



C*環を用いたカーネル法の拡張

1. C*環を用いたデータ解析の概要

NTT / RIKEN AIP

橋本悠香

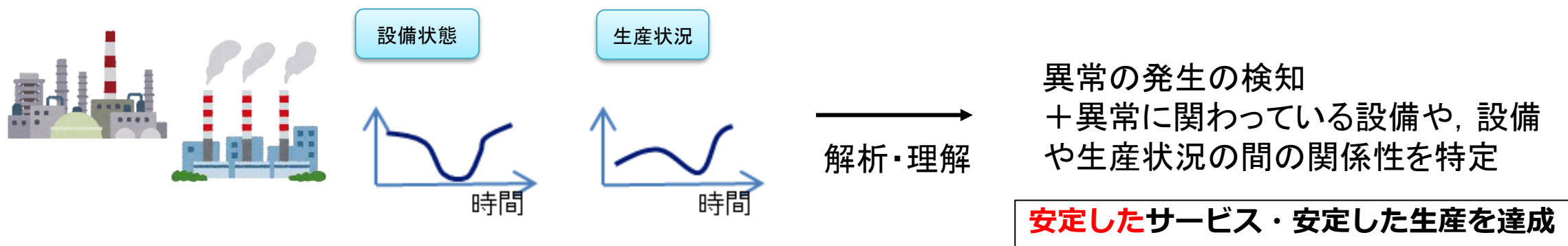
■ 橋本悠香 (NTT)

- 2016-2018 慶應義塾大学理工学研究科(数理科学科) 修士課程
- 2018-2023 NTTネットワークサービスシステム研究所 研究員
- 2019-2022 慶應義塾大学理工学研究科 博士課程(社会人博士)
- 2022- 理研AIP 客員研究員
- 2023- NTTネットワークサービスシステム研究所(NTTコミュニケーション科学基礎研究所) 特別研究員
- 2024- 金沢大学 客員准教授

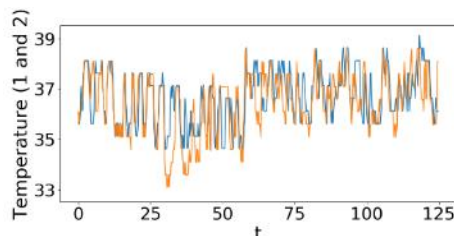
■ 研究の興味

- 作用素論・作用素環論(C^* 環)を, 世の中の現象を理解することへの応用
- 数値線形代数 (大規模な計算をどのようにコンピュータで処理するか)

- 世の中は、複雑なデータ(手に入れることができる事実・情報)で溢れている。データを理解する(解析する)ことで、産業の発展や自然法則の発見につながる。



生体データ(人間・動物の体温・血圧・呼吸の状態など)



