Free Random Projection for In-Context Reinforcement Learning

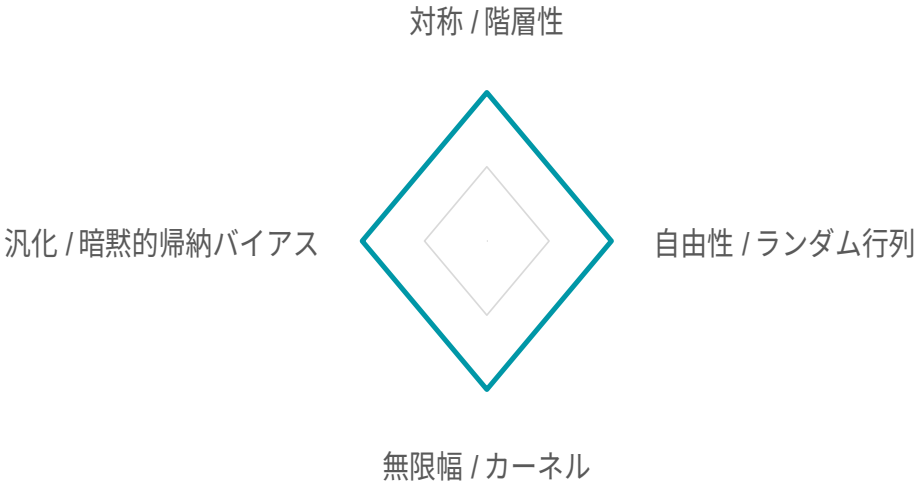
数理・情報系研究集会 2025@ 京都大学
早瀬 友裕

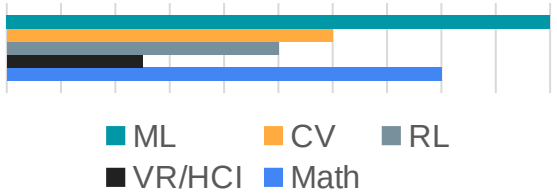
arXiv:2504.06983 **[cs.LG]**
Code available at https://github.com/ThayaFluss/frp_rl
Joint work with Benoit Collins, Nakamasa Inoue.

自己紹介

- 早瀬 友裕， 博士（数理科学）
- 専門：数理科学 / 機械学習
- 根源的な面白がり：ランダム性や無限次元といった，本来捉えきれないものを有限の概念で捉え，計算できるようにすること．
- ACT-X との関わり：１期生＋加速「自由確率論による深層学習の研究」

研究軸と関連分野





関連分野

1. 導入：研究軸と関連分野
2. **背景：汎化と暗黙的バイアス**
3. 数理への回帰：自由群とランダム行列
4. FRP の紹介
5. 今後の発展

Total 41 pages

背景と問題意識

汎用的な機械学習モデルは、訓練環境と異なる状況でも高い性能を発揮すること(汎化)が求められる。

訓練とテスト環境に何らかの共通点を仮定して、モデルを対応させるのが帰納バイアス.

ところが、モデル構造が弱くても、DNNが膨大なパラメータにも関わらず汎化性能を維持できるケースが存在. 背景には、モデル設計に明示的に含まれていない、学習アルゴリズム全体の**暗黙的帰納バイアス**が働いていると考えられる.

機械学習研究者はさまざまなデータ（画像、言語、系列）において、モデルを汎化させる暗黙的バイアスを探している. 可搬性や汎用性、融合可能性を考えると、**柔軟な暗黙的バイアス**を選べることが重要ではないか？

創発的対称性・階層性：
明示的にモデル内部に組み込まなくても、学習アルゴリズムから創発される対称性・階層性

ー＞柔軟なバイアスの設計には**高次統計量**に気を配ると良いケースがあり、数理がその手掛かりとなる。

背景 1: 汎化 / 暗黙的帰納バイアス

モデルと基礎レイヤー :

MLP: full-linear

MLP-Mixer: tensor of linear

Vision Transformer: attention

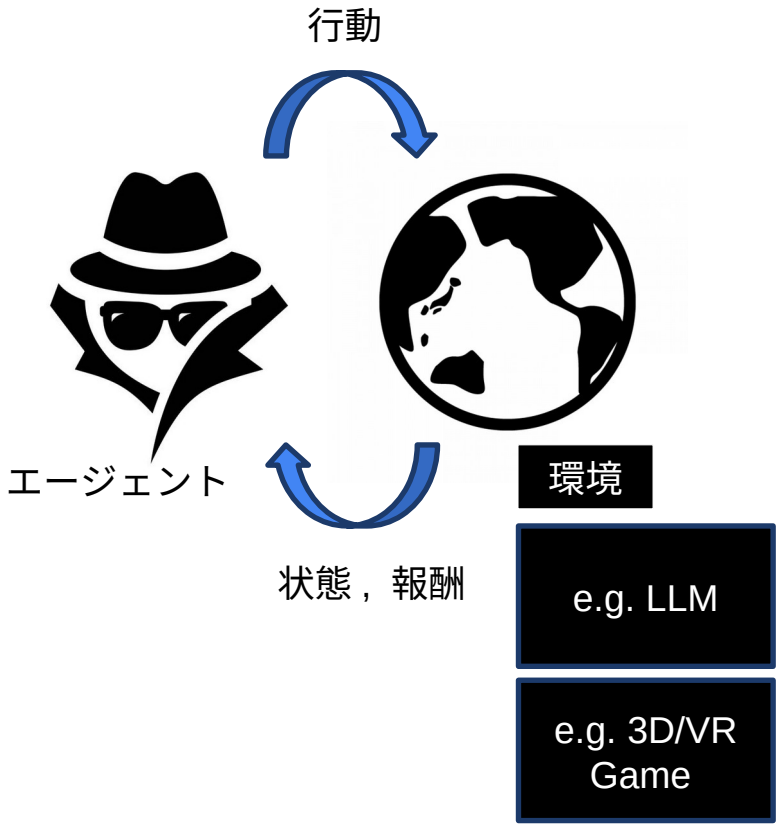
CNN: convolution

- [T & R. Karakida, ICML2024]Understanding MLP-Mixer as a wide and sparse MLP
- [R. Karakida+, ICML2023]Understanding Gradient Regularization in Deep Learning: Efficient Finite-Difference Computation and Implicit Bias
- [G. Backman, NeurIPS2023]Scaling MLP: a talk of inductive bias

