

令和 4 年 5 月 17 日現在

機関番号：14301

研究種目：基盤研究(C) (一般)

研究期間：2018～2021

課題番号：18K11463

研究課題名(和文) スパースモデリングと情報統計力学の共進化による柔軟な大規模逆問題解法の開発と応用

研究課題名(英文) Statistical mechanics and sparse modeling approach to large-scale inverse problems

研究代表者

小淵 智之 (Obuchi, Tomoyuki)

京都大学・情報学研究科・准教授

研究者番号：40588448

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本課題では、スパースモデリング(SpM)と情報統計力学の技術を拡張することで、変数選択・パラメータ推定問題の、より深い数理解と実用的数値解法を得ることを目標として研究してきた。その結果、多くの成果(複数のhigh impact journalを含む15本の査読あり論文、3つの近似アルゴリズムパッケージのGithubでの公開)を得ることが出来た。特に、交差検証法を低計算量で実行する近似アルゴリズムの開発と関連する理論解析、構造学習に関連する逆イジング問題の理論解析、ブートストラップ法を利用した変数選択の近似アルゴリズムとその理論解析が独創的な成果としてあげられる。

研究成果の学術的意義や社会的意義

近年、機械学習や人工知能といったキーワードで、ある種の数学的モデリング方法が注目を集めている。本研究では、そのモデリング方法の中でも、スパースモデリングという、様々な事物(変数)の中から重要なものを半自動的に取り出す技術に着目し、その技術を高めるべく研究を行った。多くの場合、計算量的問題から適切な変数の取り出しが難しくなるのだが、それを情報統計力学という物理分野にある手法で問題を近似することにより、計算量を削減するという方針で研究を行った。その結果、正確性を保ったまま以前よりも速やかに変数選択をすることが可能になった。機械学習を大規模に運用する上で重要な成果と成る。

研究成果の概要(英文)：This project has been aimed at obtaining a deeper mathematical understanding and practical numerical solutions to variable selection and parameter estimation problems by extending the techniques of sparse modelling (SpM) and statistical mechanical informatics. As a result, a number of results (16 peer-reviewed papers, including several high impact journals, and three approximation algorithm packages published on Github) have been obtained. In particular, ingenious results include the development and related theoretical analysis of approximation algorithms that perform cross-validation methods with low computational complexity, theoretical analysis of inverse Ising problems related to structural learning, and an approximation algorithm for variable selection using the bootstrap method and its theoretical analysis.

研究分野：統計物理学

キーワード：情報統計力学 スパースモデリング 統計的逆問題 機械学習 近似推論

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

(1) 計測技術のたゆまない進歩により、大規模高次元データを処理する必要が様々な分野に現れてきた。高次元データから低次元な情報を抽出する枠組みは、一般に多変量解析と呼ばれる。その系統的定式化は存在するが、その枠組みに忠実に従うと、多くの高次元データにおいて、モデルの解釈上の困難と解析を実行する上での計算量的困難に直面する。

(2) このような困難を避ける上で注目されている技術がスパースモデリング(SpM)である。SpMとは、データの“スパース”な表現を見つけることで、ノイズ除去・高効率データ圧縮・結果の解釈性向上などを可能にする方法論である。SpMは比較的新しい枠組みであり、スパース表現抽出のための理論的枠組の整備やアルゴリズム開発、及びそれらの実データ解析への応用が世界的に活発に進められている。

(3) SpMの素直な定式化に従うと、やはり計算量的困難に直面する。従って、大規模な高次元データから科学的知見・工学的応用を見出すには、そのような困難を克服することが必要であり、システムティックに適用可能な方法論が模索されている。

(4) その計算量的困難を克服する上で、情報統計力学とよばれる分野技術が注目されている。情報統計力学とは、物理の一分野である統計力学に蓄積している大規模確率モデルをハンドリングする技術を、情報科学の問題に適用することで成果を得るという分野であり、これまでに情報理論や計算量理論における様々な問題に適用され、その有効性を実証してきている。