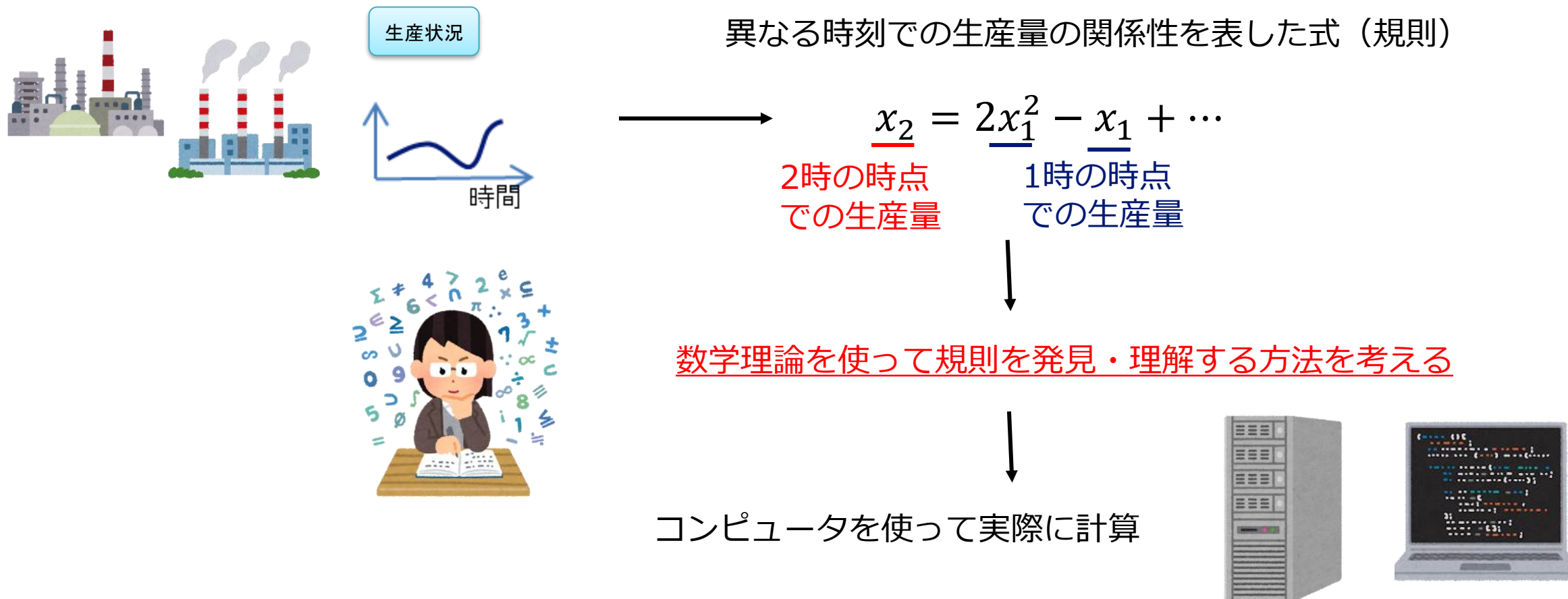
解析・理解

- ・病気の発見
+異常のある機能や, 機能の間の関係性を特定
- ・生物同士の同期の有無を発見
+同期の強弱や時間変化を特定

新たな自然界の法則を発見

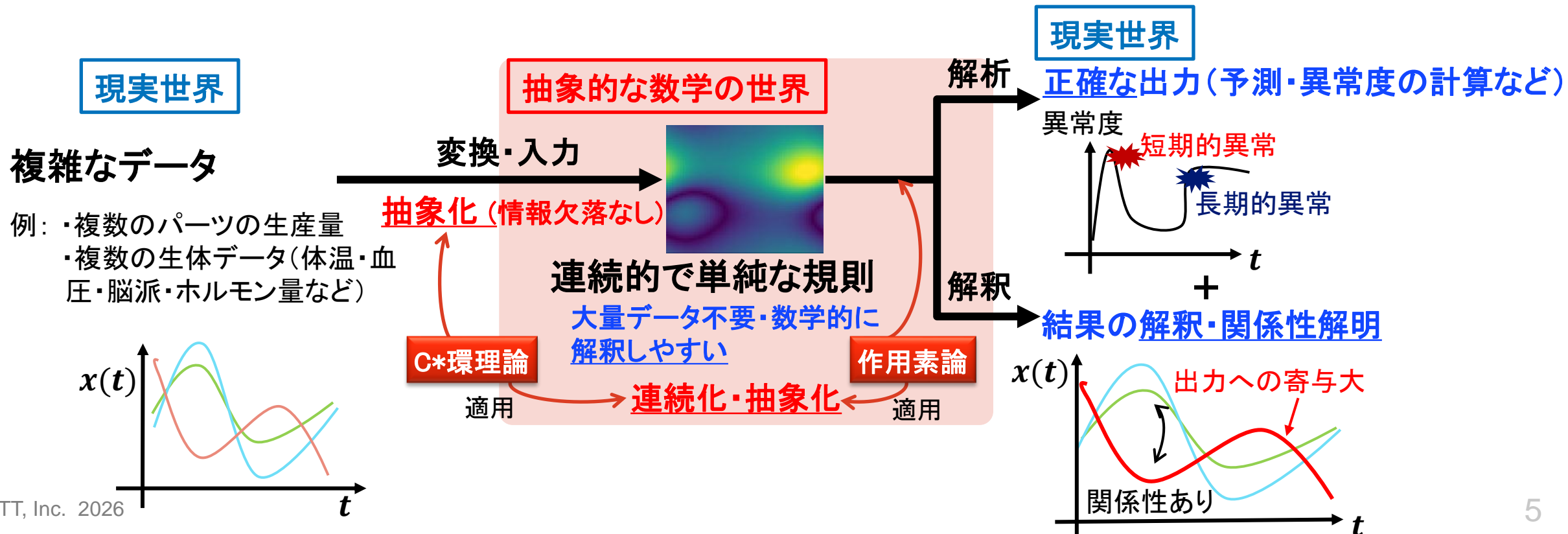
データ解析への数学の応用

- データの間には、関係性がある。関係性を数式で表現し、数学理論を使って数式の性質を調べることで、データを解析し、理解することが可能になる。

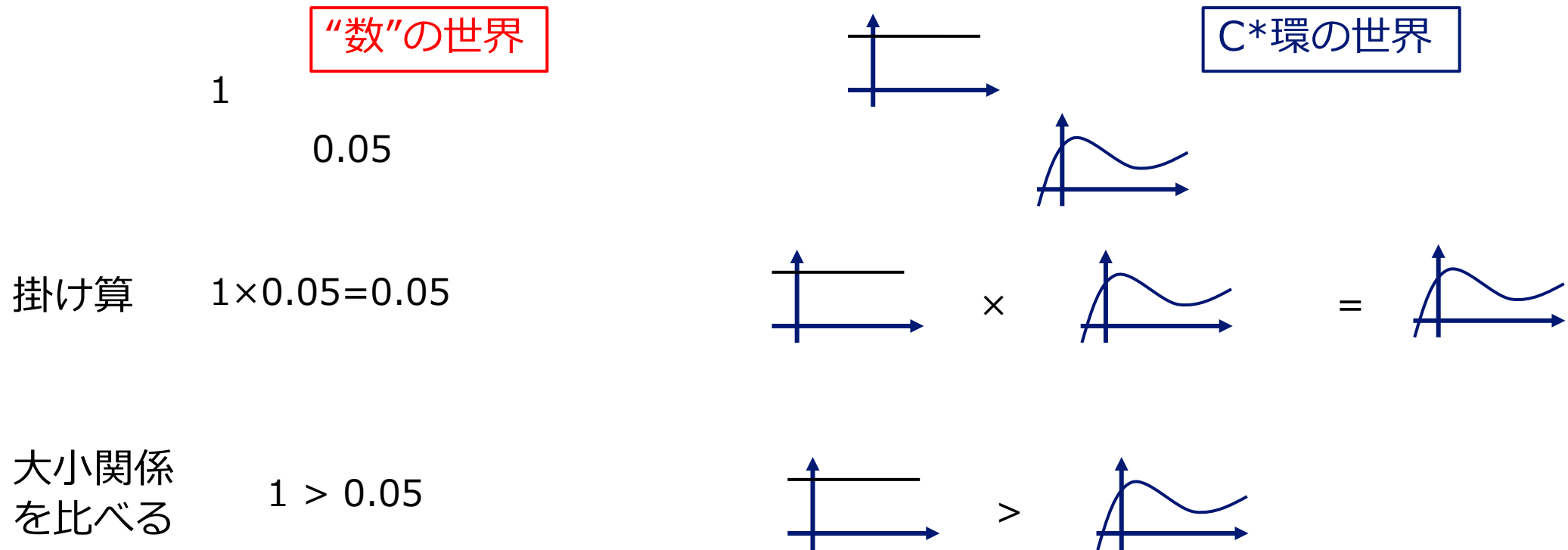


作用素論・C*環理論を用いたデータ解析

- 自然界に存在するデータは、**ノイズ**を含む、時間に関して**連続的な振る舞い**をする、**相互作用がある**、など、複雑であり、必ずしも**十分な量のデータ**が得られるとは限らない。
- これまでは学術的研究が中心であったC*環・作用素論などの**高度な数学理論**を応用し、モデルやデータを**連続化・抽象化**して**単純化**する新しいデータ解析の枠組みを提案する。これにより、**少量で複雑なデータ**でも正確に解析し、さらに**解釈性**や**性能**を理論的にも保証することを目指す。



