

# 有限長解析情報理論と最適化理論による高信頼高効率通信の相乗的基礎研究

A synergistic fundamental study of high reliability and high efficiency communication based on non-asymptotic information theory and optimization



松嶋 敏泰 (Toshiyasu MATSUSHIMA, Dr.Eng.)  
早稲田大学 基幹理工学部 応用数理学科 教授  
(Professor, Waseda University, School of Fundamental Science and Engineering, Department of Applied Mathematics)  
電子情報通信学会 IEEE 他  
研究専門分野：データ科学、情報理論、符号理論、暗号理論

## あらまし

本研究の目的は、通信の基礎理論として大きな成功を収めてきた情報理論・符号理論をさらに推し進め、有限の長さのデータ系列や符号長が有限の場合など、より実用に則した条件のもとでの通信システムの性能の理論限界を導出し、その理論限界を達成する符号、復号法の組を構成することである。その際、符号化・復号システム全体を最適化理論の枠組みに落とし込む点に特色がある。結果として、符号化レートや誤り確率の理論限界を、実用的なデータ長や実用的な誤り確率を許したもつで導出することができた。また、符号化・復号システムにおいて一部を最適化問題とみなした際の目的関数やヒューリスティクスとして用いられうる量の整理、拡張を行うことができた。

## 1. 研究の背景

情報理論・符号理論は、通信の数理的基礎理論としてのみならず、情報を扱う科学及び工学の基礎理論として現代の情報化社会の発展に寄与してきた。情報理論・符号理論では、雑音のある通信路を通して送信者から受信者に情報を効率的に高信頼度で送る問題を数理モデル化して研究を行う。通信の効率性を測る評価基準として符号化レート<sup>\*1</sup>が、信頼性を測る評価基準として復号誤り確率<sup>\*2</sup>が用いられる。両者はトレードオフの関係があり、伝統的に以下の(アプローチ a-0)、(アプローチ b-0)により研究が行われてきた。

(アプローチ a-0) ある理想的な状況を仮定したもつで、符号化レートと誤り確率の一方を制約条件として、他方を最適化することにより理論限界を導出する。

(アプローチ b-0) 上述した(アプローチ a-0)で示された理論限界を達成する符号器(符号化関数)や復号器(復号関数)を構成する。

(アプローチ a-0)により理論限界を示す事自体、情報を科学する意味で意味があり、また(アプローチ b-0)を行う上でも理想的目標が掲げられる意味は大きい。しかしながら、これまで(アプローチ a-0)と(アプローチ b-0)はほとんど独立に発展し、融合することはほとんどなかった。これに対して、近年、以下の(アプローチ a-1)、(アプローチ a-1')、(アプローチ b-1)のような、新たなアプローチによる研究が行われるようになってきた。

(アプローチ a-1) 有限のデータ長や微小な誤り確率を許したもつで、符号化レートや誤り確率の理論限界を導出する。

(アプローチ a-1') 符号と復号のクラスを固定したもつで、データの長さが限りなく大きくなることを許した際の平均誤り確率を導出する。

(アプローチ b-1) 最適化理論を用いた復号アルゴリズムの構築や符号の探索による構成を行う。

## 2. 研究の目的

従来の伝統的なアプローチである(アプローチ a-0)と(アプローチ b-0)の関係に比べれば、(アプローチ a-1)、(アプローチ a-1')、(アプローチ b-1)の研究アプローチは接近してきたと言える。(アプローチ a-0)と違い、(アプローチ a-1)は実際に(アプローチ b-1)で到達すべき工学的目標となりうる。また、(アプローチ a-1')は(アプローチ b-1)の符号のクラスの設定や符号の探索に直接役立ち、2つのアプローチが融合してきている。

そこで本研究では、この流れをさらに加速させ、以下の(アプローチ a-2)と(アプローチ b-2)により、理論

# 有限長解析情報理論と最適化理論による高信頼高効率通信の相乗的基礎研究

A synergistic fundamental study of high reliability and high efficiency communication based on non-asymptotic information theory and optimization

