

令和 4 年 6 月 9 日現在

機関番号：62615

研究種目：若手研究(A)

研究期間：2017～2020

課題番号：17H04676

研究課題名（和文）連続的な対象に対する定数時間アルゴリズム

研究課題名（英文）Constant-Time Algorithms for Continuous Objects

研究代表者

吉田 悠一（Yoshida, Yuichi）

国立情報学研究所・情報学プリンシプル研究系・准教授

研究者番号：50636967

交付決定額（研究期間全体）：（直接経費） 6,900,000円

研究成果の概要（和文）：定数時間アルゴリズムとは、判定問題や最適化問題を入力サイズに依らない定数時間で近似的に解く枠組みである。これまで多くの離散的な対象、例えばグラフや文字列などに対する定数時間アルゴリズムは研究されていたが、連続的な対象、例えば実関数、実数上の行列・テンソル、ユークリッド空間上の確率分布など、に対する定数時間アルゴリズムの研究は限定的であった。本研究課題ではこれらの連続的な対象に取り組み、二次関数最小化、実関数が線形関数かや低次の多項式かの判定、テンソルのタッカー分解、ガウス過程回帰、微分不能な確率分布の確率密度推定を行う定数時間アルゴリズムを構築することに成功した。

研究成果の学術的意義や社会的意義

離散的な対象と同様に連続的な対象の場合も、入力を構造部分と擬似ランダム部分に分け、構造部分に着目することで定数時間アルゴリズムが設計できる場合があることが分かってきた。ただし離散的な対象の時のようにそれが唯一の方法であるとまでは言えていないので、引き続き研究を行っていく必要がある。またテンソル分解や確率密度推定のように実用的に使われている問題に対しても定数時間アルゴリズムが有効に活用できることがわかった。

研究成果の概要（英文）：Constant-time algorithms are algorithms for solving a decision problem or optimization problem in constant time, independent of the input size, in an approximate manner. Although constant-time algorithms have been studied for many discrete objects, such as graphs and strings, research on constant-time algorithms for continuous objects, such as real functions, matrices and tensors over real numbers, and probability distributions in Euclidean space, has been limited. In this research project, I tackled these continuous objects and succeeded in constructing constant-time algorithms for quadratic function minimization, determining whether a real function is a linear function or a polynomial of low degree, Tucker decomposition of tensors, Gaussian process regression, and probability density estimation of nondifferentiable probability distributions.

研究分野：理論計算機科学

キーワード：定数時間アルゴリズム 性質検査 実関数 確率密度推定 テンソル分解

1. 研究開始当初の背景

定数時間アルゴリズムとは、判定問題や最適化問題を入力サイズに依らない定数時間で近似的に解く枠組みである。一見すると定数時間で問題を解くことは不可能に思える。しかし、エラーに関するパラメータ $0 < \varepsilon < 1$ を導入して、入力の ε 倍の誤差を許すことで、多くの問題が ε のみに依存する計算時間で解けることが知られている。ここで入力全体を読み込むと線形時間が掛かってしまうので、一部ずつ入力を読み取れるオラクルが与えられるものとする。定数時間アルゴリズムの研究は盛んに行われており、有限体上の関数や(密)グラフなどの離散的な対象に関する問題であれば、どのような問題が定数時間で解けるのかの特徴付けがほぼ完了している。

2. 研究の目的

定数時間アルゴリズムは入力サイズに依らない計算時間が特徴である。よって実世界に現れる非常に巨大なデータ、いわゆるビッグデータ、を扱う上での特效薬になるように思われる。しかしながら、定数時間アルゴリズムが実世界の問題に適用された例はほぼ無かった。その理由として以下の二つがある。

- 機械学習・信号処理・制御などの実世界に応用を持つ多くの問題が、実数上のベクトルや行列などの連続的な対象を扱っている。しかし定数時間アルゴリズムの研究のほとんどは離散的な対象に関するものである。
- 実世界の(最適化)問題を解くときは、得られた解(= 各変数にどのような値を割り振るか)が重要である。しかし既存の定数時間アルゴリズムのほとんどは目的関数の値しか得られない。

