Finding morphisms with gradient descent

Urban Jezernik, Ljupčo Todorovski, Urh Primožič [1]

[1]Fakulteta za matematiko in fiziko, Inštitut Jožef Stefan

Representations of finite groups

Let $|G| < \infty$ and $V = \mathbb{C}^n$.

Definition

$\rho\colon G \to \mathrm{GL}(V)$ is *irreducalbe*, if there is no $0 < W < V$, such that $G \cdot W \leq W$.

Theorem

Every representation $\rho\colon G \to GL(V)$ is equal to $\bigoplus_{i \in I} \rho_i$, where every ρ_i is irreducable.

Modeling representations

Take any model

$$\begin{aligned} \hat{\rho} \colon G &\to \mathbb{C}^{n \times n} \\ g &\mapsto \begin{bmatrix} g_{1,1} & \cdots & g_{1,n} \\ \vdots & & \vdots \\ g_{n,1} & \cdots & g_{n,n} \end{bmatrix} = \phi_g \end{aligned}$$

and keep changing its parameters $\phi = \{\phi_g \mid g \in G\}$ until $\hat{\rho}$ becomes a representation.

Just on generators

Let $G =< S|R >$. Define $\hat{\rho}\colon G \to \mathbb{C}^{n\times n}$ as

$$\hat{\rho}(s) = \begin{bmatrix} s_{1,1} & \cdots & s_{1,n} \\ \vdots & & \vdots \\ s_{n,1} & \cdots & s_{n,n} \end{bmatrix} = \phi_s$$

on generators $s \in S$ and $\hat{\rho}(s_1 s_2 \cdots s_m) = \hat{\rho}(s_1)\hat{\rho}(s_2)\cdots\hat{\rho}(s_m)$ elsewhere.

Relation loss

Let $G =< S|R >$. $\hat{\rho}$ is a homomorphism iff

$$\hat{\rho}(r) - I = 0$$

for every $r \in R$.

$$\mathcal{L}_{\text{rel}}(\hat{\rho}) = \frac{1}{|R|} \sum_{r \in R} ||\hat{\rho}(r) - I||_F^2$$

