

派生研究を支援する再構成容易な実験数学環境

横山 重俊^{1,3,a)} 浜元 信州¹ 長久 勝² 藤原一毅³ 政谷 好伸³ 竹房 あつ子³ 合田 憲人³

概要：データ駆動型科学研究分野で研究データの再利用を進めるためには、研究データの公開だけでなく、論文成果につながる実験結果にたどり着くまでの実験プロセス（実験手順や実験環境構築手順）を公開し共有する必要がある。これらが揃うことで、第三者である別の研究者がいつでも研究データを再利用し、研究の再現検証が可能となる。さらに、そのオリジナル研究に加えて自らの研究を派生させることがスムーズにできるようになる。本稿では、特に数学分野に焦点をあて、その派生研究の支援として、派生研究のために再構成容易な実験数学環境の実装方法と具体的な派生研究例への適用について報告する。

キーワード：研究環境再構成, 実験数学, 計算機実験, Jupyter Notebook, クラウドコンピューティング

A Mathematical Experiment Environment that Supports Derivative Researches

SHIGETOSHI YOKOYAMA^{1,3,a)} NOBUKUNI HAMAMOTO¹ MASARU NAGAKU² IKKI FUJIWARA³
YOSHINOBU MASATANI³ ATSUKO TAKEFUSA³ KENTO AIDA³

Abstract: In order to promote the reuse of research data in the field of data-driven scientific research, not only the publication of research data but also the experimental process (experimental procedure and experimental environment construction procedure) until reaching the experimental results leading to the paper results are disclosed. By having these, other researchers can reuse the research data at any time, and it is possible to reproduce and verify the research, and it is possible to smoothly derive his / her own research in addition to the original research. In this paper, we focus on the field of mathematics, and report on the implementation method of the experimental mathematical environment that is easy to reconstruct and its application to a specific derived research example.

Keywords: Research Environment Reconstruction, Experimental Mathematics, Jupyter Notebook, Cloud Computing

1. はじめに

オープンサイエンスを推進するための活動の進展やクラウド技術の発展に伴い、特にデータ駆動型科学研究の再現性を支える基盤技術が整備されつつある。本稿では、再現性をさらに進めて、オリジナル研究から派生研究への繋が

りを支える基盤技術について議論する。研究分野によって、その派生研究の発展パターンに多様性があると考え、まず一つの研究分野を固定してこの議論を展開し、その後他の研究分野への横展開を議論することとした。

その具体的な研究分野として設定したのは、数学、特に数論研究である。これは歴史的にも各種数値実験から定理予想が想起され、いくつかのケースではその後証明、理論化されるという活動が盛んに行われてきたことによる。

¹ 群馬大学

Gunma University

² ライフマティクス株式会社

Lifematics

³ 国立情報学研究所

National Institute of Informatics

a) yoko@gunma-u.ac.jp

2. 背景

2.1 データ駆動型科学研究の再現性

データ駆動型研究分野の研究プロセスを単純化すると、(1) データの蓄積、(2) 実験（データ分析と発見）、(3) 論文化となり、(1) と (3) についてはオープン化が進んでいる。残る (2) についてのオープン化を進める手段を提供する必要がある。

図1に「オープンデータのみ研究スタイル」と「データビリティプラットフォームがある研究スタイル」の比較を示す。

データビリティ (Datability) とは、Data と Ability を合わせた造語で「大規模なデータを持続可能 (Sustainability) かつ責任ある形 (Responsibility) で活用する能力」であると定義されており、データ駆動型の研究推進において鍵となる概念である [1]。

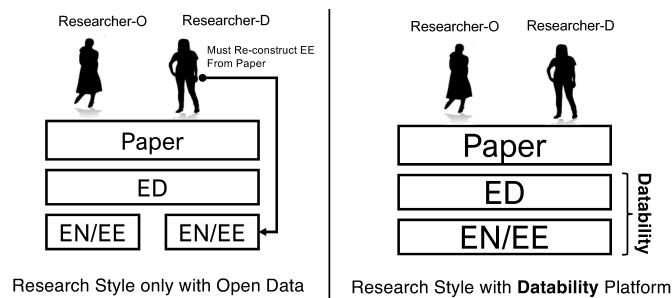


図1 研究スタイルの比較

オープンデータ化によるビッグデータの共有・流通に基づくデータ駆動型研究の推進やデータ駆動型研究人材の育成が、図1「オープンデータのみ研究スタイル」に示すように進んでいる。論文とともに対象となる研究データや研究の途中で生成される中間データも付加価値データとしてオープン化される。データビリティプラットフォームはこのデータのオープン化の動きを促進するために不可欠な要素である。

データのオープン化を促進する一つの鍵は、共有・流通しているデータを活用する計算実験の再現性向上である。データビリティの Ability 軸で言うと、アカデミッククラウドを代表とする情報基盤とそれを活用する人という Ability に加えて、図1に示すように、各実験を Datability として定義し、実験再現のための情報も付加価値データに加えて、それを共有・流通させるこのような仕組みをアカデミックコミュニティとして持つことが、データ駆動型の研究推進を成功させる鍵となると考える。

本報告では実験環境 (Experiment Environment) 図2に示すような三層 (Domain Software, Operating System, Hardware) に分けて考える。