### 2. 研究の目的

(1) 情報統計力学の技術を利用して、SpMを実行する上での計算量的困難を回避する枠組み・アルゴリズムを構築する。

(2) 情報統計力学の技術を利用して、SpMの枠組みにおける情報抽出の理論限界を明らかにする。

(3) SpM自身が現在進行系で進化する手法である。その適用限界を具体的な問題群の解析を通じて拡大すると同時に新しい手法・枠組みを開発する。

(4) 上記の問題の解析を通じて、情報統計力学における方法論そのものを進化させる。

### 3. 研究の方法

まず、取り扱う問題の範囲を一般化線形モデル、あるいは局所的にはそうみなせるモデルに限定することで、現実的に解析・開発を行えるようにする。その上で

(1) モンテカルロ法という統計力学のアルゴリズムを利用して、SpMの変数選択アルゴリズムを開発・改良する。

(2) 摂動論、あるいは線形応答近似と呼ばれる統計力学の方法を利用することで、モデルのハイパーパラメータ選択を近似的に軽量に計算する方法を開発する。

(3) レプリカ法という統計力学の方法を利用することで、統計モデルの典型的な性能を解明する。

(4) 近似的メッセージ伝搬法という、統計力学の平均場近似を改良したアルゴリズムの枠組みを利用し、低計算量で推定を行うアルゴリズムを開発する。

(5) レプリカ法と近似的メッセージ伝搬法を併用することで、ブートストラップ法などの発展的な方法による変数選択法を近似的に軽量に実行するアルゴリズムを開発する。

### 4. 研究成果

#### (1) 圧縮センシングにおける変数選択アルゴリズムの開発

圧縮センシングの素直な定式化では、推定において計算量爆発を起こすため、L1緩和などの緩和法が使われる事が多い。一方、緩和によって推定が正しくできる領域が狭くなることも明らかにされているため、緩和されていない元々の問題を解くアルゴリズムがあれば便利である。この考えに基づき、モンテカルロ法を用いたアルゴリズムを考案し、その性能解析を実データへの適用を行った。結果として、この方法がL1緩和を凌駕するとともに、より多くの事前情報を要求するベイズによる定式化に迫る性能を発揮することが、実験と理論の双方から示された。この

結果により、論文を2本出版することが出来た。

#### (2) 非凸正則化による圧縮センシングの理論性能の解明とアルゴリズムの開発

圧縮センシングにおいて、L1 緩和とは異なる方法として、非凸正則化を用いた緩和法が近年注目されている。その理論性能を解明すると同時に近似的メッセージ伝搬法に基づくアルゴリズム開発を行った。結果として、この方法がL1 緩和を凌駕するとともに、より多くの事前情報を要求するベイズによる定式化に迫る性能を発揮することが、実験と理論の双方から示された。またこの結果は場合によってはベイズによる定式化さえ上回る可能性が示唆された。この結果により、論文を1本出版することが出来た。

#### (3) 圧縮センシング/機械学習におけるハイパーパラメータ選択の軽量化

圧縮センシングや近年の機械学習モデルにおいて、ハイパーパラメータを選択することは重要である。多くの場合、この目的でホールドアウト検証法や交差検証法が利用される。この方法は非常に汎用的だが、最適化を何度もやり直す必要が生じるため計算量的な負荷が高い。その負荷を軽減すべく、摂動論/線形応答近似を用いて交差検証法を近似的に実行することで計算量増加なくハイパーパラメータ選択を行う方法を開発した。非凸圧縮センシング・多クラス分類でこれを実装して、2本論文を出版することが出来た。さらに、これとは別に、ブートストラップ法を利用した変数選択・ハイパーパラメータ選択の方法を近似的に行う方法をレプリカ法を用いて開発した。これによりさらに1本の論文を出版することが出来た。

#### (4) 逆イジング問題の理論解析

変数間の相互作用ネットワークを推定する問題が、データ解析手法の一つとして近年注目を集めている。特に、バイナリ変数の場合は、物理におけるイジングモデルとの相性がよく、イジングモデルを用いてネットワーク推定を行う逆イジング問題が注目されている。逆イジング問題を用いた場合の性能限界の解析はあまり行われおらず、それをレプリカ法を用いて行った。推定モデルとデータの生成モデルが合致するか否か、正則化を用いるか否かなどでいくつかのバージョンが考えられ、それらの網羅的な解析を行った。結果としてこれまでに3本の論文を出版することが出来た。他の場合もまだ研究を進めており、今後も論文が複数出る予定である。

#### (5) 機械学習モデルにおける理論性能解析

情報統計力学の方法を用いて、圧縮センシングだけではなく機械学習モデルの理論解析も行った。グラフニューラルネットワークやランダム PCA の解析を行いその理論性能限界を明らかにした。これにより2本論文を出版することが出来た。