Theorem

$\hat{\rho}$ is homomorphism iff $\mathcal{L}_{\text{rel}}(\hat{\rho}) = 0$

Gradient descent

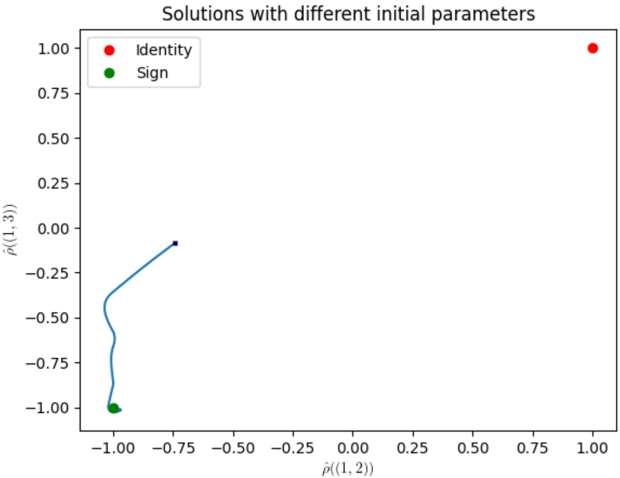
Optimisation using

$$\phi_{i+1} = \phi_i - \eta \nabla \mathcal{L}(\phi_i)$$

Gradient flow equation

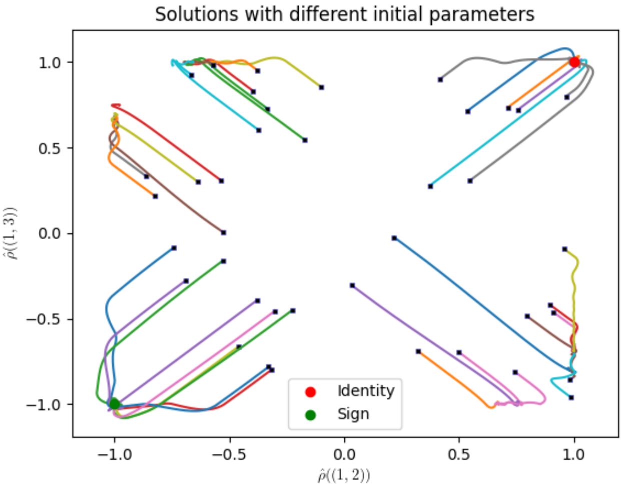
$$\frac{d\phi}{dt} = -\nabla \mathcal{L}(\phi), \quad \phi(0) = \phi_0$$

One dimensional representations of S_3



Path of model $\hat{\rho}\colon S_3 =< (1,2),(1,3) > \to \mathbb{R}$

One dimensional representations of S_3



Path of model $\hat{\rho}\colon S_3 =< (1,2),(1,3) > \to \mathbb{R}$ for different initial values.

Irreducibility loss

Theorem

ρ is irreducible iff $|\chi_\rho| = \frac{1}{|G|}\sum_{g\in G} \mathrm{tr}(\rho(g))\mathrm{tr}(\rho(g^{-1})) = 1.$

$$\mathcal{L}_{\mathsf{irr}}(\hat{\rho}) = (|\chi_{\hat{\rho}}| - 1)^2$$

Unitary loss

Theorem

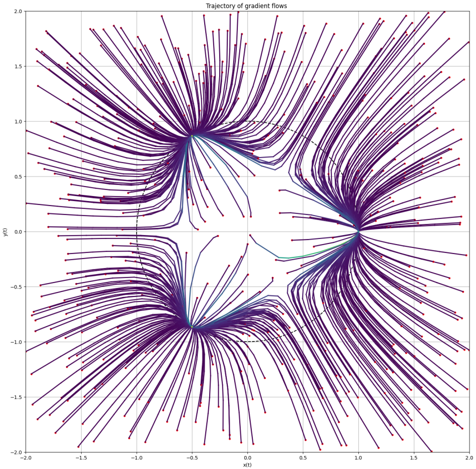
Every representation $\rho \colon G \to GL_n$ *of finite group is equivalent to unitary representation* $\rho_{\text{unitary}} \colon G \to U_n$.

$$\mathcal{L}_{\text{unitary}} = \frac{1}{|S|} \sum_{s \in S} ||\hat{\rho}(s)\hat{\rho}(s)^H - I||$$

$$\mathcal{L} = \mathcal{L}_{\text{rel}} + \mathcal{L}_{\text{irr}} + \mathcal{L}_{\text{unitary}}$$

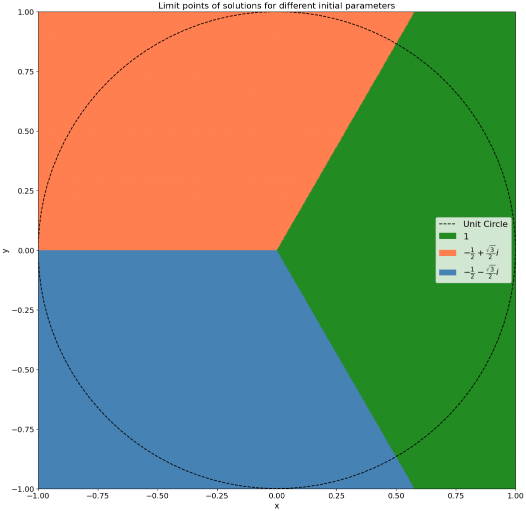
Cyclic groups

$C_n =< z \mid z^n = 1 >$, model $\hat{\rho}(z) = x + iy$ with parameters x, y.