背景 2: 強化学習における汎化

次世代 AI のひとつの候補： 言語だけでなく，様々な環境を認識しつつ自律的に行動できるような人工知能：

　LLM ・3 DCG 環境等，ブラックボックス関数として扱い最適化する必要性が増している．



何が課題として残っている？：

　環境・タスクに関する汎化．

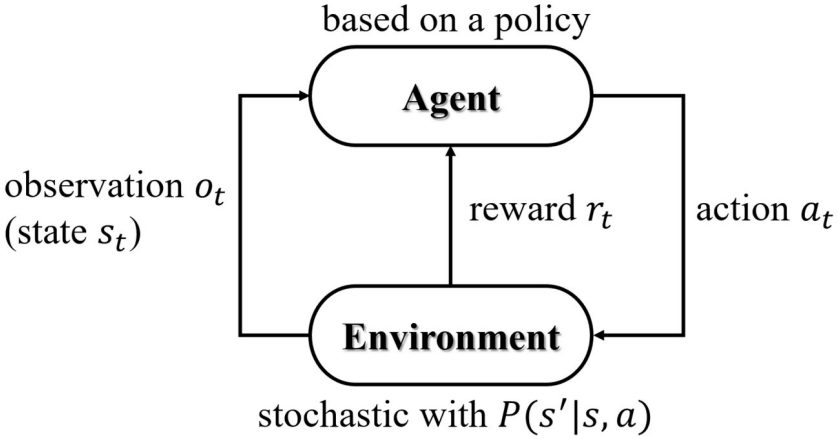
強化学習の基本設定

MDP: Markov Decision Process

(S, A, R, P, \gamma)

POMDP: 部分観測マルコフ決定過程

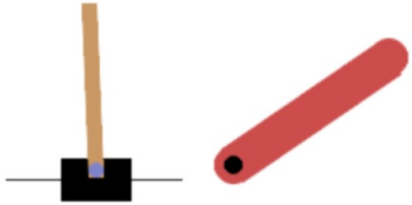
環境の一部分しか観測できない .



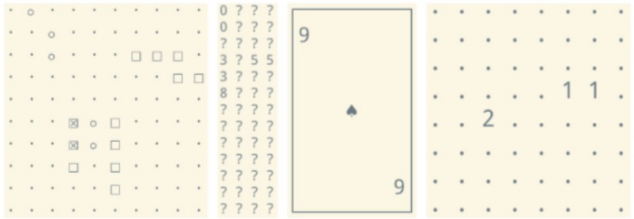
メタ強化学習

メタ・ In-Context 強化学習： 小〜中規模タスクで検証の段階

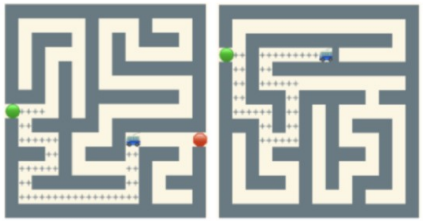
POPGYM での環境・タスク例 :



(a) Stateless Cartpole and Stateless Pendulum



(b) Battleship, Concentration, Higher Lower and Mine Sweeper



(c) Labyrinth Escape and Explore

e.g. Morad+,ICLR2023,
Lu+, NeurIPS2023.

メタな状態表現と
行動ポリシー

少数実行
サンプル

素早い適応

未知タスク

In-Context RL

Meta-RL の一種 .

学習後のモデルに少数の事例を見せ、（パラメータ更新なしで）未知タスクに適応させる

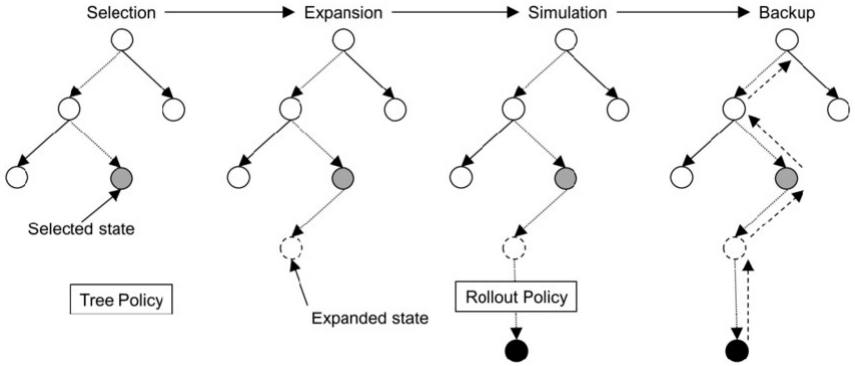
例：囲碁将棋で学習させたあと、チェスの対局例を見せ適応させる

例：学習後の LLM にプロンプトで回答例を見せ適応させる

例：コントロールタスク・ボードゲーム・迷路などカテゴリーのことなるタスクで学習後、新しいタスクの例を見せ適応させる

状態の表現方法

MCTS: 広大な探索空間が得意
(e.g. AlphaZero, AlphaGo)



S5 (状態空間モデル), Mamba: 長系列の扱いが得意 . 自然言語で話題にあがるが、強化学習でも活用される .[Lu+, NeurIPS2023]

環境や行動を表す状態の適切な表現が必要で、木構造・状態空間モデルは有望

状態表現にメタ的に共通する構造はなにか？

仮説： 階層性 or 双曲性 .

木構造以外にも，これまで，強化学習では H 双曲的な距離構造をいれることにより性能が改善することが見えていた .

しかし、基本的に双曲構造を明示的にいれるため、汎用性に乏しい

- 最新モデルにすぐに融合できない
- **弱めの帰納バイアス**のほうが汎化が期待される

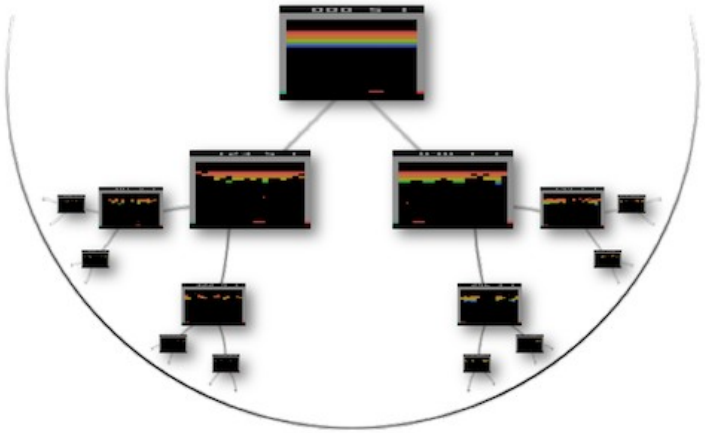


Figure 1: Hierarchical relationship between states in *breakout*, visualized in hyperbolic space.