- 実数・複素数を一般化する数学の概念. 掛け算・複素共役・大小関係を比べるなどの基本的な性質を持っている. (例: 関数, 行列)



■ データ解析でよく考えられる問題:

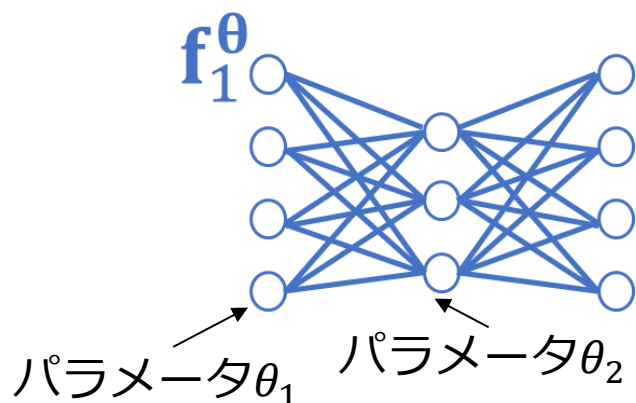
何かを何かに移す, 最も良い規則を見つける(最適化問題と呼ばれる)

例: x_1 (1時の時点での生産量) から x_2 (2時の時点での生産量) を予測する

$$x_2 = 2x_1^2 - x_1 + \dots$$

x_1 を x_2 に移す規則 $f(x_1)$

数学的な定式化: $\min_{f \in \mathcal{F}} \sum_{i=1}^n |f(x_i) - x_{i+1}|$



多くの場合, 規則はパラメータの集合とそれらの関係性で表現される.

x_1 や x_2 が実数や複素数である場合は, 十分に研究されている

C*環を用いたデータ解析1：統合

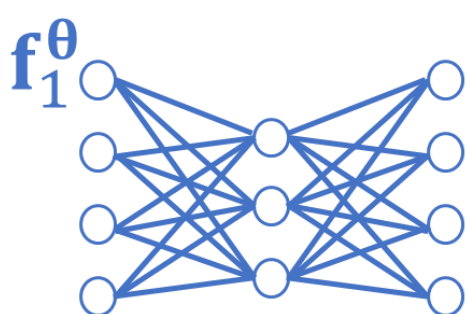
- これまで”数“として考えられていたデータを，C*環に一般化．これにより，複数のデータや複数の規則をまとめて扱うことが可能になり，データからより多くの情報を引き出せる．

複数のデータをまとめて扱う（入出力を関数とみなす）

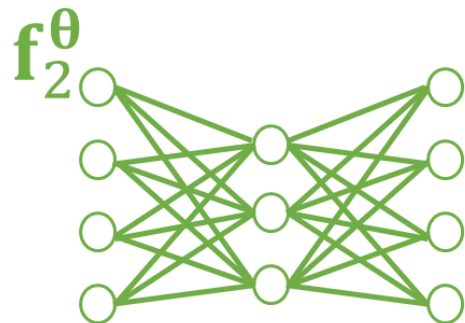
例： x_1 （1時の時点でのA工場・B工場・・・の生産量）から x_2 （2時の時点でのA工場・B工場・・・の生産量）を予測する

$$x_2 = 2x_1^2 - x_1 + \dots$$

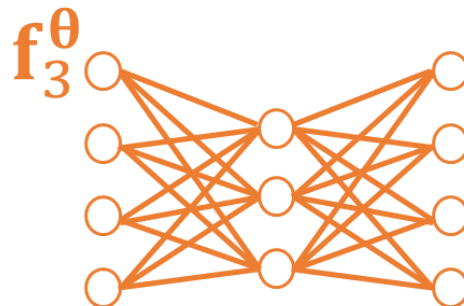
複数の規則をまとめて扱う（各パラメータを関数とみなす）



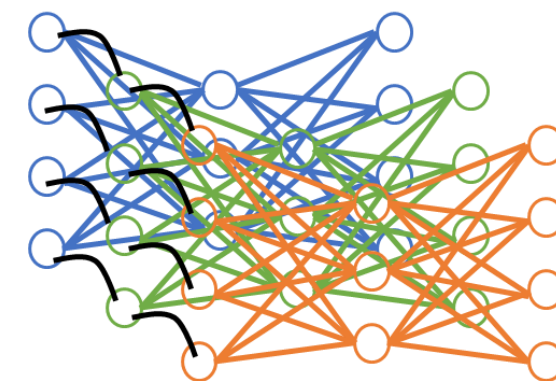
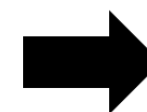
規則1



