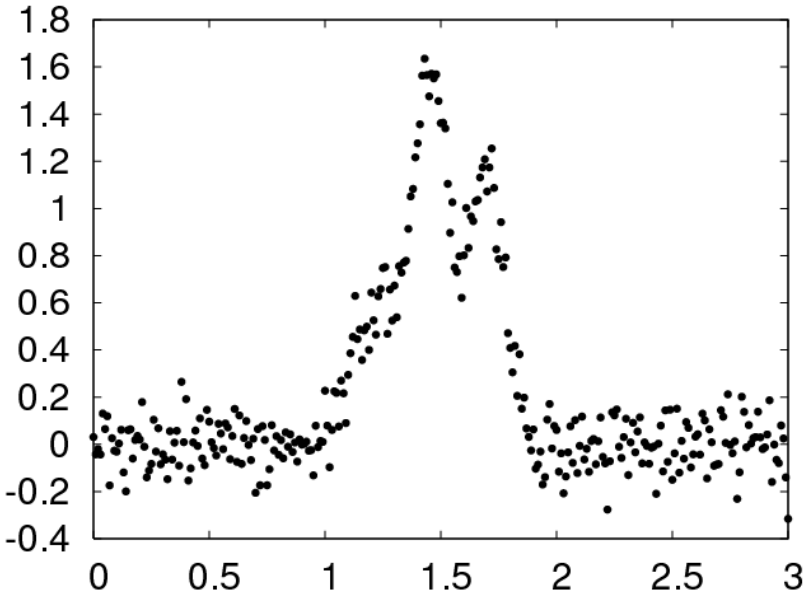


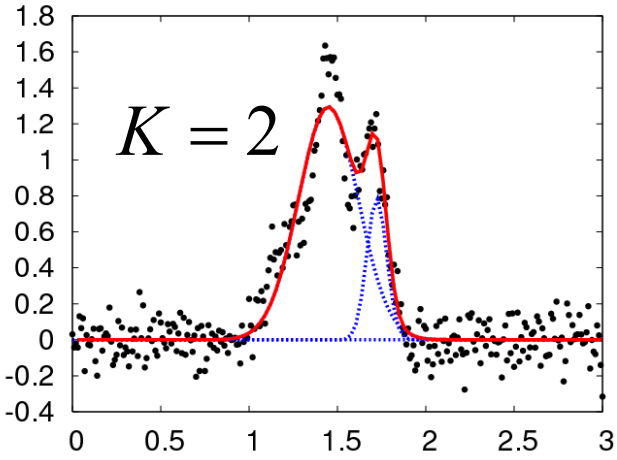
レプリカ交換モンテカルロ法

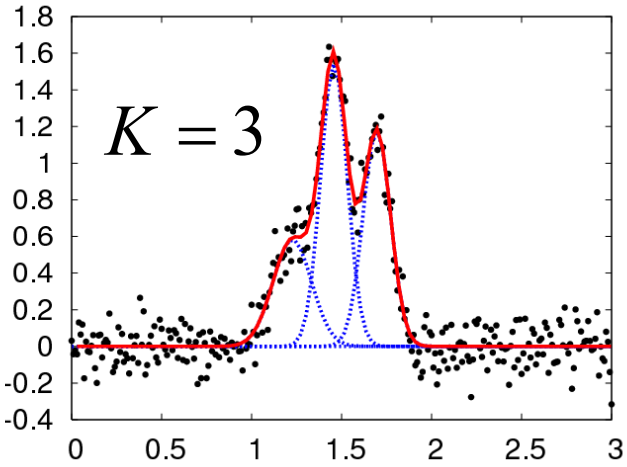


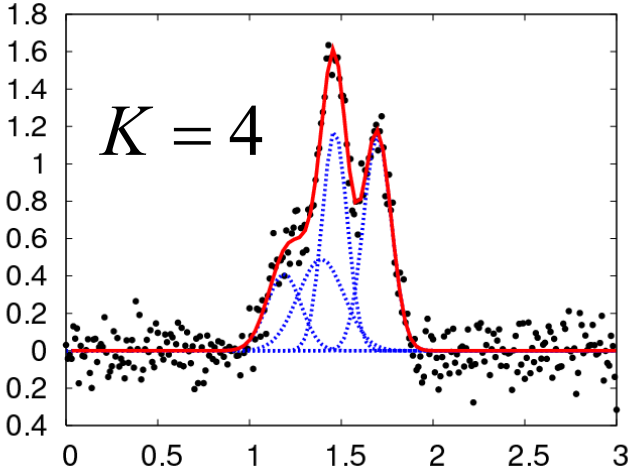
K. Hukushima, K. Nemoto, *J. Phys. Soc. Jpn.* **65** (1996).