Figure from Cetin+ ICLR2023.

背景 3 ：ドメインシフト

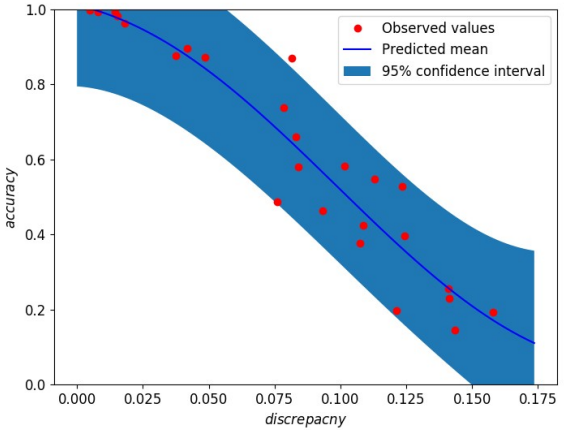
ドメインシフト：データセットを切り替えたときの latent の分布がずれること

Maximum Mean Discrepancy (MDD):

source と target における分布間の距離

Maximum Classifier Discrepancy (MCD) [Sait

classifier を使って , MDD を測る .



MDD は二次統計量：**系列データセットでは MDD では捉えきれないことがある**
[後述]

MCD + Bayes Regression
[T+, 2019]

13

高次相関・高次統計量

良く使われる相関・統計量

平均、分散、共分散、カーネル値 K(x,x’) 、分布間距離 (e.g. MMD)

高次相関

歪度、尖度、高次 cumulant,

カーネルの固有値分布 , 特異値分布

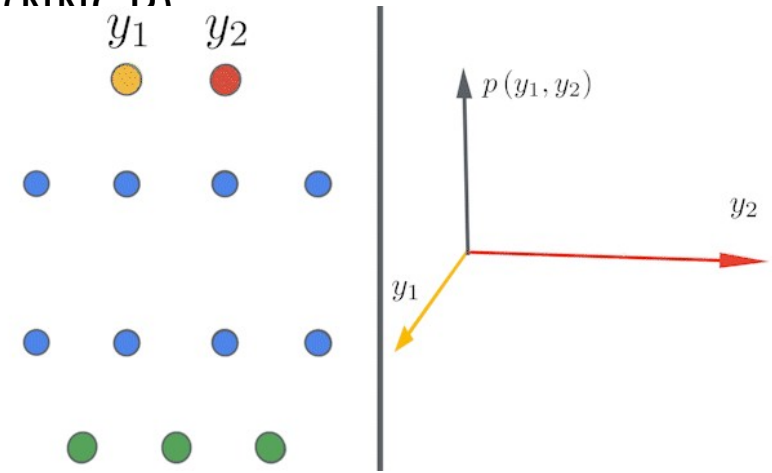
14

背景 4: 深層回路の幅無限極限

無限幅極限で出力 / 勾配はガウス分布

出力： Neural Network Gaussian Process (NNGP)

勾配： Neural Tangent Kernel (NTK)



このジャンルで高次統計が必要になるケース

- 固有値分布
- 4-th Cumulant

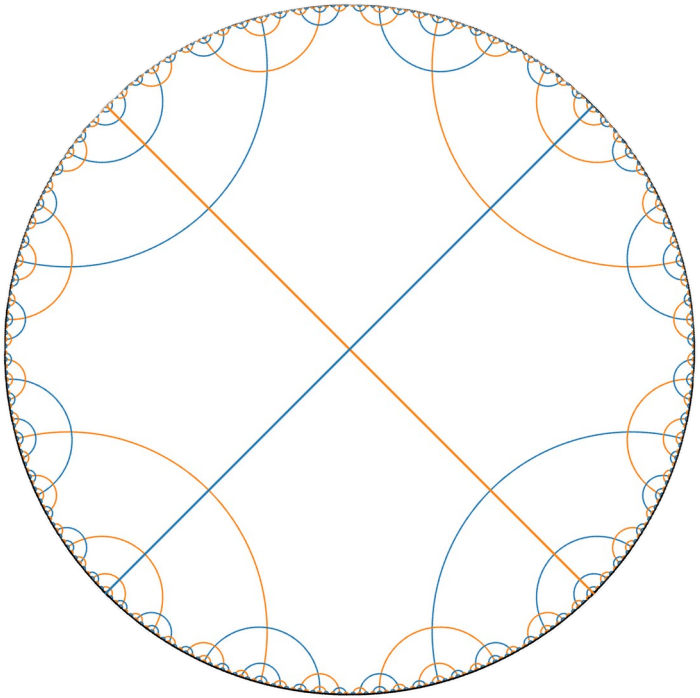
背景まとめ

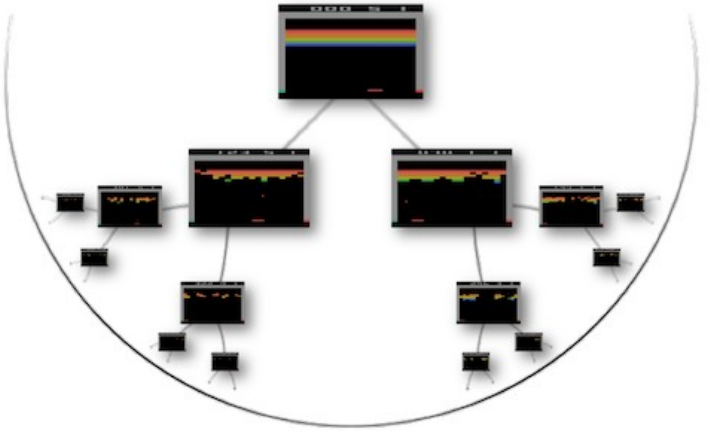
1) 暗黙的バイアスと汎化：暗黙的バイアスがほしい
2) 強化学習における汎化：双曲性や階層性がほしい
3) ドメインシフト：汎化を捉える上では高次統計量がほしい
4) 深層回路の幅無限極限：深層回路を理解する上で高次統計量が使われてきた