柔軟性を持った実験を共有・流通させるための仕組みと

Layer	Instance
Human	Researcher
Knowledge	Paper
Data	Experiment Data(ED)
Workflow	Experiment Note(EN)
Domain Software	Workflow Engine(WF)
Operating System	Operating System(OS)
Hardware	Hardware(HW)

図2 データ駆動型研究を構成するスタック

して、まず実験を「各分野で標準化された実験環境テンプレート (以後 実験環境テンプレート, EET: Experiment Environment Template)」を使って行う実験環境の構築とその実験環境の上で「各実験を実行するための再現情報 (以後 実験ノート, EN: Experiment Note)」を使って行う実験実施より構成する。前者で構築する実験環境は Controlled Environment であり、後者の実験ノートはそれを制御する Control Data である。図3に示すように、論文と同時に各分野においては標準化された実験環境テンプレートが共有化され、実験ノートを共有・流通することで、実験の再現性が確保される可能性を追求する。Data の Sustainability や Responsibility はこの実験環境テンプレートや実験ノートについても適用されるべきである。

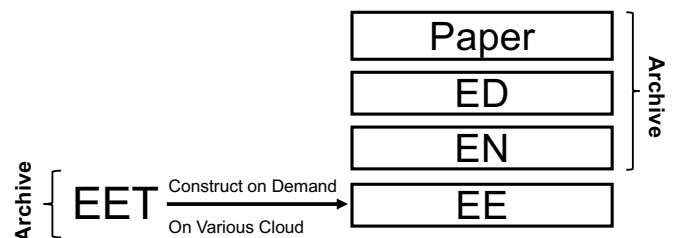


図3 テンプレート利用による実験環境再構成

2.2 実験数学

計算機環境が一般化し身近なものとなることで、様々な学問分野においてそれを活用した取り組みが浸透し、経験科学、理論科学に継ぐ第3のパラダイムとして計算科学、そして第4のパラダイムであるデータセントリック科学の台頭が現実のものとなって来た。

数学分野も特に例外ではなく、数学研究および数学教育におけるデータセントリック科学に相当する実験数学の意味が益々重要になって来ている。

以下に我々の道標とした実験数学についての考え方を示した、2000年に発刊された「実験数学入門」[2]冒頭の実験数学に関する説明文を引用させていただく。

“物理学に、物理学実験と実験物理学の使い分けがある

ように、数学の実験にも、「数学実験」と「実験数学」が考えられます。2つの間に明確な区別があるわけではありませんが、次のように考えると分かりやすいのではないのでしょうか。数学実験：的確な実験により、知識、技術を修得し、密着させる。実験数学：多くの実験を通して、新しい知見を得る。”

古来から数学者は手計算により実験数学に取り組み、そこから数学法則を予想をして数学世界を広げて来た。近年では手計算を出発点にするものの、計算機の力を借りて実験数学を加速する例が珍しくない。

有名な事例は佐藤予想 (Sato-Tate 予想とも呼ばれる) に関する実験数学である。佐藤幹夫は、1962年暮れから始めたラマヌジャンのL関数の素数 p での局所因子の零点の虚部に関する値の分布に関する計算の計算機実験を、1963年に入って難波完爾らの協力を得て行った [3]。その結果、「非常にきれいな実験的な結果が出た」のでその分布に関する佐藤予想の定式化がなされた [4], [5], [6], [7]。このようにして定式化された予想は43年を経た2006年にリチャード・テイラーを中心とする研究により本質的な部分について解決された [8]。

楕円曲線を対象にした当初の実験の後も、対象を超楕円曲線に拡張した実験を実施したり、証明されて定理となった後からも、その意味についての考察からさらに広い観点への拡張が続いている [9], [10], [11], [12], [13]。

この例にとどまらず手計算から始まる実験数学は、計算機実験による拡大の効果がより広い数学世界を作り出す力になっている。この際、計算機を援用することで派生研究の実施範囲を広げ易く、仮説の定式化やその拡張に役立っている。

具体的な派生のパターンとして、佐藤予想の場合を例に分類すると以下ようになる。1963年当初の計算機実験での計算範囲をはるかに超えた実験がその後なされており、このような派生は精度向上に当たる。当初の楕円曲線 (種数=1) という対象を超楕円曲線 (種数>1) に拡大するという方向は曲線の定義パラメータの範囲を拡大するというパラメータ変更に相当する。これ以外にも素数 p のとり範囲を素数全体に設定せずに、その部分集合に限定するという実験対象データの変更も一つの派生方向である。これ以外にそもそも仮説あるいは予想自体を変更するという派生も考えられるけれど、このような方向についてはさらに一段高いレベルの派生にあたるので本報告では扱わないこととする。佐藤予想の派生研究のパターンを元に派生研究の派生の方向性について図4に整理する。

3. 課題

これらの実験数学を起点とする数学研究の可能性は計算機環境の普及と同時に大きく膨らんで行くはずであるけれど、実際には各研究者の利用する実験数学環境が流通しな

佐藤予想の派生実験

素数 $p < N$ の範囲での実験の N を変更
実験対象とする代数曲線を変更
保型形式による計算に変更
基礎体を有理数体以外に変更
グラフ表現への変更

派生の方向性

●————→ ・ 実験精度向上
●————→ ・ パラメータ変更
●————→ ・ アルゴリズム変更
●————→ ・ データ変更
●————→ ・ データ表現変更