スペクトル分解



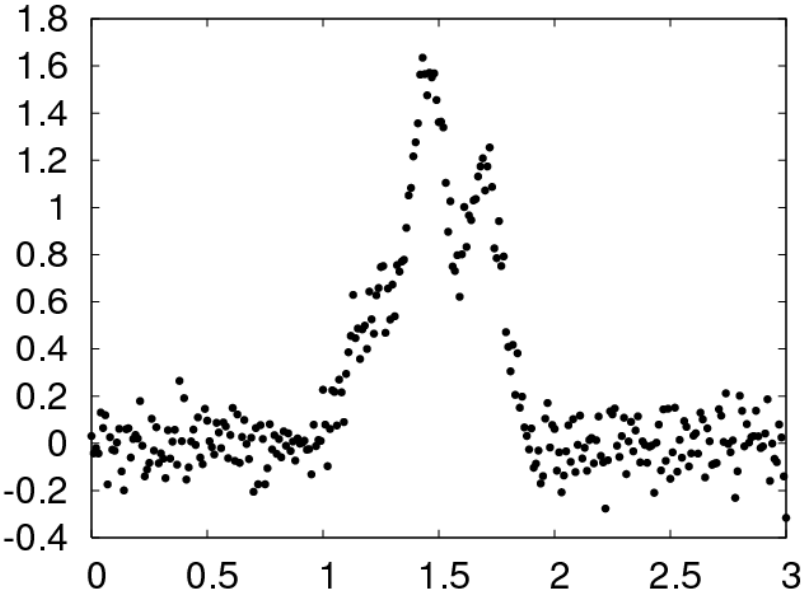


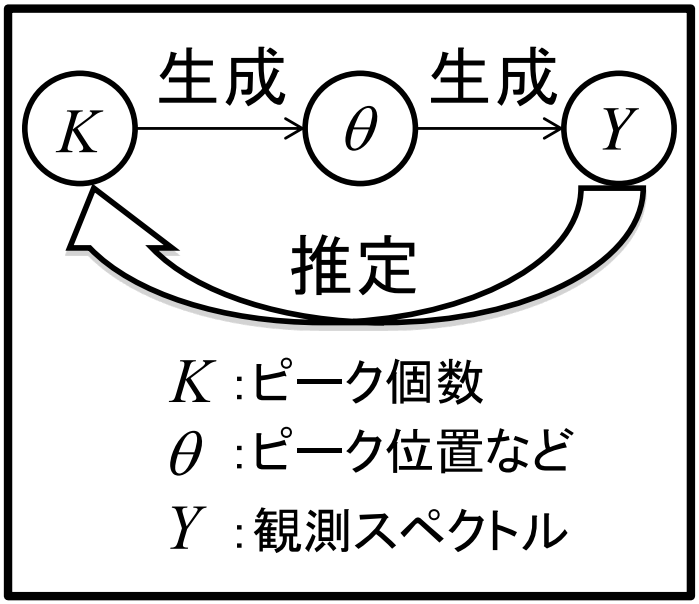


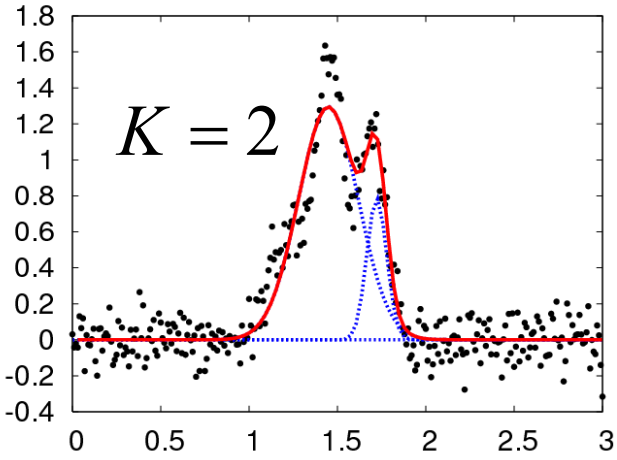


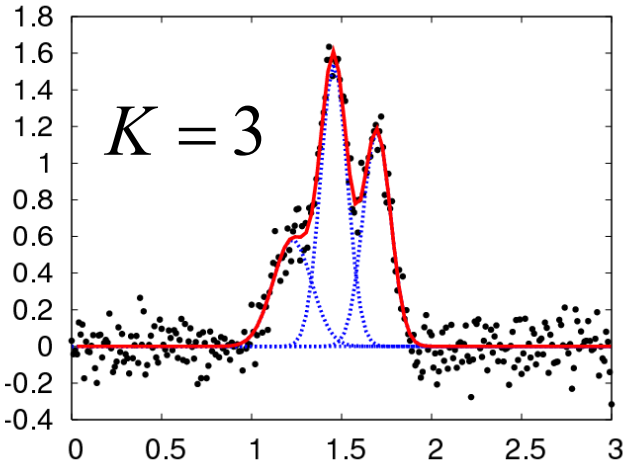
Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