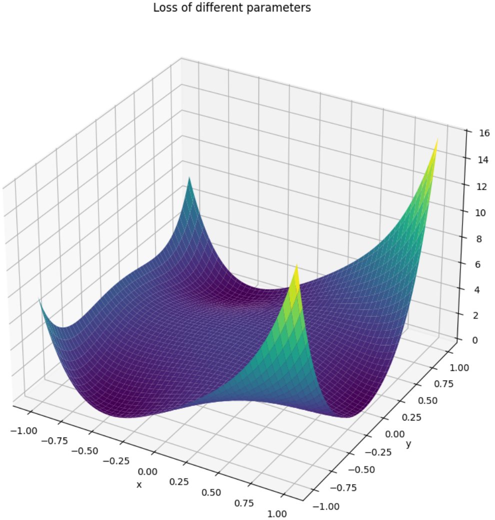
Solutions with different initial values.

Dynamics



Limit points of different initial values

Loss function



Plot of $\mathcal{L}$ for $x, y \in [-1, 1]$

Dihedral Groups

$$D_{2n} = \langle r, s \mid r^n = s^2 = (rs)^2 = 1 \rangle$$

Theorem

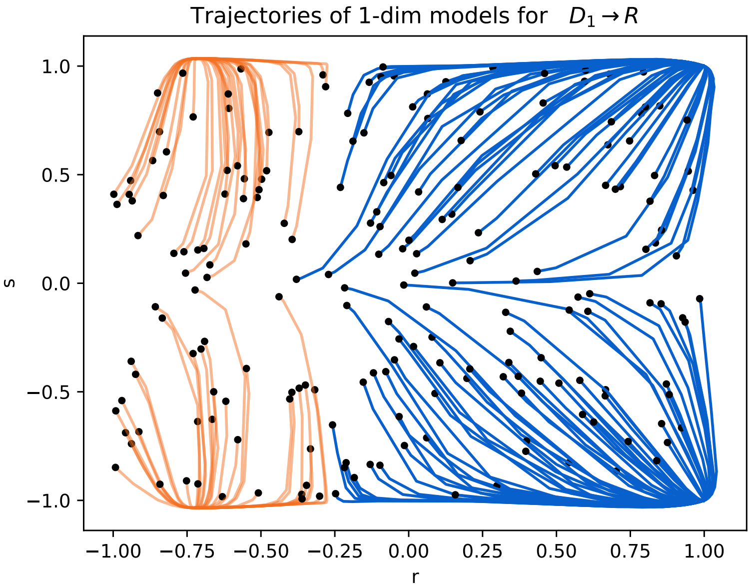
The irreducible representations of the dihedral group D_{2n} are:

$$\rho_k : \; s \mapsto \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad r \mapsto \begin{bmatrix} \cos\left(\frac{2\pi k}{n}\right) & -\sin\left(\frac{2\pi k}{n}\right) \\ \sin\left(\frac{2\pi k}{n}\right) & \cos\left(\frac{2\pi k}{n}\right) \end{bmatrix}$$

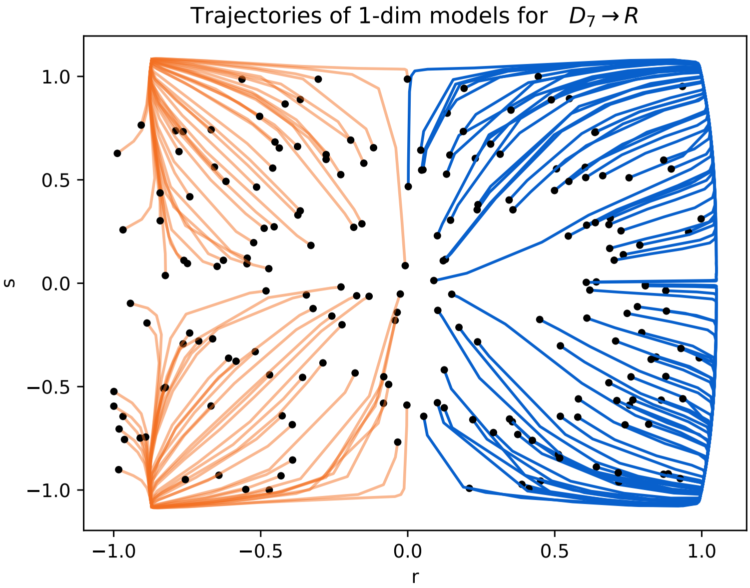
and

$$\chi_{\varepsilon,\delta} : \quad s \mapsto \varepsilon, \quad r \mapsto \delta$$