限界の導出と符号・復号法の構築の研究をさらに接近・融合させ、相乗的研究を目指すことを目的とした。

(アプローチ a-2) 符号化レートや誤り確率の理論限界を、実用的なデータ長や実用的な誤り確率を許して導出する。

(アプローチ b-2) 符号化復号システム全体を大きな最適化問題として定式化し、より実用的な制約のもと(準)最適な符号と復号の組として求める。

例えば、(アプローチ a-2)で導出された実用に近い条件での理論限界は、(アプローチ b-2)の最適化による符号・復号法の構成の直接的な目標となる上、ヒューリスティック関数<sup>\*3</sup>として、アルゴリズムの高速化や大局的最適性の保証に役立つ。この研究により、より工学的見地から有用な高効率・高信頼度の通信や情報処理システムの設計に直接的に役立つ基礎理論の構築を目指した。

## 3. 研究の方法、研究の結果

### 3. 1. 研究の進め方

従来、(アプローチ a-1)による研究(例えば、[1]、[2])では、確率論における中心極限定理等を用いて、理論限界の解析が行われている。しかしながら、データが独立同分布に従うという理想的な仮定や、数値的に求めにくい限界が導出されている。

したがって、(アプローチ a-2)の研究では、実用に近い仮定のもとでの理論限界の導出や、より精密に数値的に誤り確率等が導出できる解析手法を構築することを目指した。また、通信路容量<sup>\*4</sup>等、最大、最小を求める最適化問題として解かなければ数値が得られない問題もあり、このような問題を解く最適化法も明らかにすることを目指した。

一方で、従来、(アプローチ b-1)による研究(例えば、[3]、[4]、[5]、[6])では復号問題、符号構成問題を制約付き最適化問題とみなして研究が行われている。しかしながら、それらは復号問題、符号構成問題をそれぞれ単体で最適化問題とみなすにとどまっている。

そこで、(アプローチ b-2)の研究では、符号化・復号

システム全体を大きな最適化問題として捉え、(アプローチ a-2)によって求められた実用的な制約のもとでの理論限界を目標とした大局的最適化を行うことを目指した。もちろん、今までの重要な成果である、高性能な代数的復号法や確率伝搬法などはアルゴリズムの中に組み入れ、高性能な符号の構造や性能が良いとされる符号のクラスも積極的に符号の探索で用いていく。

### 3. 2. 現在までに得られた研究成果

現在までに、(アプローチ a-2)に関しては以下の成果が得られた。

( $\alpha$ -1) 情報理論においては、情報源から確率的にデータが出力されると仮定したもとで議論が行われる。この情報源から出力されたデータ $x$ と、復号後のデータ $y$ との間に違いを許容することで、データ $x$ と $y$ が完全に一致する場合に比べて圧縮の効率を高めることが可能である。このようなデータ圧縮は、有歪みデータ圧縮と呼ばれる。例えば、画像や音声のデータは、もとのデータと復号後のデータが多少異なっても問題ない場合が多いため、これらのデータの圧縮には有歪みデータ圧縮の研究成果が用いられる。データ圧縮の効率を測る代表的な評価基準として平均符号語長(情報源の確率分布により符号語長の期待値を取った量)があり、これを一般化した尺度として符号語長のキュムラント母関数がある。本研究では、有限の実用的なデータ長に対する符号語長のキュムラント母関数の理論限界を、レニーエントロピーに基づいた情報量により明らかにした[12]。

( $\alpha$ -2) データ圧縮の研究において、平均符号語長以外にも幾つかの評価基準が考えられている。その中の一つに、オーバーフロー確率がある。これは、データ1シンボル当たりの符号語長が、あるレート(しきい値)を超過する確率として定義される。本研究では、有歪みデータ圧縮問題に対して、オーバーフロー確率を微小量以下にしたもとでの最小のレートが、有限の実用的なデータ長に対して、スムーズ最大エントロピーという情報量を用いて特徴付けられることを明らかにした[9]、[13]。