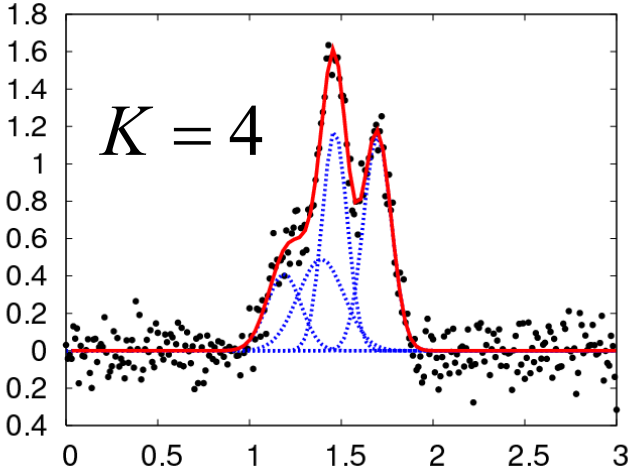
より深い構造をさぐる: モデル選択











Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

モデル選択

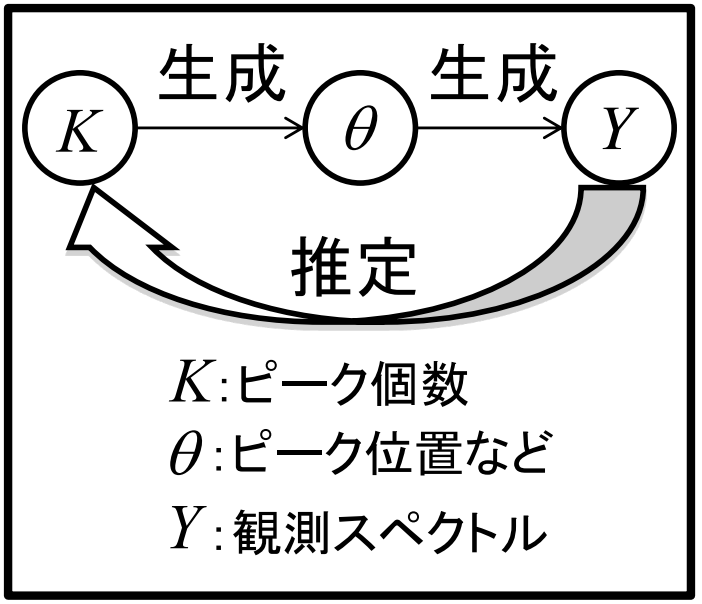
1. 欲しいのは　$p(K|Y)$

2. θがないぞ

3. $p(K,\theta,Y)$ の存在を仮定

$$p(K,\theta,Y) = p(Y|\theta,K)p(K)$$

$$p(Y|\theta,K) = \prod_{i=1}^{n} p(y_i|\theta) \propto \exp\left(-nE(\theta)\right)$$



4. 無駄な自由度の系統的消去：周辺化，分配関数

$$p(K,Y) = \int p(K,\theta,Y)d\theta$$

$$p(K|Y) = \frac{p(Y|K)p(K)}{p(Y)} \propto p(K)\int \exp\left(-nE(\theta)\right)p(\theta)d\theta$$