図4 佐藤予想の派生の方向性

いことが第一の原因で、未だに数学研究を支える柱とはなり得ていない。普及を妨げる障害には大きく分けて二つのものがある。一つ目の課題 (課題1) は、オリジナルな研究で用いられた実験数学環境を再構築することでオリジナル研究の再現を行う際に発生する問題である。どうしても実験数学環境は各研究者が独自に構築・利用することが多く、派生研究者はまずそのオリジナル研究の再現から実施することになるけれど、その際オリジナル研究で利用した実験数学環境の再構築に大きな手間を要するのが現実である。

実験数学環境に限定せず、研究環境の再現性については筆者らの報告 [14] にある通り、オープンサイエンスの促進のための情報基盤システムとクラウド関連技術を組み合わせることで、この手間を大幅に軽減する手法が現れて来ている。

そこで筆者らは、各所でオンデマンドにデータ分析環境を再現する方式を提案した [15]。具体的には、情報基盤上に構築された計算実験 (以後 実験) の再現性をコンテナ技術を用いたオーバーレイクラウド方式の仮想クラウド技術 [16], [17] と実行可能な構築手順書構成技術 (LC4RI: Literate Computing for Infrastructure)[18] の組合せで実現する。

もう一つの課題 (課題2) は、オリジナル研究を出発点とした派生研究を進める際の実験数学環境をいかに派生研究の方向性に合わせて確保するかということである。本報告では、こちらの課題について焦点をあてて報告する。

図4で整理したそれぞれの派生研究の方向性に沿った支援をする再構成容易な実験環境をどのように構成すべきかを課題と設定し、以下に各章で、その課題解決に関連する既存の研究例、我々の提案する解決策、そしてその解決策に沿った実践例について述べる。

4. 関連研究

公開情報を元にデータ分析環境を再現する取り組みとして、Materials Cloud[19], BinderHub[20] と Chameleon Cloud データ分析環境再現サービス [21] について紹介する。Materials Cloud は研究データとともにデータ分析を再現するために必要な情報を合わせてアーカイブするクラウドサービスである。実際の再現環境自身はそれらの情報を元に再現者が自ら構築することを前提としている。

一方 BinderHub は分析ノートが Jupyter Notebook[22]

で書かれていることを前提として、その分析ノートを実行できる環境を提供するクラウドサービスである。

Chameleon Cloud データ分析環境再現サービスは、NSF Cloud の一つである Chameleon Cloud 上の一つのサービスとして、Jupyter Notebook 環境を活用してこのクラウド上にデータ分析環境再現サービスを提供する。

Materials Cloud は、再現環境自身はそれらの情報を元に再現者が自ら構築する必要があり、この再構築手順は標準化されていないため、論文毎に様々で再現者の研究再現の障害となる可能性がある。また、BinderHub は一つのコンテナイメージでデータ分析ノートが実行される場合には有効であるけれど、マルチノードからなる計算機クラスタのような実行環境が必要な場合には適用できないという制約がある。

一方、Chameleon Cloud 上のデータ分析環境再現サービスは、オンデマンドで Chameleon Cloud 上にデータ分析環境再現するというアプローチで、分野毎の要求や時間を経過することによる仕様変更などは再現手順を記述する Jupyter Notebook の中で吸収する方式である。

本報告で述べるアプローチは、オンデマンドでクラウド上にデータ分析環境を再現するという意味では、Chameleon Cloud 上のデータ分析環境再現サービスに類似するものである。但し、派生研究の促進を視野に入れ、分野毎のデータ分析環境に対する雛形を用意するという点での相違点がある。

5. 解決策

課題 1 への解決策は、オーバーレイクラウド方式の仮想クラウド技術と実行可能な構築手順書構成技術の組合せで実現しているため、これを前提に課題 2 の解決策を提案する。つまり実験のプロセスを記述した Jupyter Notebook とそれを実行するためのデータ分析環境を構築するための Jupyter Notebook は論文や研究データと同様に公開されているという前提を置く。

課題 1 のみに着目すると、ここで止まるのであるけれど、派生研究の支援を視野に入れると、この解決策にさらに追加の構造を付加する必要がある。このような前提の元に追加の構造を考えるので、まず、派生の方向性と実験のプロセスを記述した Jupyter Notebook の関係を図 5 のように整理した。

派生の方向性

- 精度向上 →
- パラメータ変更 →
- アルゴリズム変更 →
- データ変更 →
- データ表現変更 →
- 仮説変更 →

Jupyter Notebookとの関係

- Rangeの変更 in Notebook
- Parameter list in Notebook
- Template Notebookの変更
- Data Path in Notebook
- Visualization in Notebook
- 扱わない

図 5 派生の方向性と Jupyter Notebook の関係

派生研究を支援するための追加の構造を付加する際の基本的な考え方は、これらの派生の方向性毎に存在する Jupyter Notebook 内の要素を各 Jupyter Notebook から分離し、Jupyter Notebook 外に間接化することである。つまり Template Notebook とそれを制御するパラメータ群に分離して管理する構造を付加する。