# 有限長解析情報理論と最適化理論による高信頼高効率通信の相乗的基礎研究

A synergistic fundamental study of high reliability and high efficiency communication based on non-asymptotic information theory and optimization

( $\alpha$ -3) 情報理論において、データ圧縮と密接に関係のある問題として **guessing** がある。これは、確率変数列の実現値を、なるべく少ない回数の質問により推測する問題である。この問題に対して、有限の系列長の確率変数を推定する場合、最小の推定回数の理論限界が、レニーエントロピーに基づいた情報量により特徴づけられることを解明した[16]。

( $\alpha$ -4) 情報源の確率分布が未知である場合のデータ圧縮をユニバーサルデータ圧縮と呼ぶ。ユニバーサルデータ圧縮において、有限のデータ長に対して圧縮率の高いデータ圧縮法として、ベイズ符号がある。ベイズ符号は平均符号語長という評価基準のもとでは多くの研究がなされてきた。しかしながら、オーバーフロー確率という評価基準のもとでは、これまで殆ど研究がなされていなかった。本研究では、**moderate deviation** の枠組み<sup>\*5</sup>において、ベイズ符号をオーバーフロー確率という評価基準のもとで理論解析を行った。その結果、ベイズ符号は、オーバーフロー確率が最小という意味で最適な符号と同様の性質を有することを明らかにした[7]。

一方で、(アプローチ b-2)に関しては以下の成果が得られた。

( $\beta$ -1) 符号構成の目的関数になりうる量として、重み分布という量がある。従来、我々は富士山型空間結合符号[8]と呼ばれる新たな符号クラスを提案していたが、この符号に対して重み分布を解明し、**Expected Graph Evolution** や **Covariance Evolution** による評価を行った[10], [11], [14], [15]。

( $\beta$ -2) 様々な最適化問題とその解法アルゴリズムについての従来研究の整理を行った。また、現状の符号化復号システムにおいてシステムの一部を最適化問題とみなした際の目的関数やヒューリスティクスとして用いられうる量の整理、拡張を行った[17]。

### 3. 3. 本研究の意義及び研究者自身の見解

本研究により、より工学的見地から高効率・高信頼

度の通信や情報処理システムの設計に直接的に役立つような基礎理論を、ある程度打ち立てることができたと考えている。

### 4. 将来展望

本研究で得られた基礎理論は今後、高信頼度かつ高効率の通信や処理が要求されるような通信工学関連分野において、システムの設計や構築に寄与することが期待される。

### 用語解説

- \*1 符号化レート・・・情報源から発生する $k$ シンボルの情報系列を符号器(符号化関数)により $n$ シンボルの符号語に変換するとき、符号化レートは $k/n$ と定義される。
- \*2 復号誤り確率・・・雑音等の影響で符号語が確率的に変化した $n$ シンボルの受信語は、復号器(復号関数)により情報系列に復号される。この際に、元の情報系列と異なる場合に復号誤りとなり、そのような事象が生起する確率が復号誤り確率である。
- \*3 ヒューリスティック関数・・・経験的な知識をもとに構成した評価関数のこと。
- \*4 通信路容量・・・符号長 $n$ を十分大きくしたとき、復号誤り確率を漸近的に0にするという条件のもとで達成可能な符号化レート $k/n$ の上限を通信路容量と呼ぶ。
- \*5 **moderate deviation** の枠組み・・・オーバーフロー確率とレート(しきい値)はどちらも小さい方が望ましいが、これらの量の間にはトレードオフの関係がある。そのため、オーバーフロー確率とレートのトレードオフをコントロールするパラメータを導入し、これを上手く設定することで、オーバーフロー確率とレートのどちらもある速度で小さくなっていくという問題設定で解析を行うことがある。このような問題設定は、情報理論の研究において、**moderate deviation** の枠組みと呼ばれる。

### 参考文献

- [1] Shota Saito, Nozomi Miya, Toshiyasu Matsushima, Evaluation of the Bayes Code from