$$F(K) = -\log \int \exp\left(-nE(\theta)\right)p(\theta)d\theta$$

自由エネルギーを最小にする個数 K を求める.

自由エネルギーの数値的計算法
レプリカ交換法の性質を巧妙に使う

自由エネルギー:
$$F = -\log \int \exp\left(-\frac{n}{\sigma^2} E(\theta)\right) p(\theta) d\theta$$

以下のように、補助変数 β を導入する。　　β : 逆温度

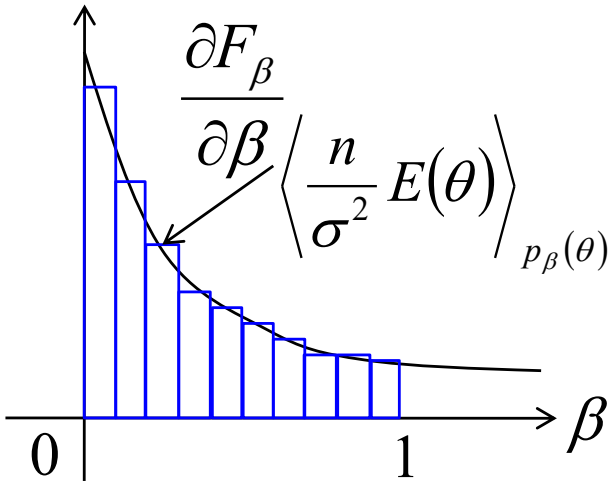
$$F_\beta = -\log \int \exp\left(-\frac{n}{\sigma^2} \beta E(\theta)\right) p(\theta) d\theta \quad \left(F_{\beta=0} = 0\right)$$

$$F = F_{\beta=1} = \int_0^1 d\beta \frac{\partial F_\beta}{\partial \beta}$$

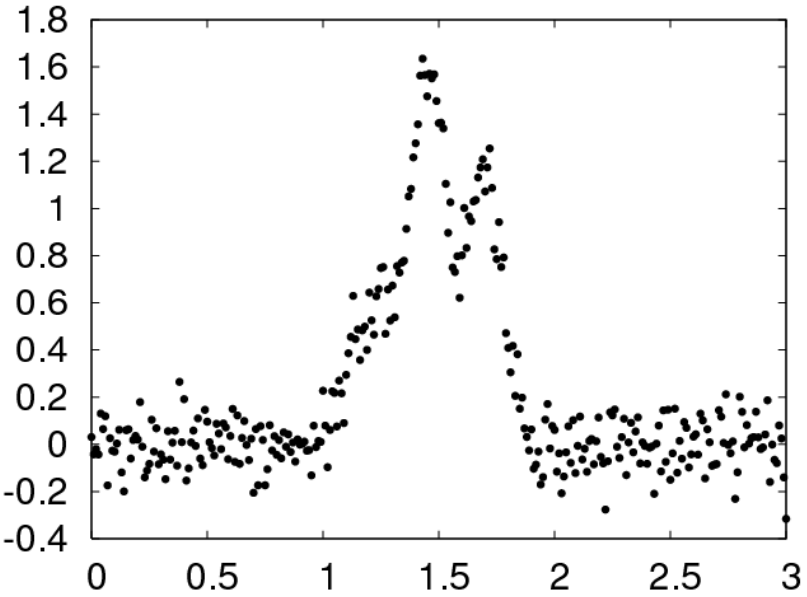
たくさんの温度でのシミュレーションが必要
→ 各温度でのエネルギー平均(すでにやってる)