例えば、パラメータの変更という派生の方向性を支援するためには Jupyter Notebook 内の該当パラメータセットを Jupyter Notebook 内での“パラメータ”として分離し、外部のデータベース内で実験のパラメータセットを管理することにより、この方向の派生研究がよりスムーズに実施できる仕組みを入れる。また、精度向上については、実験の実行範囲の変更もパラメータ化することで対応する。

データの変更はデータへのパスを変更することで対応できると考える。データ表現の変更については表現したいデータをデータフレームなどに格納することで、Jupyter Notebook が持つ機能で分離が可能である。アルゴリズムの変更については、Template Notebook 作成時に中核となるアルゴリズム部分を明確にすることで対応することとする。

本報告では、特にパラメータ変更と精度向上という方向性を中心に報告する。この双方とも勘所はパラメータセット毎の並列実行や実験の実行範囲内の並列実行などの仕組みであり、分離されたパラメータを元にクラウド環境内で並列に実験を遂行する仕組みを付加する方法を提案する。

この仕組みを組み込めば、派生研究の実施者は自分の実験意図に従ったパラメータセット群を決め、必要な精度を決めることでほぼ自動的に派生研究の実験結果が得られる。別の言い方をすると、ここで提案する方法はオリジナル研究の実施者にとってもオリジナル研究を行う上で有用な仕組みであり、そもそもオリジナル研究の公開前の自らの環境にも適用することでメリットが得られる仕組みである。

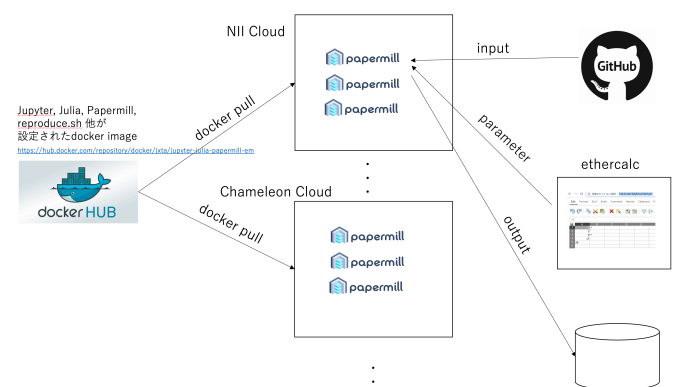


図 6 解決策 (現状)

この並列実行のために我々はコマンドラインから Jupyter Notebook を自動実行できる Papermill[23] と呼ばれるツールを活用する。Papermillを使って Jupyter Notebook を起

動する際にはコマンドパラメータとして Jupyter Notebook 内で用いるパラメータを与えることができるので、この機能を活用する。また、その際与えるパラメータセット群を管理するデータベースとしては扱いやすい API を持っているクラウドサービス ethercalc[24] を用いる。図 6 に課題 2 への解決策の全体の構成を示す。

なお、最終的には課題 1 への解決策の中にこの構成を埋め込んだ図 7 のような構成にする必要がある。この統合化については、国立情報学研究所で提供しているオンデマンドクラウド構築サービスのアプリケーションテンプレートとして提供するなどの方策をとることで技術的には実現できる見通しがあるため、本報告ではこの部分については述べない。

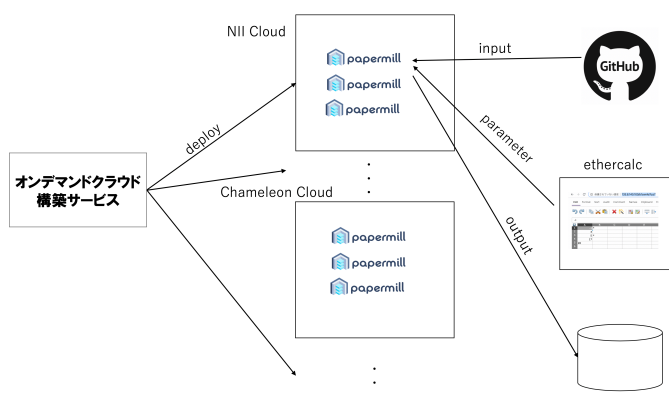


図 7 解決策 (統合版)

6. 派生研究への適用例

本章では我々の解決策を一つの実験数学環境に適用した事例について説明し、この解決策の可能性について議論したい。背景で述べた佐藤予想についてはすでに主要な部分については数学的に解決されてきているので、この報告では、まだ数学的に未解決な「深リーマン予想」[25], [26], [27], [28], [29], [30], [31] についての実験数学を扱うこととする。

本報告で適用例とする論文 [28] は数学の有名な未解決問題リーマン予想を含む予想である深リーマン予想についての数値実験を含んでいる。この数値実験の派生研究を本報告の解決策に沿った実践例として紹介する。この数値実験は数体上の L 関数 $L(s)$ のオイラー積表示についての収束を扱っている。深リーマン予想はこの関数の $s = 1/2$ における無限オイラー積の収束を言っている。つまり相当する積表示の収束に関する実験は、 $p < N$ で N を無限大に持って行くときの有限オイラー積の値の振る舞いを実験する。

オリジナル研究では $p < 10^7$ の範囲の有限オイラー積と L 関数の $s = 1/2$ での値を比較している。こちらは精度向上の観点では、さらに大きな素数までの有限オイラー積を計算することになる。