数理の話をしよう

1. 導入：研究軸と関連分野
2. 背景：汎化と暗黙的バイアス
3. **数理への回帰：自由群とランダム行列**
4. FRP の紹介
5. 今後の発展

階層性 / 双曲性のために丁度良いものが自由群





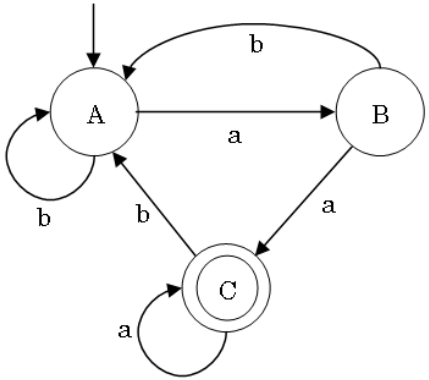
自由モノイドと自由群

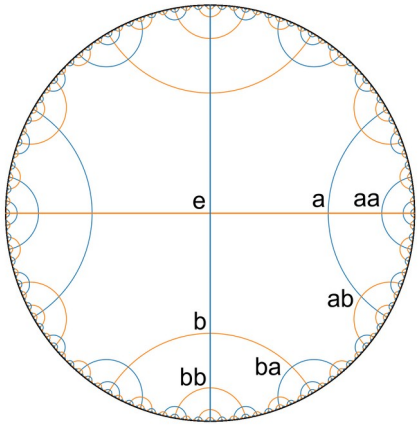
自由モノイド：生成元 a,b を用意して語をつくり，　積を語の連接で定義する：

　w=ab,　 v=aaa　=>　 wv = abaaa.

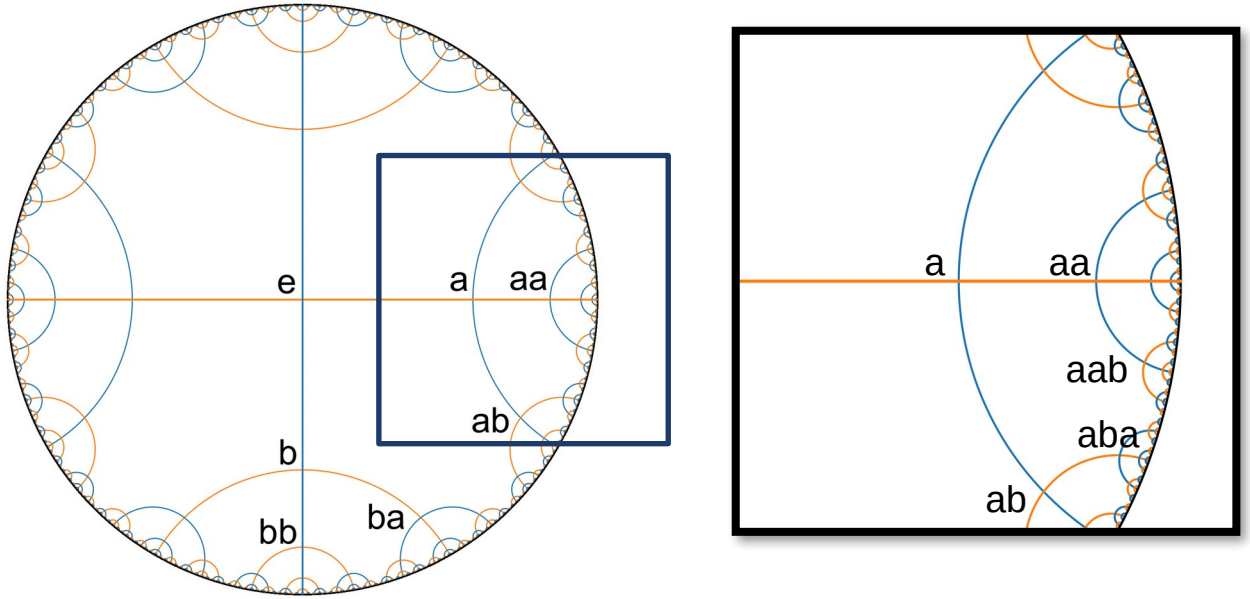
形式言語における言語は自由モノイドの部分集合．

自由群：逆元もいれて、逆元が隣り合った時の簡約だけは入れる．これら語の集合は群になり，自由群と呼ばれる．
その Cayley グラフは木であり、 Gromov 双曲空間の代表例．



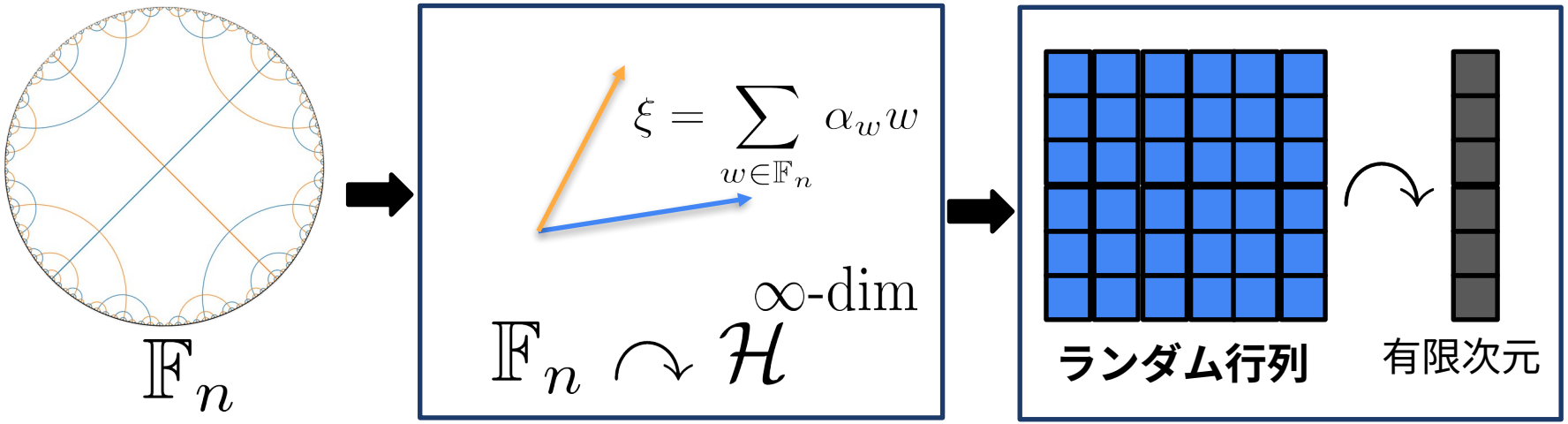


自由群は無限に階層性を持っている



Key Point: 線形化と有限次元化による汎用性

- 群のままでは使いづらい：**ベクトル化**して類似度をいれる（群ヒルベルト空間）
- 無限次元では計算できない：**ランダム行列で近似**する (Voiculescu, 86~)