# 有限長解析情報理論と最適化理論による高信頼高効率通信の相乗的基礎研究

A synergistic fundamental study of high reliability and high efficiency communication based on non-asymptotic information theory and optimization

- Viewpoints of the Distribution of Its Codeword Lengths, *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E98-A, no. 12, pp. 2407-2414, 2015.
- [2] Y. Polyanskiy, H. V. Poor, and S. Verdú, Channel Coding Rate in the Finite Blocklength Regime, *IEEE Trans. Inf. Theory*, vol.56, no.5, pp.2307-2359, 2010.
- [3] Shunsuke Horii, Toshiyasu Matsushima, Shigeichi Hirasawa, Linear Programming Decoding of Binary Linear Codes for Symbol-Pair Read Channels, *Proceedings of 2016*
- [4] Shunsuke Horii, Toshiyasu Matsushima, Shigeichi Hirasawa, Linear Programming Decoding of Binary Linear Codes for Symbol-Pair Read Channels, *Proceedings of 2016*
- [5] 堀井俊佑, 松嶋敏泰, 平澤茂一, シンボルペア通信路における 2 元線形符号の線形計画復号法について, *電子情報通信学会技術研究報告*, vol.115, no.214, IT2015-34, pp.1-6, 2015.
- [6] E. Sharon and S. Litsyn, Constructing LDPC Codes by Error Minimization Progressive Edge Growth, *IEEE Trans. Commun.* vol.56, no.3, pp.359-368, 2008.
- [7] Shota Saito, Toshiyasu Matsushima, Evaluation of overflow probability of Bayes code in moderate deviation regime, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E100A(12), pp.2728 - 2731, 2017.
- [8] Yuta Nakahara, Shota Saito, Toshiyasu Matsushima, Spatially "Mt. Fuji" Coupled LDPC codes, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E100A(12), pp.2594 - 2606, 2017.
- [9] Shota Saito, Hideki Yagi, Toshiyasu Matsushima, Variable-Length Lossy Compression Allowing Positive Overflow and Excess Distortion Probabilities, *IEEE International Symposium on Information Theory 2017*
- [10] 中原悠太, 松嶋敏泰, 富士山型空間結合符号の重み分布に関する一考察, 第 40 回情報理論とその応用シンポジウム, 2017.
- [11] Yuta Nakahara, Toshiyasu Matsushima, A Note on Weight Distributions of Spatially "Mt. Fuji" Coupled LDPC Codes, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E101.A(12), pp.2194 - 2198, 2018.
- [12] Shota Saito, Toshiyasu Matsushima, Cumulant Generating Function of Codeword Lengths in Variable-Length Lossy Compression Allowing Positive Excess Distortion Probability, *IEEE International Symposium on Information Theory*, 2018.
- [13] Shota Saito, Hideki Yagi, Toshiyasu Matsushima, New Results on Variable-Length Lossy Compression Allowing Positive Overflow and Excess Distortion Probabilities, *International Symposium on Information Theory and Its Applications*, 2018.
- [14] Yuta Nakahara, Toshiyasu Matsushima, Expected Graph Evolution for Spatially "Mt. Fuji" Coupled LDPC Codes, *International Symposium on Information Theory and Its Applications*, 2018.
- [15] 中原悠太, 松嶋敏泰, 富士山型空間結合符号に対する Covariance Evolution, 第 41 回情報理論とその応用シンポジウム, 2018.
- [16] Shota Saito, Toshiyasu Matsushima, Non-Asymptotic Fundamental Limits of Guessing Subject to Distortion, *IEEE International Symposium on Information Theory*, 2019.
- [17] 中原悠太, 松嶋敏泰, Spatially Coupled Uneven LDPC Codes, 第 42 回情報理論とその応用シンポジウム予稿集, 2019.
- この研究は、平成 28 年度 S C A T 研究助成の対象として採用され、平成 29 年～令和元年度に実施されたものです。