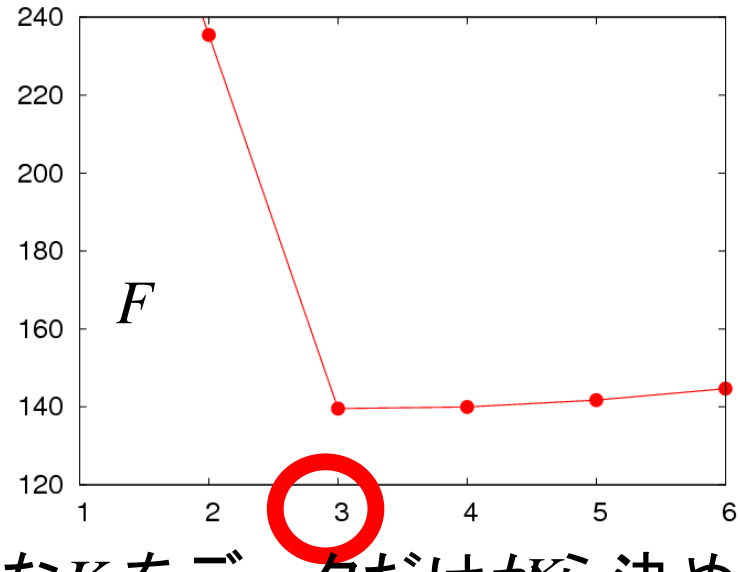
$\frac{\partial F_\beta}{\partial \beta}$ … 確率分布 $p(\theta;\beta)$ に従う二乗誤差 $\frac{n}{\sigma^2} E(\theta)$ の期待値

$$p_\beta(\theta) \propto \exp\left(-\frac{n}{\sigma^2} \beta E(\theta)\right) p(\theta)$$