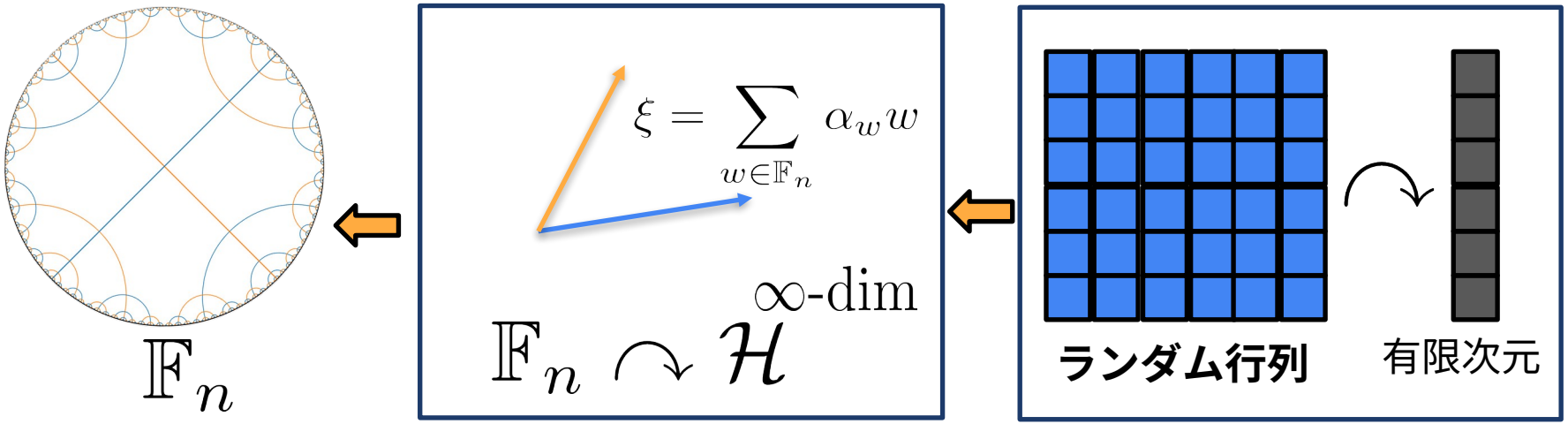


* ランダム行列は、これまで深層学習の研究でも使われてきた : e.g. MFT/NNGP/NTK

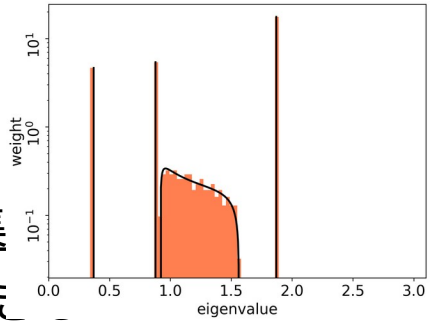
これまでは逆向きだった

- ランダム行列は、これまで深層学習の研究でも使われてきた：

e.g. 深層回路の MFT/NNGP/NTK/DI （に関連する量）を、ランダム行列を自由群で近似して計算する



逆向きの例 : Dynamics & Learning Rate



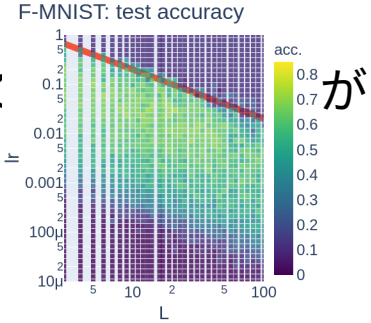
- **Dynamical Isometry**

  Dynamical Isometry （Jacobian Spectrum が１周辺に集
  設計すると，超深層（＞ 10000 層）ネットワークも学習で

- **Fisher 情報行列**

  DNN の FIM は学習ダイナミクスの解析に使える . D.I. のもとて が
  最大固有値＝層数に集中す [T, Ryo Karakida, AISTATS2021]

- 学習率予測 .



$$\frac{2}{\lambda_{\max}(H_L)} \sim \frac{2}{L}.$$

23

漸近的自由性（内積 ver.）

Let $U_1, \ldots, U_n \sim \mathrm{Unif}(\mathrm{O}(d))$, i.i.d., $\lambda : \mathbb{F}_n \to \mathrm{O}(d)$; $\lambda(a_i) = U_i$ $(i = 1, 2, \ldots, n)$. Then

$$\lim_{d\to\infty} \mathbb{E}[\langle \lambda(v), \lambda(w) \rangle_{\mathrm{O}(d)}] = \langle v, w \rangle_{\mathbb{F}_n} \quad \text{for any } v, w \in \mathbb{F}_n,$$

where $\langle U, V \rangle_{\mathrm{O}(d)} = d^{-1}\mathrm{Tr}U^{\mathsf{T}}V$, and $\langle v, w \rangle_{\mathbb{F}_n} = 1$ if $v = w$ and 0 otherwise.

要するに、**群の語の問題（いつ単位元になるか？）**判定を表現先のランダム行列で再現したもの。これをまず群環上の計算に拡張し、次に作用素環に一般化すると Voiculescu の漸近的自由独立性になります。

自由独立性は確率論の独立性と並列していて、たとえば自由独立なら分布が自由畳み込みに分解するとか、自由中心極限定理（収束先は自由ガウス分布＝半円分布）も成立

1. 導入：研究軸と関連分野
2. 背景：汎化と暗黙的バイアス
3. 数理への回帰：自由群とランダム行列
4. **FRP の紹介**
5. 今後の発展

Free Random Projection

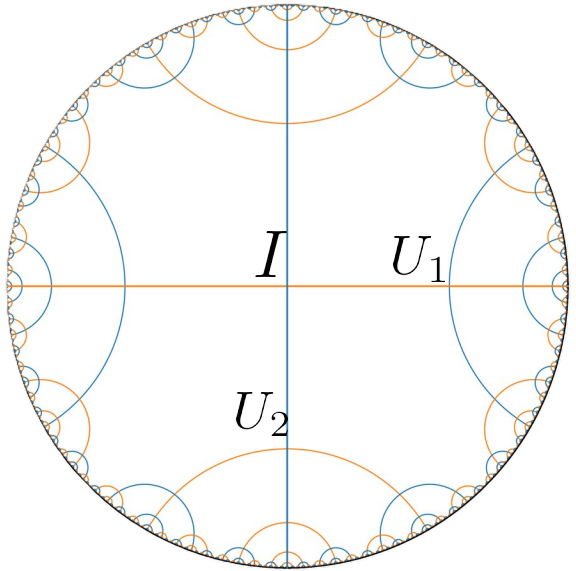
まずは IID 一様直交ランダム行列をサンプリング（これは自由群の生成元を近似）.