#### (6) ボルツマンマシンにおける経験ベイズ法の近似アルゴリズム

経験ベイズ法とは、ベイズ推定においてハイパーパラメータを決定する手法の一つである。一般にこれを実行するには経験ベイズ尤度関数という計算量的負荷の高い量を計算する必要があるが、ここにレプリカ法と Plefka 展開を利用することで、低計算量でこれを計算する手法を構成した。これにより1本論文を出版することが出来た。

#### (7) 生物データの解析

生物のデータを解析して有用な情報(関係のある素子(ニューロン・遺伝子・化学種など)や伝達される情報量)を抽出することは、生物の体内で何が行われているかを理解する上で重要である。現実の生物データを機械学習の手法を用いて解析することで実際に有用な情報を抽出することを試みた。逆イジング問題を用いてニューロン間ネットワークを抽出する問題や、生体内細胞において化学種同士でやり取りされている情報の定量化などを行った。前者で2本、後者で1本、論文を出版することが出来た。

## 5. 主な発表論文等

〔雑誌論文〕 計15件（うち査読付論文 15件 / うち国際共著 1件 / うちオープンアクセス 10件）

1. 著者名 Tomoyuki Obuchi, Yoshiyuki Kabashima	4. 巻 20
2. 論文標題 Semi-Analytic Resampling in Lasso}	5. 発行年 2019年
3. 雑誌名 Journal of Machine Learning Research	6. 最初と最後の頁 1--33
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Imaizumi Takuya, Umeki Nobuhisa, Yoshizawa Ryo, Obuchi Tomoyuki, Sako Yasushi, Kabashima Yoshiyuki	4. 巻 105
2. 論文標題 Assessing transfer entropy from biochemical data	5. 発行年 2022年
3. 雑誌名 Physical Review E	6. 最初と最後の頁 1-13
掲載論文のDOI (デジタルオブジェクト識別子) 10.1103/physreve.105.034403	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Sakata Ayaka, Obuchi Tomoyuki	4. 巻 2021
2. 論文標題 Perfect reconstruction of sparse signals with piecewise continuous nonconvex penalties and nonconvexity control	5. 発行年 2021年
3. 雑誌名 Journal of Statistical Mechanics: Theory and Experiment	6. 最初と最後の頁 093401 ~ 093401
掲載論文のDOI (デジタルオブジェクト識別子) 10.1088/1742-5468/ac1403	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Tagashira Fumito, Obuchi Tomoyuki, Tanaka Toshiyuki	4. 巻 none
2. 論文標題 Sharp Asymptotics of Matrix Sketching for a Rank-One Spiked Model	5. 発行年 2021年
3. 雑誌名 2021 IEEE International Symposium on Information Theory (ISIT)	6. 最初と最後の頁 250-255
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/isit45174.2021.9517724	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Xiangming Meng, Tomoyuki Obuchi, Yoshiyuki Kabashima	4. 巻 34
2. 論文標題 Ising Model Selection Using $\ell_1$ -Regularized Linear Regression: A Statistical Mechanics Analysis	5. 発行年 2021年
3. 雑誌名 Advances in Neural Information Processing Systems	6. 最初と最後の頁 6290-6303
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Meng Xiangming, Obuchi Tomoyuki, Kabashima Yoshiyuki	4. 巻 2021
2. 論文標題 Structure learning in inverse Ising problems using $\ell_2$ -regularized linear estimator	5. 発行年 2021年
3. 雑誌名 Journal of Statistical Mechanics: Theory and Experiment	6. 最初と最後の頁 053403 ~ 053403
掲載論文のDOI (デジタルオブジェクト識別子) 10.1088/1742-5468/abfa10	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hayashi Kao, Obuchi Tomoyuki, Kabashima Yoshiyuki	4. 巻 89
2. 論文標題 Reconstructing Sparse Signals via Greedy Monte-Carlo Search	5. 発行年 2020年
3. 雑誌名 Journal of the Physical Society of Japan	6. 最初と最後の頁 124802 ~ 124802
掲載論文のDOI (デジタルオブジェクト識別子) 10.7566/JPSJ.89.124802	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Terada Yu, Obuchi Tomoyuki, Isomura Takuya, Kabashima Yoshiyuki	4. 巻 32
2. 論文標題 Inferring Neuronal Couplings From Spiking Data Using a Systematic Procedure With a Statistical Criterion	5. 発行年 2020年
3. 雑誌名 Neural Computation	6. 最初と最後の頁 2187 ~ 2211
掲載論文のDOI (デジタルオブジェクト識別子) 10.1162/neco_a_01324	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Abbara Alia, Kabashima Yoshiyuki, Obuchi Tomoyuki, Xu Yingying	4. 巻 2020
2. 論文標題 Learning performance in inverse Ising problems with sparse teacher couplings	5. 発行年 2020年
3. 雑誌名 Journal of Statistical Mechanics: Theory and Experiment	6. 最初と最後の頁 073402 ~ 073402
掲載論文のDOI (デジタルオブジェクト識別子) 10.1088/1742-5468/ab8c3a	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Yasuda Muneki, Obuchi Tomoyuki	4. 巻 53
2. 論文標題 Empirical Bayes method for Boltzmann machines	5. 発行年 2019年
3. 雑誌名 Journal of Physics A: Mathematical and Theoretical	6. 最初と最後の頁 014004 ~ 014004
掲載論文のDOI (デジタルオブジェクト識別子) 10.1088/1751-8121/ab57a7	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Obuchi Tomoyuki, Sakata Ayaka	4. 巻 52
2. 論文標題 Cross validation in sparse linear regression with piecewise continuous nonconvex penalties and its acceleration	5. 発行年 2019年
3. 雑誌名 Journal of Physics A: Mathematical and Theoretical	6. 最初と最後の頁 414003 ~ 414003
掲載論文のDOI (デジタルオブジェクト識別子) 10.1088/1751-8121/ab3e89	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Yu Terada, Tomoyuki Obuchi, Takuya Isomura, Yoshiyuki Kabashima	4. 巻 31
2. 論文標題 Objective and efficient inference for couplings in neuronal networks	5. 発行年 2018年
3. 雑誌名 Advances in Neural Information Processing Systems	6. 最初と最後の頁 4976--4985
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Tatsuro Kawamoto, Masashi Tsubaki, Tomoyuki Obuchi	4. 巻 31
2. 論文標題 Mean-field theory of graph neural networks in graph partitioning	5. 発行年 2018年
3. 雑誌名 Advances in Neural Information Processing Systems	6. 最初と最後の頁 4366--4376
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Tomoyuki Obuchi, Yoshinori Nakanishi-Ohno, Masato Okada, Yoshiyuki Kabashima	4. 巻 103401
2. 論文標題 Statistical mechanical analysis of sparse linear regression as a variable selection problem	5. 発行年 2018年
3. 雑誌名 J. Stat. Mech.	6. 最初と最後の頁 103401
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Tomoyuki Obuchi, Yoshiyuki Kabashima	4. 巻 19
2. 論文標題 Accelerating Cross-Validation in Multinomial Logistic Regression with L1-Regularization	5. 発行年 2018年
3. 雑誌名 Journal of Machine Learning Research	6. 最初と最後の頁 1--30
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計25件 (うち招待講演 4件 / うち国際学会 7件)