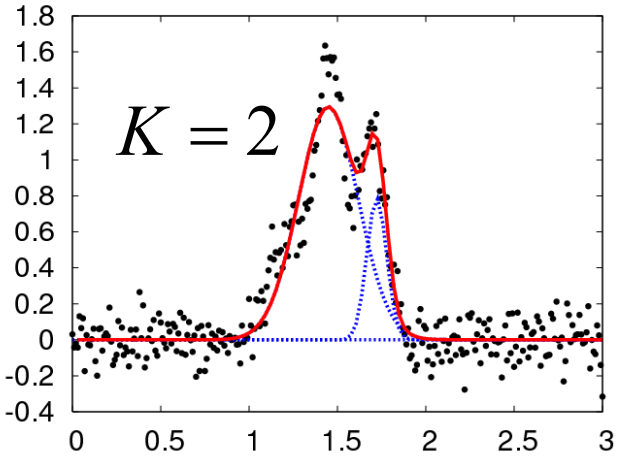
スペクトル分解

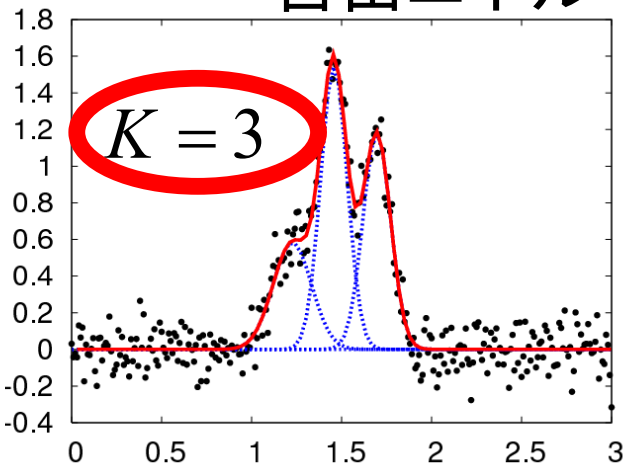


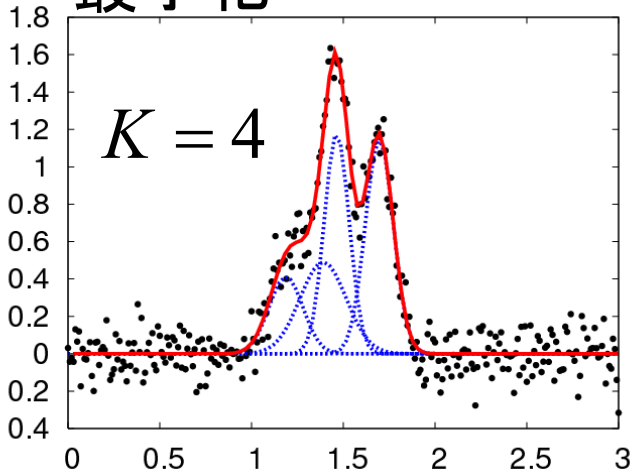


最適なKをデータだけから決める

自由エネルギー最小化







Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

ベイズ計測はこのような問題も取り扱える

- スペクトルや画像データからフィッティングを行なっている
- そのフィッティングの際に、パラメータを手打ちで決めている。最急降下法などを使っているが、うまくいかない。
- フィッティング用のモデルが複数あって、事前にどれを使うかを決めておかないといけない。
- S/Nが悪いデータや欠損データをなんとかした。
- 複数計測の統合を行いたい。