$$U_1, U_2, \ldots, U_n.$$

これらをかけ合わせて変換行列を構成：

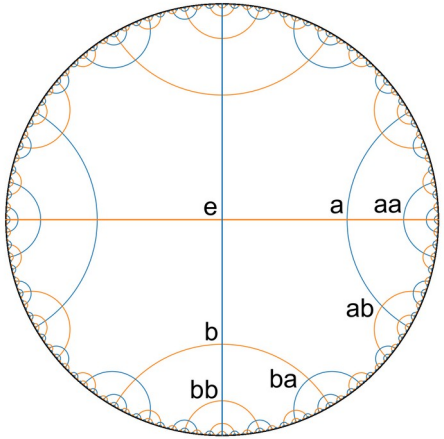
$$A = U_{i_1} U_{i_2} \cdots U_{i_\ell}$$

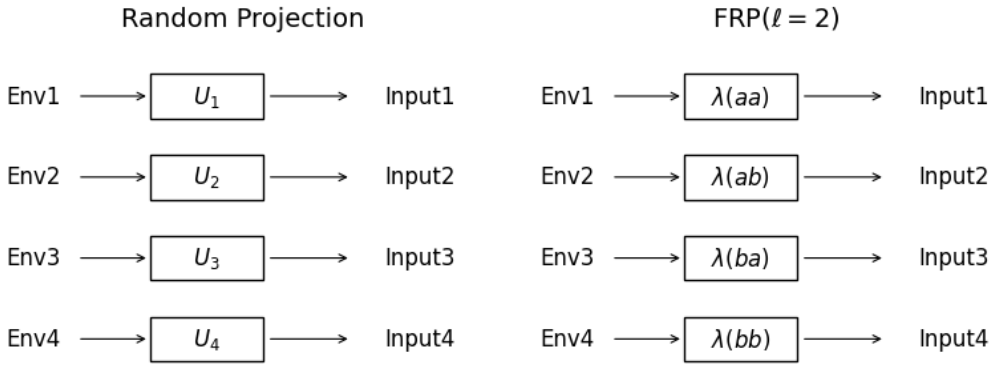
$$i_1, i_2, \ldots, i_\ell \in [n].$$



指数的パターンの変換行列が完成

Free Random Projection





学習中定期的にベースのランダム行列をサンプリング
各環境には、語の分布からサンプリングして変換行列とする。

$$U_1, U_2, \ldots, U_n.$$
$$A = U_{i_1} U_{i_2} \cdots U_{i_\ell}$$

Word Distribution

Word Distribution. We define the following collection of words to investigate the impact of word length for $n, \ell \in \mathbb{N}$:

$$\Lambda_n^\ell := \{a_{i_1} \dots a_{i_\ell} \in \mathbb{F}_n \mid i_1, \dots, i_\ell \in [n]\}. \tag{3.3}$$

We define the word distribution as $\rho_{\mathbb{F}} := \mathrm{Unif}(\Lambda_n^\ell)$, and sample words $\{w_1, \dots, w_{n_e}\}$ independently from $\rho_{\mathbb{F}}$ for the n_e parallel environments. In our analysis, we fix the total number of possible words $n_w > n_e$ and compare various pairs of (n, ℓ) that satisfy $|\Lambda_n^\ell| = n_w$ to ensure a fair comparison.

環境ステップでの使い方

Algorithm 1 Meta RL environment step with FRP

Require: Distribution of environments $\rho_{\mathcal{E}}$, Agent action a and Environment termination 1^{done}
Require: Distribution of words $\rho_{\mathbb{F}}$, Matrix representation $\lambda : \mathbb{F}_n \to \mathrm{O}(d)$.

function StepEnvironment(a, 1^{done})
  if the environment terminated (1^{done}) **then**
    Sample random environment $E \sim \rho_{\mathcal{E}}$
    Sample random word $w \sim \rho_{\mathbb{F}_n}$
    Initialize random observation projection matrix $M_o = \sigma_w T_2 \lambda(w) T_1^E$
    Initialize random action projection matrix M_a
    Reset E to receive an initial observation ξ
    Apply the random observation projection matrix to the observation $\xi' = M_o \xi$
    Initialize $r = 0$ and $1^{\text{done}} = 0$
    return $\xi', r, 1^{\text{done}}$
  else
    Apply the projection matrix $a' = M_a a$
    Step E using a' to receive the next observation ξ, reward r, and done signal 1^{done}
    Apply the projection matrix $\xi' = M_o \xi$
    return $\xi', r, 1^{\text{done}}$
  end if
end function

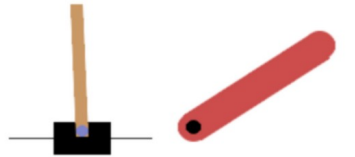
予備実験

Cartpole （位置観測のみ） + Resettable S5(状態空間モデル) でのメタ強化学習実験 .

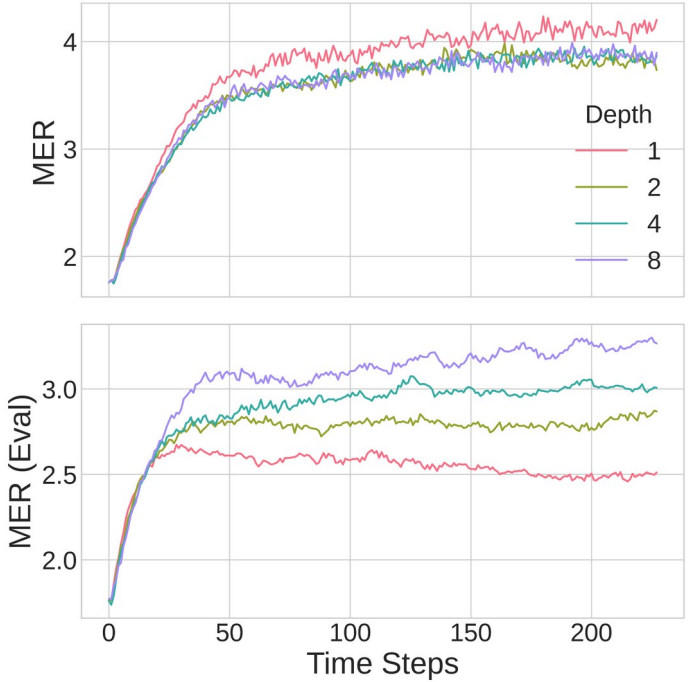
Depth の異なる Free Random Projection を使って実験し比較 . （ depth=1 が従来手法 .)