また、オリジナル論文で扱っている数値実験の L 関数は 2 次指標 (d/\cdot) に対応するものを扱っていて、 $d = -3, -4, \dots, 29$ に対応する 19 種類の指標に対応した実験を行なっている。

図 8 にオリジナル研究の論文 [28] の中から、ここで再現および派生させる数値計算に関係する部分の一部を引用させていただく。 $s = 1/2$ がもっとも重要であるけれど、他の s でのオイラー積表示の値やオイラー有限積収束の様子も興味がある話題でありオリジナル研究でも触れられている。こちらの方向もパラメータ変更に対応する。

3. NUMERICAL CALCULATIONS

In this section we show some numerical data supporting the Deep Riemann Hypothesis (Conjecture 2). If this conjecture is true, the partial Euler product

$$L_x(s, \chi) = \prod_{p \leq x} (1 - \chi(p)p^{-s})^{-1},$$

converges to $L(s, \chi)$ or $\sqrt{2}L(s, \chi)$ as $x \rightarrow \infty$ even on the critical line $\text{Re}(s) = 1/2$. We formally put $L_x(s, \chi) = L(s, \chi)$ for $x = \infty$.

First we give Table 1, which shows the accuracy of Conjecture 2 at $s = 1/2$. We find that the ratio of $\sqrt{2}L(\frac{1}{2}, \chi)$ and $L_x(\frac{1}{2}, \chi)$ is almost equal to 1 for $x = 10^7$, when χ is quadratic.

d	$\sqrt{2}L$	E	$(\sqrt{2}L)/E$
-3	0.680049	0.688002	0.988440
-4	0.944258	0.945909	0.998254
5	0.327745	0.320619	1.022223
-7	1.621517	1.640320	0.988536
8	0.528479	0.539992	0.978680
-8	1.556230	1.521663	1.022716
-11	1.402301	1.342967	1.044181
12	0.705066	0.729170	0.966942
13	0.621678	0.618558	1.005044
-15	2.612093	2.791265	0.935809
17	1.020601	1.066235	0.957201
-19	1.137621	1.173052	0.969795
-20	2.375413	2.356696	1.007942
21	0.703235	0.724051	0.971250
-23	3.472406	3.320551	1.045732
24	1.003325	1.057376	0.948881
-24	2.223023	2.130498	1.043428
28	1.162994	1.199957	0.969196
29	0.658655	0.683281	0.963958

TABLE 1. $L := L(\frac{1}{2}, \chi)$, $E := \prod_{p \leq 10^7} (1 - \frac{\chi(p)}{p})^{-1}$.

図 8 オリジナル研究の数値計算実験 ([28] p.8 より転載)

実際、この数値計算の再現だけを実現するのであれば図 9 に示すような Jupyter Notebook を流通させることで実現可能である。実際、この Notebook を実行することで $p < 10^9$ の範囲あるいはそれを超える範囲での精度向上が見込める。

ただ、 $p < 10^9$ までの計算結果である図 10 をオリジナル論文の図 8 にある L 関数の $s = 1/2$ での値と比較すると、その収束の様子をつかむことが容易でないことが分かる。従って精度向上の仕方もこのような簡単な方法では派生研究を支えるのには不十分であると考えられる。さらに派生の範囲を指標の種類を広げる方向に考える場合、単純に指標を決める d の値を増やすだけでは不十分で、並列実行も考慮した仕組みを導入する必要がある。

我々の行なった派生実験は、精度向上と 2 次指標の種類増加と s の変動である。また、データ表現についてもオリジナル研究は数値表になっているのに加えて収束を示すグラフ表現を追加した。

```
function l_function3(d, n)
    P = primes(n)
    L = Vector{Float64}(undef, length(P))
    size = length(P)
    for i in 1: size
        p = P[i][1]
        L[i] = 1/(1- legendresymbol(d, p)*(1/sqrt(p)))
    end
    l_p = 1.0
    for j in 1: size
        l_p = l_p * L[j][1]
    end
    l_p
end

function reproduce(n)
    for j in [-3 -4 5 -7 8 -8 -11 12 13 -15 17 -19 -20 21 -23 24 -24 28 29]
        println(l_function3(j, n))
    end
end

println("-----")
reproduce(10^7)
println("-----")
reproduce(10^8)
println("-----")
reproduce(10^9)
```

図 9 単純な派生研究の数値計算実験のための Notebook

```
-----
0.688002116449104 0.6936570793407454 0.6711973131895831
0.9459092752863326 0.9609346838995422 0.9809041624466672
0.32061990362634746 0.33207675395084485 0.3239718304721752
1.6403206160987356 1.5631052179715497 1.633811404203082
0.5399923534021607 0.5232348100901406 0.5294723192300219
1.5216639420290181 1.580485009936785 1.6219302548144625
1.3429676236522667 1.3015969087435706 1.357335322483763
0.7291707632246083 0.7145076308453128 0.7103990113929324
0.6185580266630368 0.5823424031171116 0.6153128920642423
2.7912653716593767 2.789935028923919 2.5475159887754915
1.0662352269904616 1.057170286995489 0.9887893005384226
1.1730529765379205 1.1179084769982124 1.1560693778382232
2.3566961388375267 2.522173320408326 2.349118491200374
0.7240516671115267 0.7206999568181199 0.7064519987931197
3.3205511582736182 3.4254955456580816 3.4675124609969026
1.0573761951305216 1.0212168932750174 1.0312212529296534
2.1304988197072445 2.3304834620625337 2.2437060603896253
1.199957120221541 1.050140852896047 1.1447982157981247
0.6832818131679798 0.6181122384651955 0.667149578468032
```