これらの課題を解決すべく、本研究では連続的な対象、具体的には実関数、行列・テンソル、ユークリッド空間上の確率分布などを扱う定数時間アルゴリズムの構築、またこれを通じて

連続的な対象に関する問題が定数時間で解ける為の条件の解明

を目指す。

3. 研究の方法

以下に研究当初に想定していた研究方法について述べる。

線形逆問題に対する定数時間検査アルゴリズム：機械学習・信号処理・統計学においてよく研究されている問題として線形逆問題がある。これは入力に対して、ある線形の演算が適用された結果(観測)が複数得られる時に、もとの入力を復元する問題である。例えば圧縮センシングは、疎な信号ベクトル $x \in \mathbb{R}^n$ に対して、ランダムなベクトル $a_1, \dots, a_k \in \mathbb{R}^n$ との内積 $o_1 = a_1^T x, \dots, o_k = a_k^T x$ が観測されるとき、ベクトル x を復元する問題である。他にも行列補完、辞書学習などの重要な問題が線形逆問題で表せる(図1)。

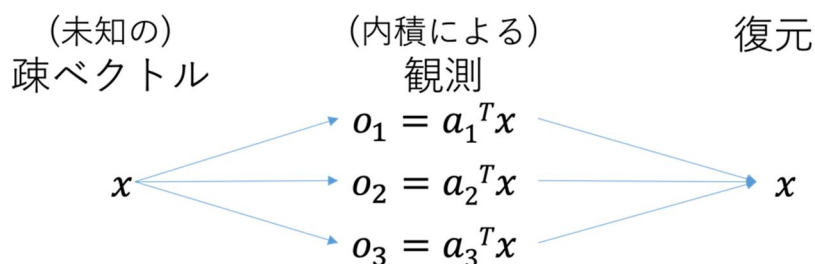


図 1: 圧縮センシング

本研究では、線形逆問題に対する定数時間検査アルゴリズムを構築する。ここで検査とは「ある性質を満たす」か「満たすには入力の割合を書き換える必要がある」かを高い確率で区別することである。圧縮センシングを例にとると「少ない観測で信号ベクトル x が疎かを検査」或いは「与えられた観測ベクトル o が正しい観測ベクトルかの検査」という二つの問題が考えられる。観測の数は少なくければ少ないほど良く、検査という枠組みにより、必要な観測数が定数になることが期待される。

前述の通り有限体上の関数やグラフなどの離散的な対象に対する問題であれば、定数時間で解ける問題の特徴付け、即ち必要十分条件がほぼ解明されている。簡単に述べると、まずどのような入力も構造化された部分(低次の多項式、二部グラフなど) + 疑似ランダム部分に分解することができる。この時、構造部分のみで決まるような判定問題や最適化問題であれば定数時間で解くことができ、そうでなければ解くことが出来ない。

連続最適化問題に対しては、[Hayashi and Yoshida, Advances in Neural Information

Processing Systems 29 (NIPS 2016)] の様に定数時間で解ける例が知られているが、それが何に根ざしたものは分かっていない。また連続的な対象に対する上記の様な分解定理も知られていない。連続的な対象に対する理解を推し進めることが本研究の最終目標である。

4. 研究成果

行列・テンソルに対する定数時間アルゴリズム

まず、テンソルのタッカー分解に対する定数時間アルゴリズムを構築した。このアルゴリズムはランクと呼ばれるパラメータを指定すると、そのランクに基づいてテンソルをタッカー分解した際の誤差を定数時間で近似する(分解に用いる行列・テンソル自体は求めない)。本アルゴリズムの一つの応用に「良いランクを高速に決める」というものがある。ランクの選び方によってその後の応用での性能が大きく変わるが、選んだランクが良いかどうかは実際にタッカー分解してみないと分からない。しかしもともとタッカー分解は実行に時間がかかるので、何度もランクを変えてタッカー分解を行うのは避けたい。しかし本アルゴリズムを用いれば、各ランクに対して誤差が定数時間で求まる為、良いランクを素早く決めることが出来る。この成果は The 31st Annual Conference on Neural Information Processing Systems (NIPS 2017) に採択され、口頭発表が認められた。