→ 階層性が汎化へ好影響

自由群の Depth による汎化性能向上を確認 .



(a) Stateless Cartpole and Stateless Pendulum



他のタスクでの比較

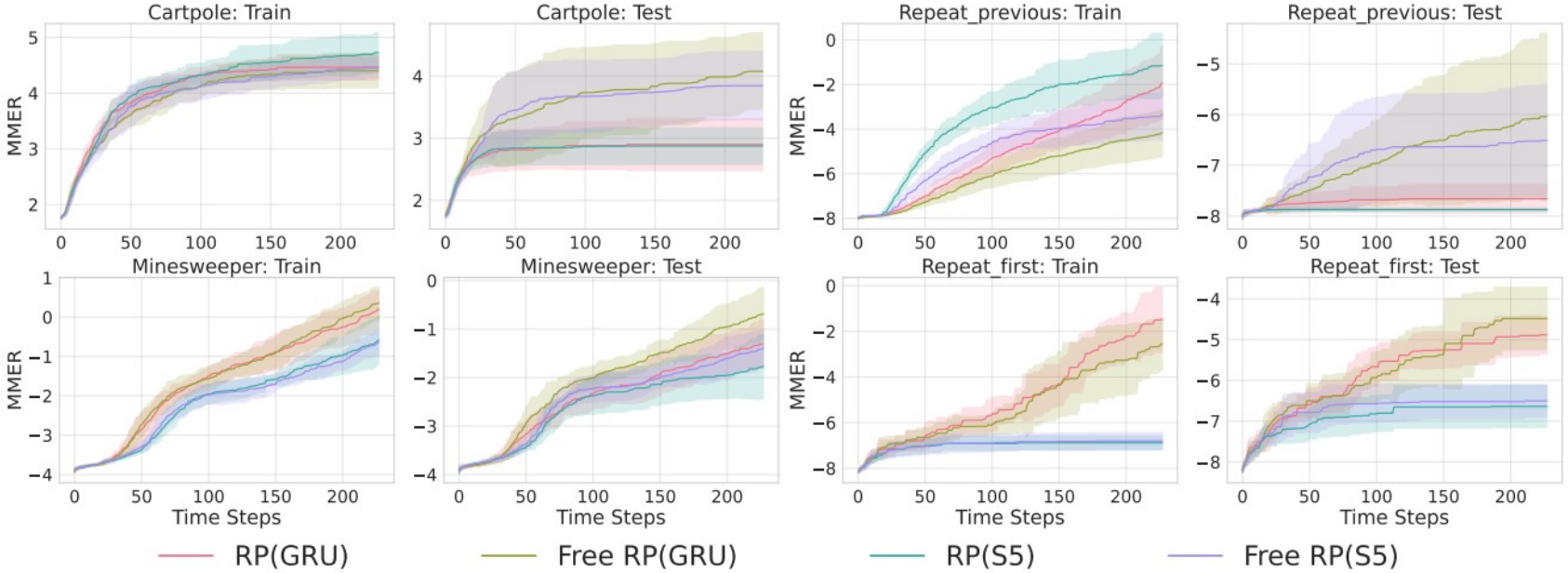


Figure 2: Performance of FRP vs. standard RP on four environments – Stateless Cartpole, Repeat Previous, Mine Sweeper, and Repeat First – shown in the top-left, top-right, bottom-left, and bottom-right subplots, respectively. We use ℓ^* in Table 2 for FRP. Each subplot plots Train MMER and ICL-Test MMER. Shaded regions indicate standard error across 10 random seeds.

定量評価

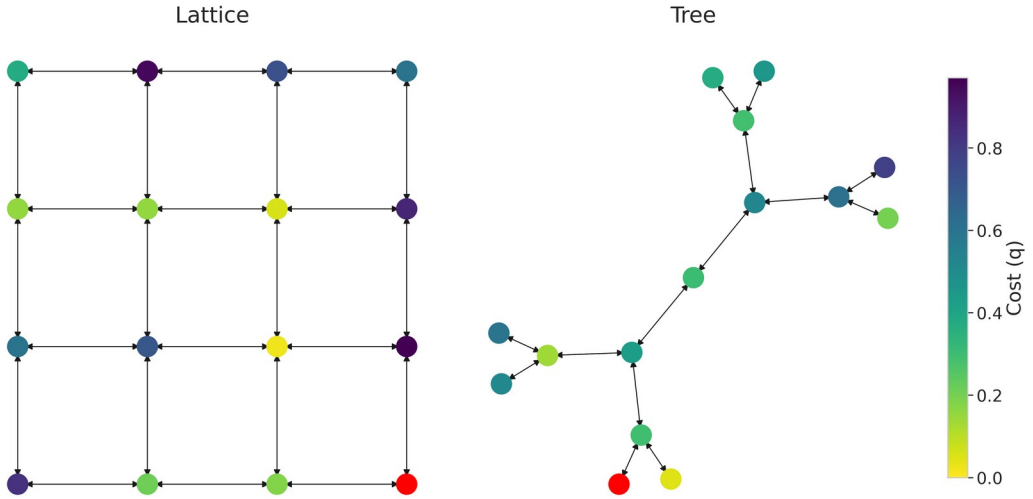
- FRP + GRU が Best であった。構造のないモデル＋暗黙的バイアスの入ったアルゴリズムが良い汎化を与えるというのは、深層学習のストーリーラインにのっている

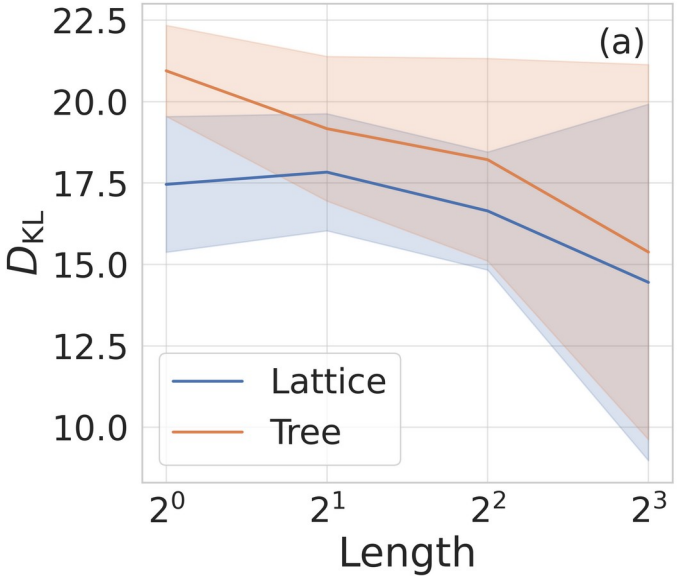
Table 1: Test performance comparison across POPGym environments; Stateless Cartpole, Higher Lower, Mine Sweeper, Repeat First, and Repeat Previous. Each value indicates the mean and standard error of ICL-Test MMER at the last step across 10 random seeds.