図 10 単純な派生研究の数値計算実験のための Notebook を使った数値計算結果 (左から $p < 10^7$, $p < 10^8$, $p < 10^9$)

本報告で紹介する派生実験の方向性を図 11 にまとめる。

図 12 に示す Template Notebook を基本にこの Notebook の中で変数化されている指標を決める d (パラメータ変更の方向) と精度向上のための素数 p の範囲を決める bank を変化させて papermill を実行スクリプト (reproduce.sh) の中で起動する。どういう変数 d , bank について実際に papermill を起動するのかをコントロールするのは各実験環境で共有されている etharcale のページ中で管理されている。その変数値に関する実行ステータスである。実際に papermill 起動時に設定された papermill のパラメータの d と bank は、この Template Notebook の最初の CELL である “Set Parameters” で Notebook 内の変数として引き渡される。

派生の方向性

本派生実験の対応

- 精度向上 → $p < \text{big number}$ の範囲の有限オイラー積
- パラメータ変更 → 指標 d の値の拡大、 $s = \frac{3}{4}, 1$, その他 複素数
- アルゴリズム変更 → オイラー積の定義式を変更
- データ変更 → p を素数全体ではなく $\text{mod } 4$ で 1 のものに限定するなど
- データ表現変更 → 収束を示すための統計的処理各種

図 11 本派生実験の方向性

以下に示す実験のための実行スクリプト (reproduce.sh)

を各実行基盤であるクラウド内のマシン上で可能な範囲で並列実行することで、派生研究の実験が実施できる。実際の並列実行状況を図 13 に示し、実行結果の一部を図 14 に示す。図 13 の右に示している各実験環境では複数の reproduce.sh が並列に実行されていると同時に各 Notebook 実行内でも Julia の並列実行機能を利用してマルチコアを活用している。現状、reproduce.sh の各実験環境での並列度は、それぞれの実行環境の性能や実験状況のモニタリングを派生研究者が手動で行う形態となっており改善が必要な点である。

なお、さらなる精度向上を実現するために必要な大きな素数表としては 10 億番目の素数までを含むものをストレージサーバ上に確保して対応している。クラウド環境内にクラスタを構築し精度向上方向およびパラメータ変更方向のどちらの方向へも派生研究が展開できる。

```
## 実験のための実行スクリプト reproduce.sh ##

#!/bin/bash
URL='http://133.8.x.x/./yyyyyyyyyy'
NOTEBOOKS='/home/yoko/notebooks/euler_products'
for i in `seq 1 1000`; do
    status=`curl --include ${URL}/cells/CS{i} 2> /dev/null|tail -1 |jq -r '
        .datavalue'`
    if [ $status = w ]; then
        d=`curl --include ${URL}/cells/AS{i} 2> /dev/null|tail -1 |jq -r '
            .datavalue'`
        prime_file_no=`curl --include ${URL}/cells/BS{i} 2> /dev/null|tail -1 |jq -r '
            .datavalue'`
        cmd_f="curl -X POST -H \"Content-Type: application/json; charset=utf-8\" -d \"${command}\""
        cmd_m="\set CS{i} text t r\"
        cmd_b="\"${URL} /dev/null2>&1"
        cmd=""
        cmd+="${cmd_f}"
        cmd+="${cmd_m}"
        cmd+="${cmd_b}"
        eval $cmd
    fi
    papermill ${NOTEBOOKS}/l-d-b-template.ipynb ${NOTEBOOKS}/l-${d}-${prime_file_no}.ipynb -p d ${d} -p bank ${prime_file_no}
    cmd_f="curl -X POST -H \"Content-Type: application/json; charset=utf-8\" -d \"${command}\""
    cmd_m="\set CS{i} text t r\"
    cmd_b="\"${URL} /dev/null2>&1"
    cmd=""
    cmd+="${cmd_f}"
    cmd+="${cmd_m}"
    cmd+="${cmd_b}"
    eval $cmd
done
exit 0
```

7. 今後の取り組み

今回の報告で述べた派生実験以外に背景のところでも例としてあげた佐藤予想についても同じ枠組みで派生実験が実施可能であることを確認している。しかしながら、まだ少数な例での確認にとどまっており、それぞれの研究分野における派生実験を継続すると同時に、別の予想についても実績を積み上げ本方式へのフィードバックを続けて行く必要がある。