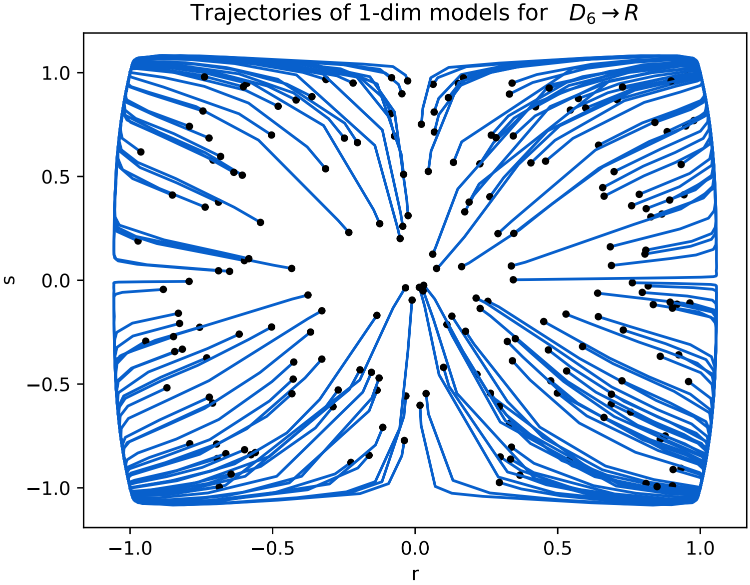
where $0 \leq k < \frac{n}{2}$, $\varepsilon \in \{-1, 1\}$, *and* $\delta \in \begin{cases} \{-1, 1\} & \text{if } n \text{ is even,} \\ \{1\} & \text{if } n \text{ is odd.} \end{cases}$



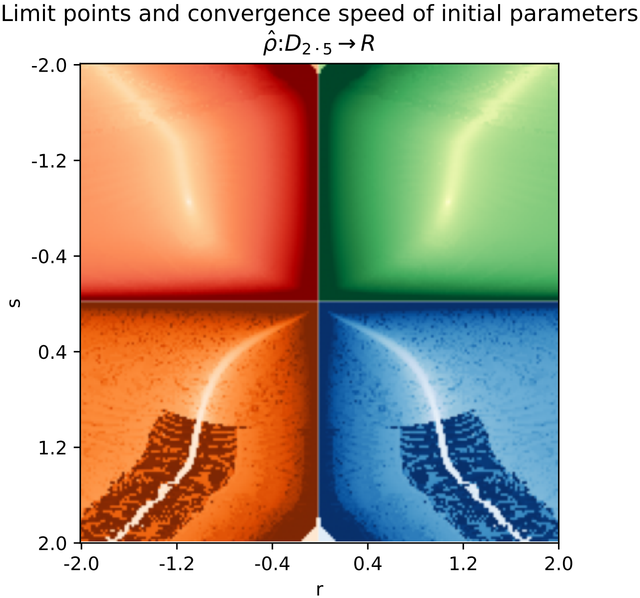
Showing 200 out of 5000 sampled trajectories