Method	S. Cartpole	H. L.	M. Sweeper	R. First	R. Previsous
RP(gru)	2.90 ± 0.42	0.18 ± 0.29	-1.31 ± 0.50	-4.88 ± 0.47	-7.66 ± 0.28
RP(s5)	2.87 ± 0.29	0.08 ± 0.14	-1.77 ± 0.65	-6.64 ± 0.52	-7.87 ± 0.03
FRP(gru)	$\mathbf{4.08 \pm 0.62}$	$\mathbf{2.13 \pm 1.12}$	$\mathbf{-0.69 \pm 0.56}$	$\mathbf{-4.48 \pm 0.76}$	$\mathbf{-6.04 \pm 1.65}$
FRP(s5)	3.85 ± 0.55	1.04 ± 0.82	-1.39 ± 0.38	-6.50 ± 0.40	-6.51 ± 1.10

FRP の階層性は本当に意味がある？可解モデルで検証

- Linearly Solvable MDP での実験：状態空間を格子と木で比較
- 木構造状態空間の方が、ワード長に適応して汎化誤差下がった。 FRP は状態空間の構造に反応した！

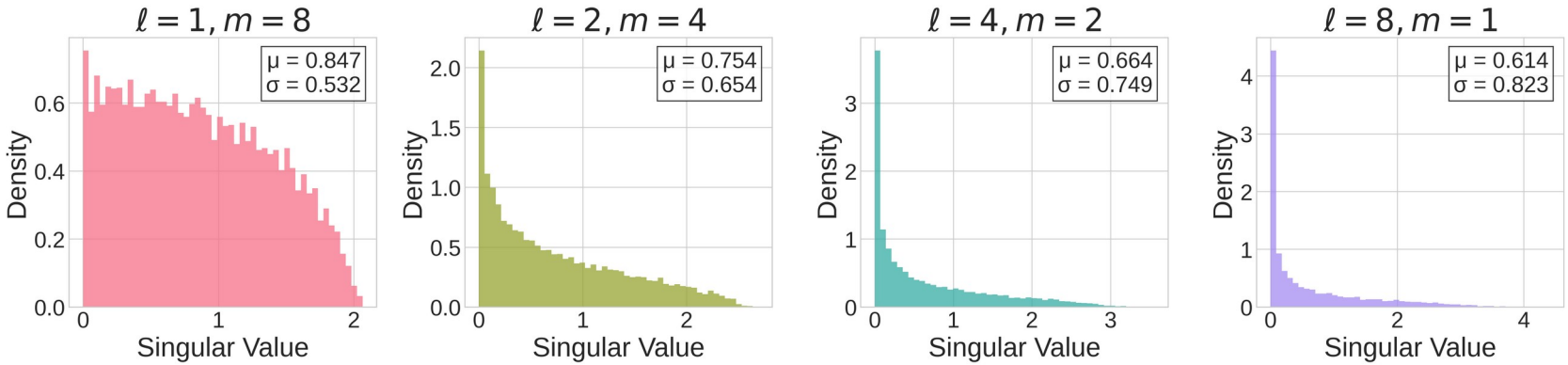




FRP の暗黙的バイアスを測る指標はあるか？

The *averaged kernel matrix* K for FRP: Given samples $X_1, \ldots, X_p \in \mathbb{R}^d$,

$$K_{i,j} = \sum_{w,w' \in \Lambda_n^\ell} \langle \lambda(w) X_i, \lambda(w') X_j \rangle / n^\ell \quad (i,j = 1,2,\ldots,p).$$



有効次元

有効次元は汎化誤差バウンドの項の一つ。 FRP のワード長が長いほど有効次元が小さく、汎化誤差が小さいと期待できる

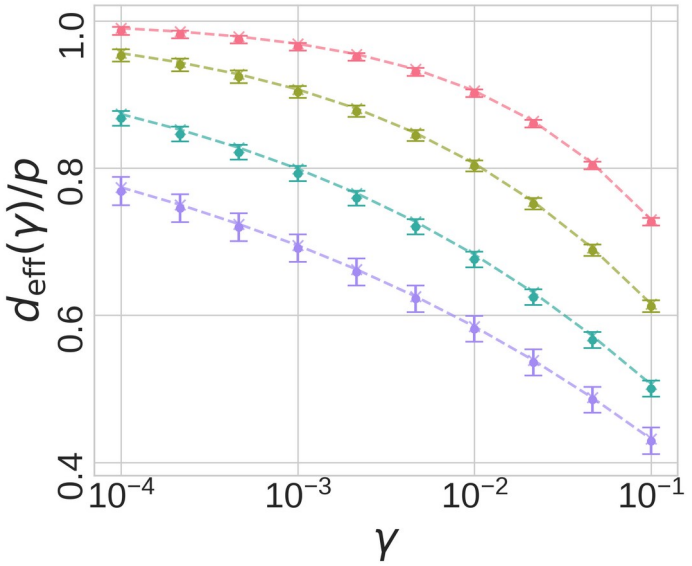
＊有効次元の理論値も自由群で計算できる

Theorem 5.1. *Fix $n, \ell \in \mathbb{N}$. Consider $\lambda : \mathbb{F}_n \to \mathrm{O}(d)$ with (3.1). Assume that $XX^\top$ has the compactly supported limit spectral distribution ν with $\int_{\mathbb{R}} t\nu(dt) \neq 0$ as $p, d \to \infty$ with $p/d \to c \in (0, \infty)$. Then, under the limit of p and d, we have*

$$\mathbb{E}[d_{\mathrm{eff}}(\gamma)/p] \to -\psi(-1/\gamma), \quad \gamma > 0, \tag{5.2}$$

where ψ is the inverse function of χ given by

$$\chi(z) = \frac{z}{z+1}\left(\frac{z/n+1}{z+1}\right)^{\ell} \mathscr{S}_\nu(z). \tag{5.3}$$



1. 導入：研究軸と関連分野
2. 背景：汎化と暗黙的バイアス
3. 数理への回帰：自由群とランダム行列
4. FRP の紹介
5. **今後の発展**