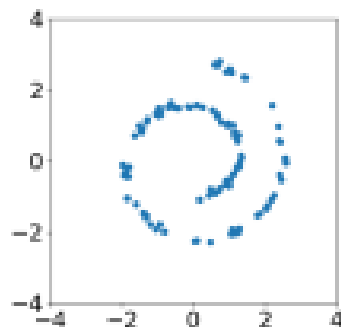
規則2



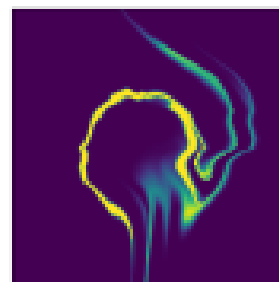
規則3



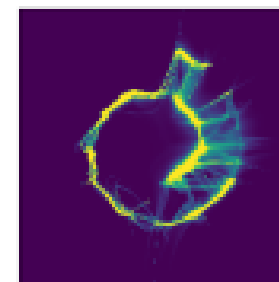
- データ(点)がどこにどのくらいあるかを予測するタスクを解いてみた結果. 色が濃いとところほど, たくさんのデータがあると予測されている.



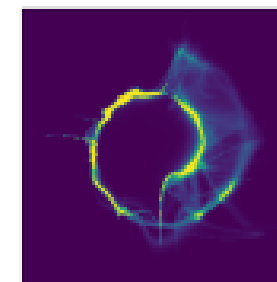
データ



C*環を使わない方法1



C*環を使わない方法2



C*環を使った方法 (複数の規則をまとめて扱う)