また、今回の報告では図 7 の解決策 (統合版) として将来形として紹介したオンデマンドクラウド構築サービスとの統合についても近い将来実現したい。さらには、reproduce.sh の並列度の調整は実行状況のモニタリングをして自動調整する必要があり、すでに検討が進んでいる実験環境の動的再構成手法 [32] を適用することでさらに利便性の高い “再構成容易な実験数学環境” を目指したい。

```

Set Parameters
d =
bank =

Template Body
#time begin
total_size = 10^8
P = zeros(Int64, total_size)
size_L_val = Int(floor(10^10/total_size))
segment_size = 10^7
segments = Int(floor(total_size/segment_size))

# Euler Product
L_val = zeros(Float64, size_L_val)
io_input = open("/primes/primes-$bank", "r")

l_val = 1.0
for l in 0:(size_L_val - 1)
    for i in 1:total_size
        P[i] = read(io_input, Int64)
    end
    L = zeros(Float64, segments)

    addedprocs = addprocs(5)
    @everywhere begin
        using Combinatorics
        # Euler Product from P[i] to P[j]
        function l_function7_segment(d, i, j)
            @distributed (*) for k in i:j
                1/(1 - legendresymbol(d, P[k])*(1/sqrt(P[k])))
            end
        end
    end

    for i in 0:segments - 1
        L[i+1] = l_function7_segment(d, 1+segment_size*i,
            1+segment_size*(i+1))
        end
    end
    rmprocs(addedprocs)

    for i in 1:segments
        l_val = l_val * L[i]
    end
    L_val[l+1] = l_val
    println(l_val)
end
println("d = $d")
println(L_val)
close(io_input)
#end

x = 1:size_L_val; y = L_val[x];
z = zeros(size_L_val)
for i in 1:size_L_val
    # These are the plotting data
    for j in 1:i
        z[i] = (z[i] + L_val[j])
    end
    z[i] = z[i]/i
end

plot(x,y)
plot!(x,z)
plot!(size=(1000,200))
savefig("/output_fig/1-$d-$bank.png")

println(L_val[size_L_val])

io_output = open("/output/1-$d-$bank", "w")
write(io_output, L_val)
close(io_output)
    
```

図 12 Template Notebook

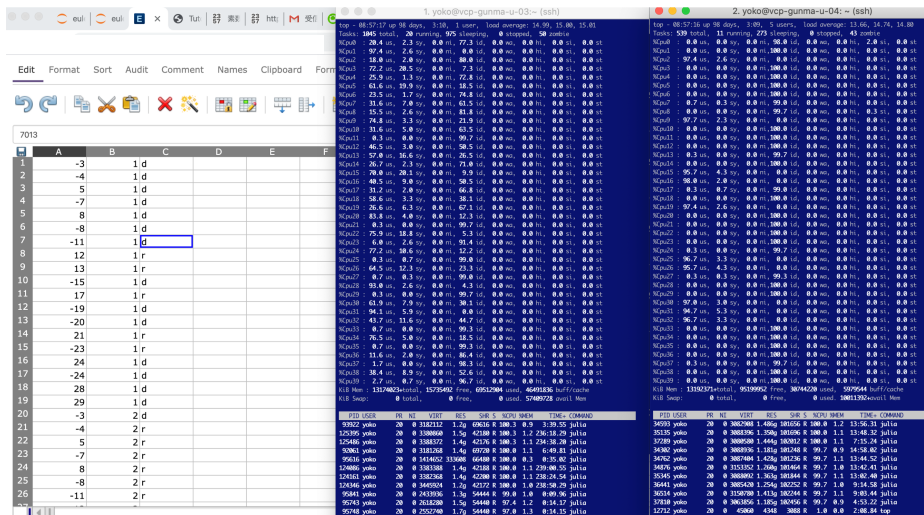


図 13 本派生実験実行状況

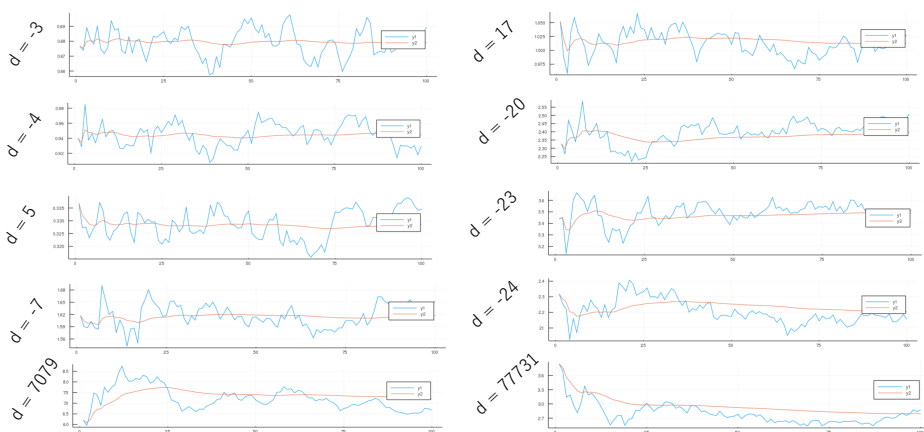


図 14 本派生実験実施結果 (一部)