- そのような方は、一度ベイズ計測をお試しください。

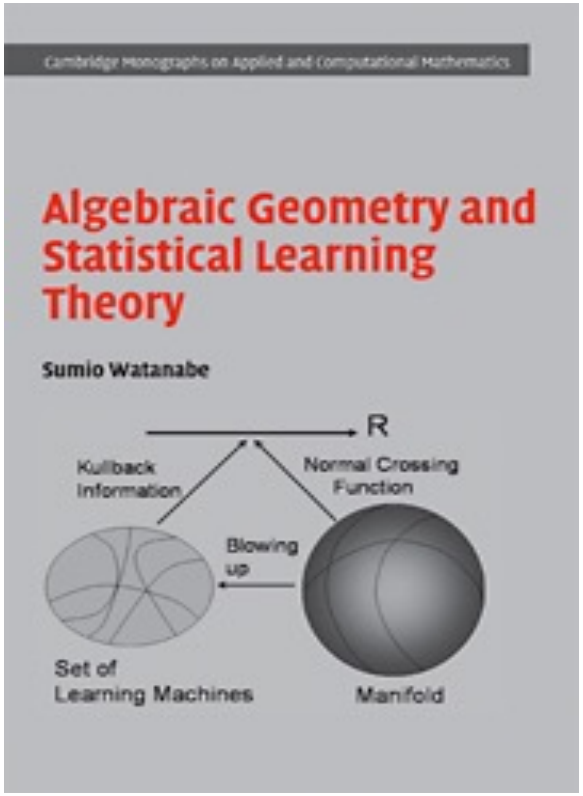
内容

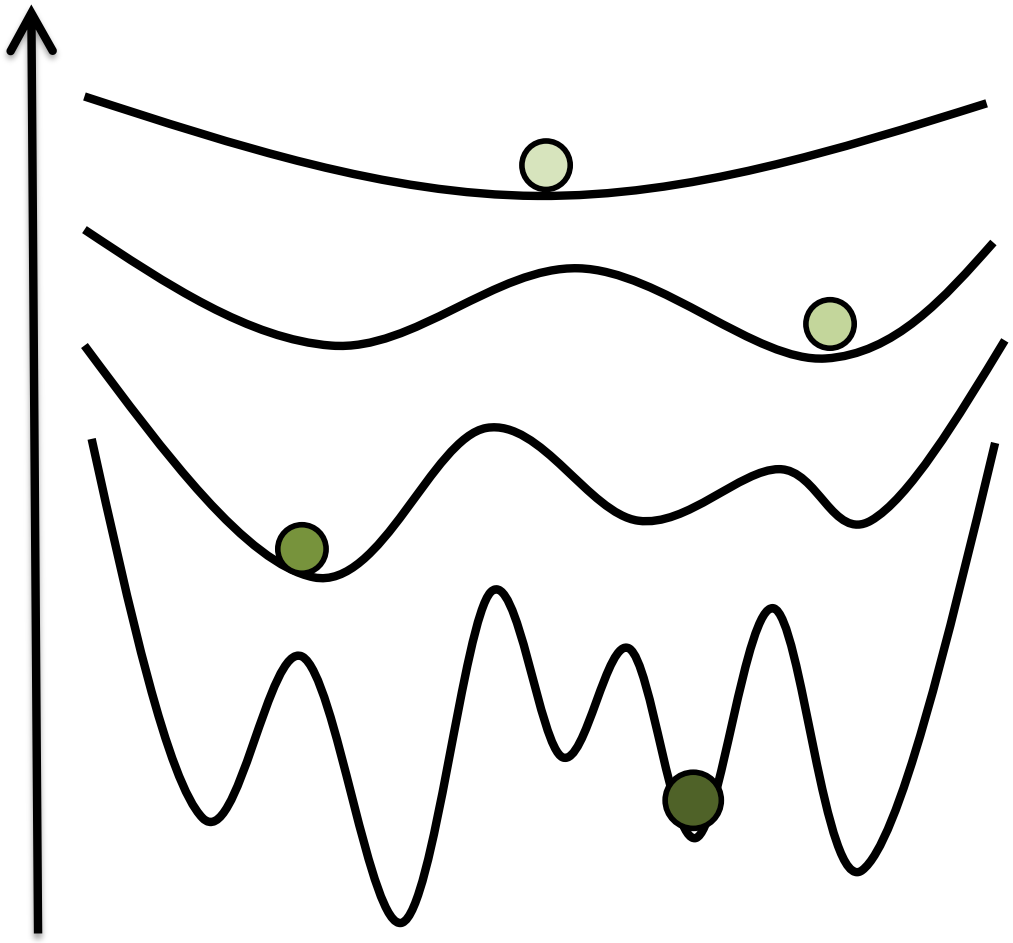
1. 自己紹介
2. 人工知能を明日から利用するには?
3. 顔応答細胞のポピュレーションダイナミクス
4. 新学術領域研究「スパースモデリングの深化と高次元データ駆動科学の創成」
5. ベイズ計測とスペクトル分解
6. SPring-8全ビームラインベイズ化計画
7. まとめ

SPring-8

- アメリカのAdvanced Photon Source (APS), ヨーロッパのEuropean Synchrotron Radiation Facility(ESRF) と合わせて，世界3大放射光施設.
- 理研はSPring-8を「データ創出基盤」であると言っている．年間延べ1万人が利用.
- APSやESRFにおいてベイズ計測は導入されていない.
- 放射光におけるベイズ計測に関しては日本が最先端である.