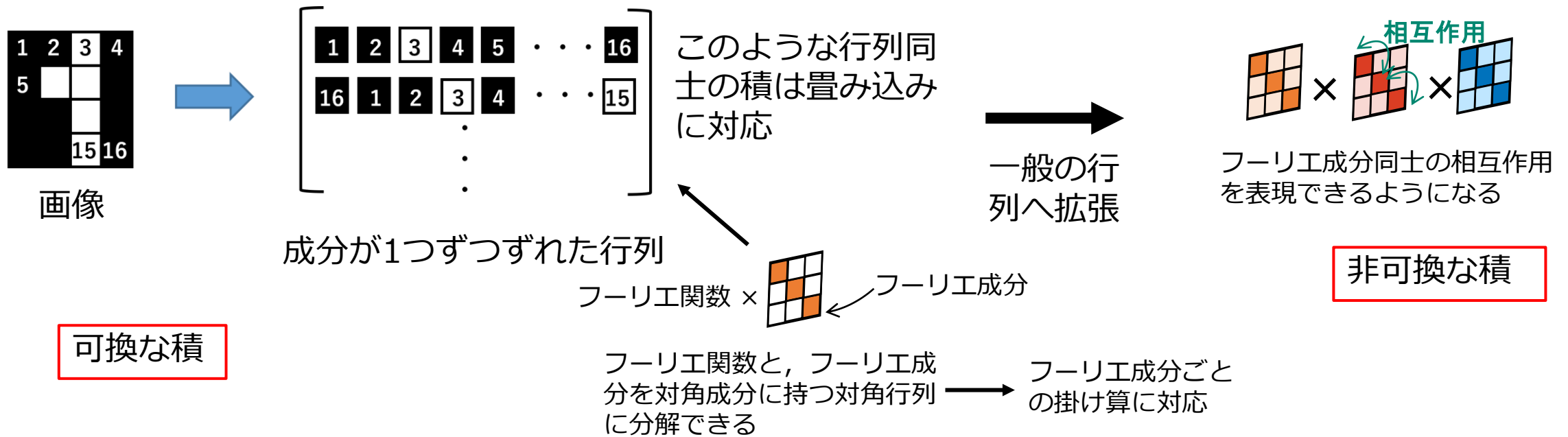
C*環を用いた方法が, 最もうまく予測できている

C*環を用いたデータ解析2：新たな演算

- 畳み込みと呼ばれる既存の演算を，行列演算を用いて一般化．これにより，フーリエ成分同士の関係性に関する情報を引き出せる．

畳み込み：変化の大きさごとにデータを分解し，分解した成分（フーリエ成分）同士の演算をする

異なるフーリエ成分同士の相互作用



- 手書き文字(0~9)の画像にノイズ(不要な情報)を加えた画像を与えた時に, それを元の画像に復元するタスクを解いた結果.



元の画像



ノイズを加えた画像



C*環を使わない方法で復元した画像

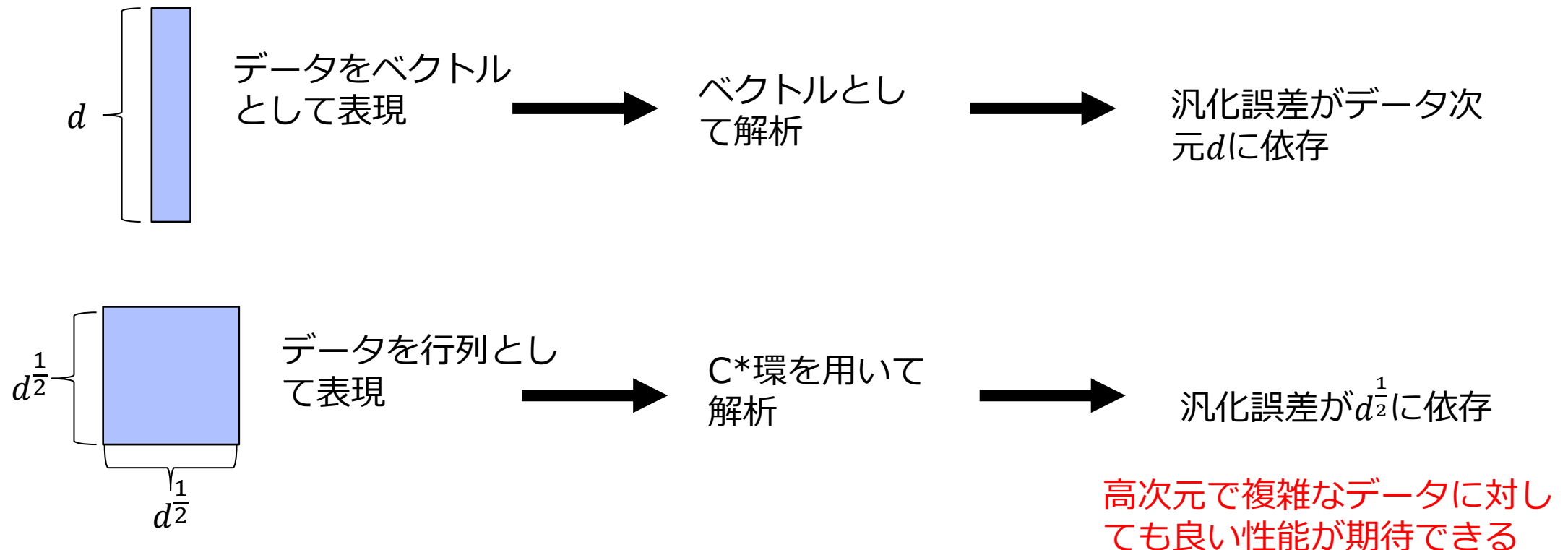


C*環を使った方法で復元した画像

C*環を用いた方法のほうが, きれいに復元できている

C*環を用いたデータ解析3：理論評価の改善

- 汎化誤差(導かれた規則の性能を表す指標)のデータ次元への依存性が既存の結果に比べて削減できることを理論的に証明.



C*環を用いたデータ解析まとめ

- C*環を用いることで、複雑なデータや規則を抽象的に表現でき、理論解析により精度向上を示すことも可能となる。