また定数時間での二次関数最小化に取り組んだ。以前の研究成果 [Hayashi and Yoshida, Advances in Neural Information Processing Systems 29 (NIPS 2016)] において、既に定数時間での二次関数最小化の研究は行なっているが、今回の研究では球面制約に対応することができる。これにより非線形最適化における重要なアルゴリズムである信頼領域アルゴリズムの一要素である信頼領域問題を高速に近似することができるようになる。解析はグラフ極限理論と新たに証明した行列のスペクトル分解を組み合わせることで行った。具体的には、任意の $n \times n$ 行列 A は構造部分と擬似ランダム部分に次のように分けることができる：構造部分は $\text{polylog}(n)$ 個のブロックからなる定数ブロック行列で、擬似ランダム部分はスペクトラルノルムが小さい。この分解は Frieze と Kannan (FOCS'16) による既存の分解よりも良いエラー評価を与える。またこの分解を使うと行列の上位の特異値も近似することができる。この結果は Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018) に採択された。

機械学習・統計の問題に対する定数時間アルゴリズム

次に機械学習の基本的な問題である回帰の典型手法であるガウス過程回帰に対する定数時間アルゴリズムの解析を行い、その成果が 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020) に採択された。ガウス過程回帰は正確に解くには逆行列を計算する必要があり計算量が大きい。そこで計算量を改善するための様々な手法が提案されている。本研究では特にサブサンプリングと呼ばれる計算量を定数にする手法について、精度の理論的評価を行った。

次に、確率分布の確率密度が微分不可能な場合に確率密度推定を行うアルゴリズムを構築した。これまでのアルゴリズムはすべて確率密度が微分可能であることを仮定していた。このアルゴリズムの手法と解析には、グラフ理論の道具であるセメレディの正則性補題を用いている。また得られたアルゴリズムの収束レート(確率分布からのサンプル数と確率密度の推定誤差の関係)がタイトであることを示した。この成果は、The 21st International Conference on Artificial Intelligence and Statistics (AISTATS 2018) に採択されベストペーパーを受賞した。

実関数の性質検査

本研究では実関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ に対する性質検査について研究を行っている。まず性質 P (d 次多項式かなど) と \mathbb{R}^n 上の分布 D を定める。ここで D の具体的な形は未知であるが、 D から点がサンプリングできると仮定する。このとき、 f が P を満たすか、満たすには ϵ -far、すなわち任意の P を満たす関数 g に対して、 $\Pr_{x \sim D}[f(x) \neq g(x)] > \epsilon$ かを区別するアルゴリズムを作ることが目標となる。特に f にクエリする回数を n に依存しない定数にしたい。本研究では D が product distribution であるときに検査アルゴリズムを構築する一般的な手法としてダウンサンプリングを提案した。これは各軸に対して複数の点を D の周辺分布からサンプリングし、それらの点の組み合わせで表される格子点に対して f の値を問い合わせ、元の問題を格子上で定義される関数の検査に帰着するというものである。これにより非常に多数の性質についての結果を同時に得ることができた。この結果は理論計算機科学のトップ会議の一つである The 49th EATCS International Colloquium on Automata, Languages and Programming (ICALP'22) に採択されている。