APSやESRFにおいてベイズ計測が導入されていない理由





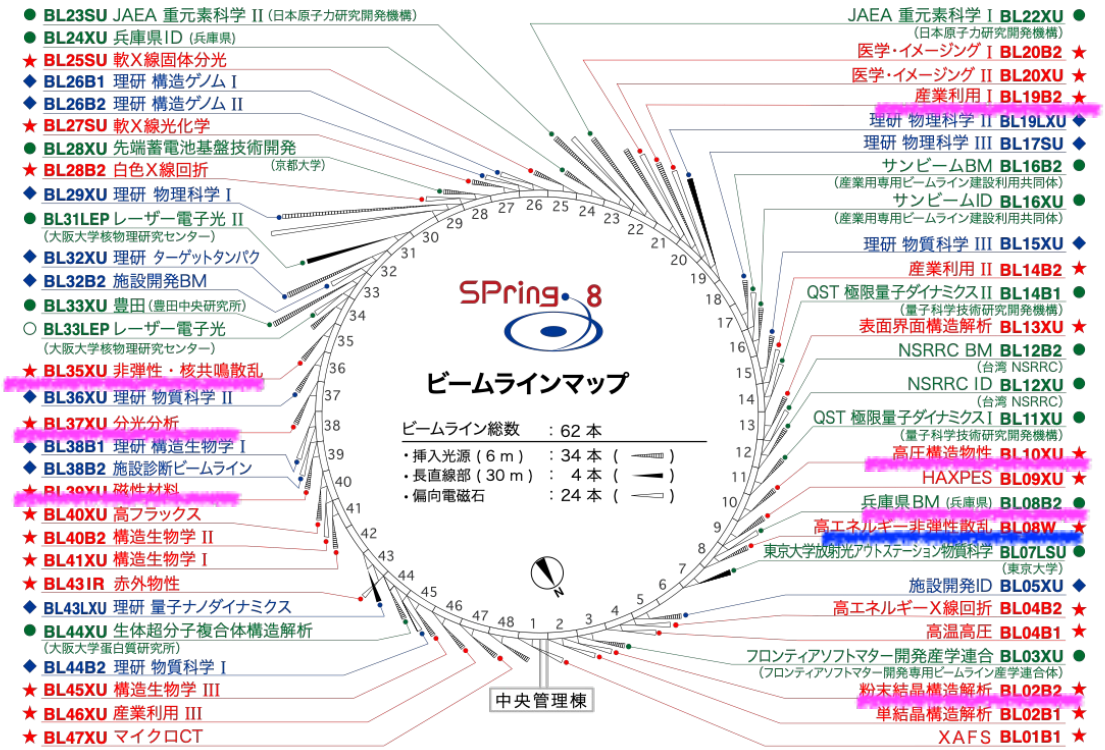
多くのスペクトル解析が統計的特異モデルであることを知らない

レプリカ交換モンテカルロ法の知見の欠如

水牧 仁一朗
公益財団法人高輝度光科学研究センター
放射光利用研究基盤センター
コーディネータ
4月1日より熊本大学大学院先端科学研究部

SPring-8全ビームラインベイズ化計画

敬称略



情報と放射光研究者のマッチング

メスバウアー

BL35XU　岡田研学生+筒井

小角散乱

BL08B2　岡田研学生+桑本

BL19B2

XAS測定

BL37XU　岡田研学生+水牧

BL39XU

放射光ユーザーへの展開

時分割XRD

BL02B2　横山優一+河口彰吾、沙織

BL10XU　ユーザー：公立大、東工大

赤色BLが共用BL(JASRI担当): 計26本

全BL本数：62本

来年度には過半数をこえる予定

年度	2021	2022	2023
導入	2	8	14
全BL	26	26	26

まとめ

1. 自己紹介
2. 人工知能を明日から利用するには?
3. 顔応答細胞のポピュレーションダイナミクス
4. 新学術領域研究「スパースモデリングの深化と高次元データ駆動科学の創成」
5. ベイズ計測とスペクトル分解
6. SPring-8全ビームラインベイズ化計画
7. まとめ