1. 発表者名 杉崎嵐, 寺田裕, 小淵智之, 樺島祥介
2. 発表標題 ニューラルネットにおける短期記憶の1ビット圧縮センシングによる想起
3. 学会等名 日本物理学会 第76回年次大会
4. 発表年 2021年

1. 発表者名 長澤莉希, 小淵智之, 吉野元
2. 発表標題 高ランクテンソル分解のBPに基づく推定
3. 学会等名 日本物理学会 第76回年次大会
4. 発表年 2021年

1. 発表者名 石井智, 小淵智之, 樺島祥介
2. 発表標題 SCAD正則化を用いた圧縮センシングに関する信号復元性能の分析
3. 学会等名 日本物理学会 第76回年次大会
4. 発表年 2021年

1. 発表者名 小淵智之
2. 発表標題 リサンプリングの近似計算アルゴリズムの理論と実践
3. 学会等名 スマートサンプリング講演会
4. 発表年 2020年

1. 発表者名 Alia Abbara, 樺島祥介, 小淵智之, 許インイン
2. 発表標題 教師の結合が疎な場合における逆イジング問題のレプリカ解析と仮説検定
3. 学会等名 日本物理学会2020年秋季大会
4. 発表年 2020年



1. 発表者名 今泉拓也, 小淵智之, 吉澤亮, 佐甲靖志, 樺島祥介
2. 発表標題 細胞内生化学反応に関する伝達情報量の解析
3. 学会等名 日本物理学会2020年秋季大会
4. 発表年 2020年

1. 発表者名 長澤莉希, 小淵智之, 吉野元
2. 発表標題 1+p体観測による多成分ベクトルの統計的推定
3. 学会等名 日本物理学会2020年秋季大会
4. 発表年 2020年