次に実関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ が線形関数かどうかを検査する定数時間アルゴリズムの構築に取り組み、その成果が 11th Innovations in Theoretical Computer Science (ITCS 2020) に採択された。線形性の検査は有限体上の一様分布に対してよく研究されてきたが、 \mathbb{R}^n 上では一様分布は定義することができない。そこで最初は n 次元のガウシアンを用いて距離を定義し、その上での検査アルゴリズムを与えた。その後、このアルゴリズムをサブルーチンとして用いることで、任意の \mathbb{R}^n 上の分布を用いた距離に対するアルゴリズムに拡張できることを示した。

次に P が d 次多項式で、 D が一般の場合に対する $\text{poly}(d, 1/\varepsilon)$ クエリ検査アルゴリズムを構築した。これまでは定義域が \mathbb{F}_p^n の様な有限の場合でしか検査アルゴリズムが知られていなかったが、これを初めて \mathbb{R}^n に拡張することに成功した。これまでの検査アルゴリズムは x, y を \mathbb{F}_p^n から一様サンプリングした際に、 $x + y$ も \mathbb{F}_p^n 上で一様に分布することを利用していたが、 \mathbb{R}^n ではそのような性質が無いため非常に込み入った議論が必要になっている。

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 7件/うち国際共著 2件/うちオープンアクセス 5件）

1. 著者名 Noah Fleming and Yuichi Yoshida	4. 巻 なし
2. 論文標題 Distribution-Free Testing of Linear Functions on R^n	5. 発行年 2020年
3. 雑誌名 Proceedings of the 11th Innovations in Theoretical Computer Science (ITCS)	6. 最初と最後の頁 22:1--22:19
掲載論文のDOI (デジタルオブジェクト識別子) 10.4230/LIPIcs.ITCS.2020.22	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 該当する
1. 著者名 Kohei Hayashi, Msaaki Imaizumi, Yuichi Yoshida	4. 巻 なし
2. 論文標題 On Random Subsampling of Gaussian Process Regression: A Graphon-Based Analysis	5. 発行年 2020年
3. 雑誌名 Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)	6. 最初と最後の頁 -
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 Amit Levi and Yuichi Yoshida	4. 巻 116
2. 論文標題 Sublinear-Time Quadratic Minimization via Spectral Decomposition of Matrices	5. 発行年 2018年
3. 雑誌名 Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)	6. 最初と最後の頁 17:1--17:19
掲載論文のDOI (デジタルオブジェクト識別子) 10.4230/LIPIcs.APPROX-RANDOM.2018.17	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 該当する
1. 著者名 Yoichi Iwata, Yutaro Yamaguchi, and Yuichi Yoshida	4. 巻 無し
2. 論文標題 0/1/all CSPs, Half-Integral A-path Packing, and Linear-Time FPT Algorithms	5. 発行年 2018年
3. 雑誌名 Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS)	6. 最初と最後の頁 462--473
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/FOCS.2018.00051	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 Kohei Hayashi and Yuichi Yoshida	4. 巻 30
2. 論文標題 Fitting Low-Rank Tensors in Constant Time	5. 発行年 2017年
3. 雑誌名 Advances in Neural Information Processing Systems 30 (NIPS 2017)	6. 最初と最後の頁 2473 ~ 2481
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 Iwama Kazuo and Yoshida Yuichi	4. 巻 9
2. 論文標題 Parameterized Testability	5. 発行年 2017年
3. 雑誌名 ACM Transactions on Computation Theory	6. 最初と最後の頁 1 ~ 16
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3155294	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 Masaaki Imaizumi, Takanori Maehara, and Yuichi Yoshida	4. 巻 84
2. 論文標題 Statistically Efficient Estimation for Non-Smooth Probability Densities	5. 発行年 2018年
3. 雑誌名 Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS)	6. 最初と最後の頁 978 ~ 987
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1件 (うち招待講演 1件 / うち国際学会 0件)

1. 発表者名 吉田悠一
2. 発表標題 連続最適化問題に対する定数時間アルゴリズム
3. 学会等名 第21回情報論的学習理論ワークショップ (招待講演)
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
---------	---------