参考文献

- [1] Isak, C.: Datability, <https://techacute.com/datability/> (accessed on 06-01-2020).
- [2] 山本芳彦: 実験数学入門, 岩波書店 (2000).
- [3] 佐藤幹夫: 私の数学 (数論へーラマヌジャン予想), 佐藤幹夫の数学, 日本評論社, pp. 46–48 (2007).
- [4] 難波完爾: 佐藤 \sin^2 -予想の話, 佐藤幹夫の数学, 日本評論社, pp. 344–367 (2007).
- [5] 加藤和也, 斎藤 毅, 黒川信重: 数論 I — Fermat の夢と類体論, 岩波書店 (2005).
- [6] 数学のたのしみ, 2008 最終号, フォーラム:現代数学のひろがり: 佐藤 - テイト予想の解決と展望, 日本評論社 (2008).
- [7] 加藤和也: 類体論と非可換類体論 I —フェルマーの最終定理・佐藤 - テイト予想解決への道, 岩波書店 (2009).
- [8] Taylor, R.: Automorphy for some l -adic lifts of automorphic mod l Galois representations. II, *Publications Mathématiques de l’IHÉS*, Vol. 108, pp. 183–239 (2008).
- [9] 難波完爾: 種数 2 の楕円曲線と \sin^2 -予想, 第 17 回数学史シンポジウム (2006).
- [10] 難波完爾: 種数 3 の超楕円曲線と \sin^2 -予想, 第 18 回数学史シンポジウム (2007).
- [11] Murty, M. R. and Murty, V. K.: The Sato-Tate conjecture and generalizations, *Current Trends in Science: Platinum Jubilee Special*, pp. 639–646 (2009).
- [12] Barnet-Lamb, T., Gee, T. and Geraghty, D.: The Sato-Tate conjecture for Hilbert modular forms, *Journal of the American Mathematical Society*, Vol. 24, No. 2, pp. 411–469 (2011).
- [13] Johansson, C.: On the Sato-Tate conjecture for non-generic abelian surfaces, *Transactions of the American Mathematical Society*, Vol. 369, No. 9, pp. 6303–6325 (2017).
- [14] 横山重俊, 浜元信州, 長久 勝, 藤原一毅, 政谷好伸, 竹房あつ子, 合田憲人: オープンサイエンス推進のためのデータ分析プロセス共有, 第 48 回インターネットと運用技術研究会 2020-IOT-48(27), 情報処理学会, pp. 1–6 (2020).
- [15] 横山重俊, 浜元信州, 長久 勝, 藤原一毅, 政谷好伸, 竹房あつ子, 合田憲人: データ駆動科学研究の再現性を支えるデータ分析環境再構成方式, 第 48 回インターネットと運用技術研究会 2020-IOT-48(13), 情報処理学会, pp. 1–8 (2020).
- [16] 竹房あつ子, 佐賀一繁, 丹生智也, 横山重俊, 合田憲人: 学認クラウドオンデマンド構築サービスの推進, AXIES 2018 年度年次大会 (2018).
- [17] 国立情報学研究所: 学認クラウドオンデマンド構築サービス, <https://cloud.gakunin.jp/ocs> (accessed on 06-01-2020).
- [18] 長久 勝, 政谷好伸, 谷沢智史, 中川晋吾, 合田憲人: Notebook を介した作業ノウハウの継承・移転を分析するための基盤, 第 44 回インターネットと運用技術研究会 2019-IOT-44(16), 情報処理学会, pp. 1–6 (2019).
- [19] AiiDA: Material Cloud, <https://materialscloud.org/> (accessed on 06-01-2020).
- [20] Jupyter: BinderHub, <https://mybinder.org/>, <https://github.com/jupyterhub/binderhub> (accessed on 06-01-2020).
- [21] Anderson, J. and Keahey, K.: A Case for Integrating Experimental Containers with Notebooks, *11th IEEE International Conference on Cloud Computing (Cloud-Com 2020)* (2019).
- [22] Jupyter: Jupyter Notebook, <http://jupyter.org/> (accessed on 06-01-2020).
- [23] Jupyter: Papermill, <https://papermill.readthedocs.io/> (accessed on 06-01-2020).
- [24] Ethercalc: Ethercalc, <https://ethercalc.net/> (accessed on 06-01-2020).
- [25] Conrad, K.: Partial Euler Products on the Critical Line, *Canadian Journal of Mathematics*, Vol. 57, No. 2, pp. 267–297 (online), DOI: 10.4153/CJM-2005-012-6 (2005).
- [26] Gonek, S.: Finite Euler products and the Riemann hypothesis, *Transactions of the American Mathematical Society*, Vol. 364, No. 4, pp. 2157–2191 (2012).
- [27] Koyama, S.-y., Suzuki, F. et al.: Euler products beyond the boundary for Selberg zeta functions, *Proceedings of the Japan Academy, Series A, Mathematical Sciences*, Vol. 90, No. 8, pp. 101–106 (2014).
- [28] Kimura, T., Koyama, S. and Kurokawa, N.: Euler Products Beyond the Boundary, *Letters in Mathematical Physics*, Vol. 104, No. 1, pp. 1–19 (online), DOI: 10.1007/s11005-013-0644-3 (2014).
- [29] Akatsuka, H.: The Euler product for the Riemann zeta-function in the critical strip, *Kodai Mathematical Journal*, Vol. 40, No. 1, pp. 79–101 (online), DOI: 10.2996/kmj/1490083225 (2017).
- [30] Kaneko, I. and ya Koyama, S.: Euler products of Selberg zeta functions in the critical strip (2018).
- [31] Kaneko, I.: Euler product asymptotics on Dirichlet L -functions, *arXiv: Number Theory* (2019).
- [32] Tanjo, T., Jingtao, S., Saga, K., Takefusa, A. and Aida, K.: Dynamic Framework for Reconfiguring Computing Resources in the Inter-Cloud and Its Application to Genome Analysis Workflows, *Proceedings of the 11th International Conference on Internet and Distributed Computing Systems (IDCS 2018)*, pp. 160–172 (2019).