1. 発表者名 Alia Abbara, 樺島祥介, 小淵智之, 許インイン
2. 発表標題 逆インジグ問題のレプリカ法による性能解析: 教師の結合が疎な場合
3. 学会等名 第23回情報論的学習理論ワークショップ (IBIS2020)
4. 発表年 2020年

1. 発表者名 Obuchi Tomoyuki, Sakata Ayaka
2. 発表標題 Cross validation in sparse linear regression with piecewise continuous nonconvex penalties and its acceleration
3. 学会等名 2019 Workshop on Statistical Physics of Disordered Systems and Its Applications (SPDSA2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Obuchi Tomoyuki, Sakata Ayaka
2. 発表標題 Cross validation in sparse linear regression with piecewise continuous nonconvex penalties and its acceleration
3. 学会等名 Statistical Physics and Neural Computation (SPNC-2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Obuchi Tomoyuki
2. 発表標題 Statistical mechanical analysis of sparse linear regression as a variable selection problem
3. 学会等名 2019 International Workshop on Glass Physics in Beijing (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 小淵 智之
2. 発表標題 高次元確率モデルの扱いについて：統計物理の視点から
3. 学会等名 電子情報通信学会総合大会 (招待講演)
4. 発表年 2020年

1. 発表者名 小淵 智之
2. 発表標題 ボルツマンマシンによる神経細胞集団の有効な非対称結合推定
3. 学会等名 情報数物研究会 (招待講演)
4. 発表年 2019年

1. 発表者名 Tatsuro Kawamoto, Masashi Tsubaki, Tomoyuki Obuchi
2. 発表標題 Mean-field theory of graph neural networks in graph partitioning
3. 学会等名 NeurIPS2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Yu Terada, Tomoyuki Obuchi, Takuya Isomura, Yoshiyuki Kabashima
2. 発表標題 Objective and efficient inference dor couplings in neuronal networks
3. 学会等名 NeurIPS2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Tomoyuki Obuchi, Y. Nakanishi-Ohno, M. Okada and Y. Kabashima
2. 発表標題 Statistical mechanical analysis of sparse linear regression as a variable selection problem
3. 学会等名 Disordered serendipity: a glassy path to discovery (国際学会)
4. 発表年 2018年

1. 発表者名 Tomoyuki Obuchi and Yoshiyuki Kabashima
2. 発表標題 Accelerating Cross-Validation in Multinomial Logistic Regression with L1-Regularization
3. 学会等名 Statistical physics and machine learning back together (国際学会)
4. 発表年 2018年

1. 発表者名 安田宗樹, 小淵智之
2. 発表標題 レプリカ法を用いたボルツマンマシンに対する経験ベイズ推定
3. 学会等名 日本物理学会2019年年次大会
4. 発表年 2019年

1. 発表者名 小淵智之, 樺島祥介
2. 発表標題 Elastic-net正則化付き多項ロジスティック回帰における交差検証法の加速
3. 学会等名 日本物理学会2019年年次大会
4. 発表年 2019年

1. 発表者名 石井奨, 小淵智之, 樺島祥介
2. 発表標題 スパース線形回帰モデルにおけるレプリカ交換モンテカルロ法を用いた変数選択
3. 学会等名 日本物理学会2019年年次大会
4. 発表年 2019年

1. 発表者名 小淵智之
2. 発表標題 半解析的ブートストラップ法とその応用
3. 学会等名 第21回情報論的学習理論ワークショップ (IBIS2018)
4. 発表年 2018年

1. 発表者名 小淵智之
2. 発表標題 統計力学的アプローチによるリサンプリング手法の軽量化
3. 学会等名 第30回RAMPシンポジウム (RAMP2018)
4. 発表年 2018年

1. 発表者名 川本達郎, 椿真史, 小淵智之
2. 発表標題 深層学習によるグラフのコミュニティ構造推定II
3. 学会等名 日本物理学会2018年秋季大会
4. 発表年 2018年

1. 発表者名 坂田綾香, 小淵智之
2. 発表標題 非凸正則化付き線形回帰における復号性能
3. 学会等名 日本物理学会2018年秋季大会
4. 発表年 2018年

1. 発表者名 小淵智之, 樺島祥介
2. 発表標題 LASSO+半解析的ブートストラップ法を用いた現実的な変数選択法
3. 学会等名 日本物理学会2018年秋季大会
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------