Showing 200 out of 5000 sampled trajectories

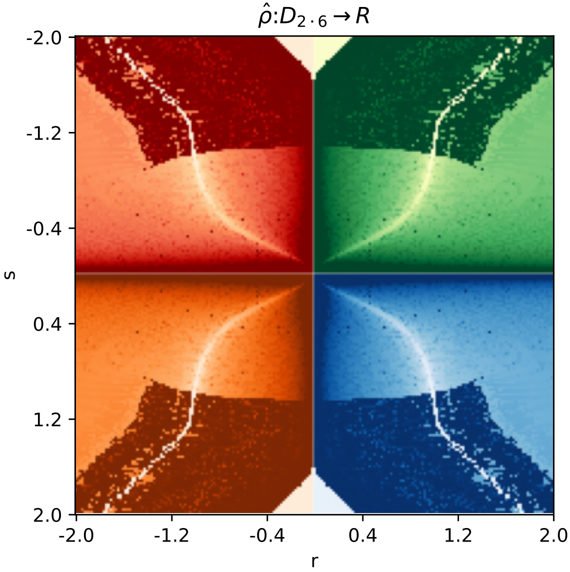


Showing 200 out of 5000 sampled trajectories

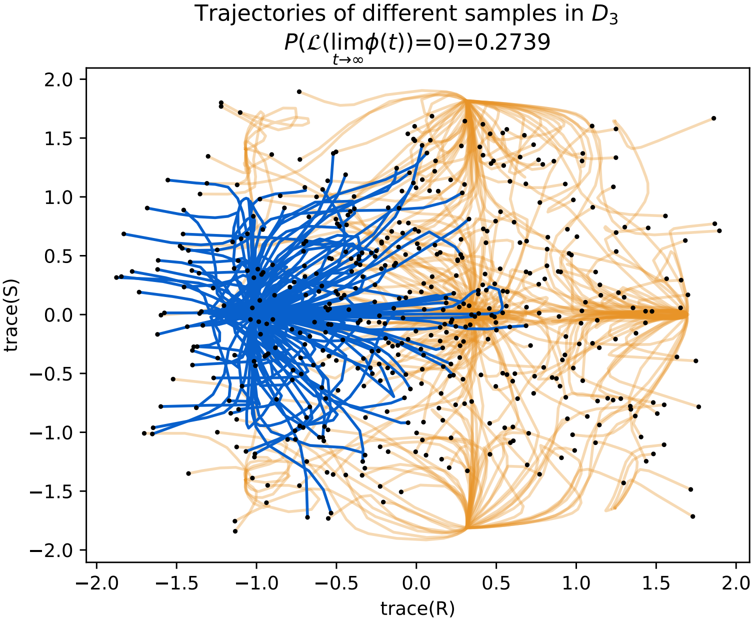


Each colormap encodes one limit point. Faster the convergence, lighter the color.

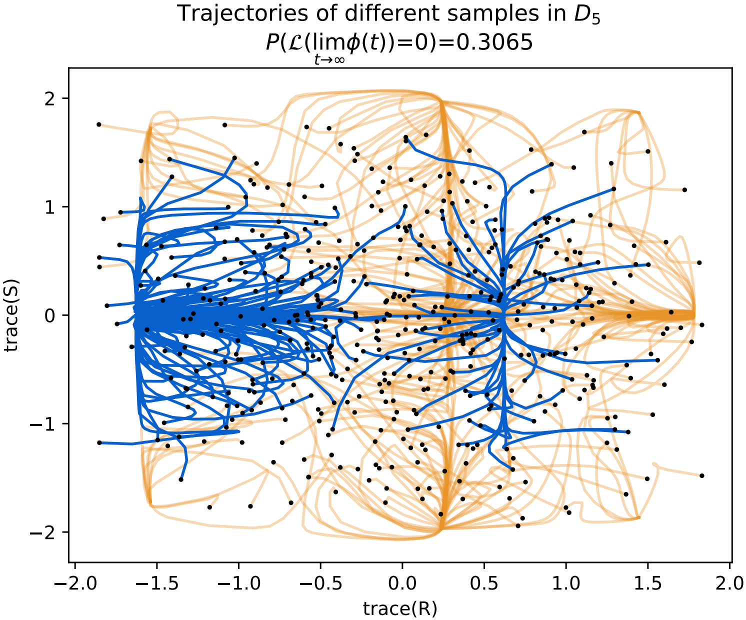
Limit points and convergence speed of initial parameters



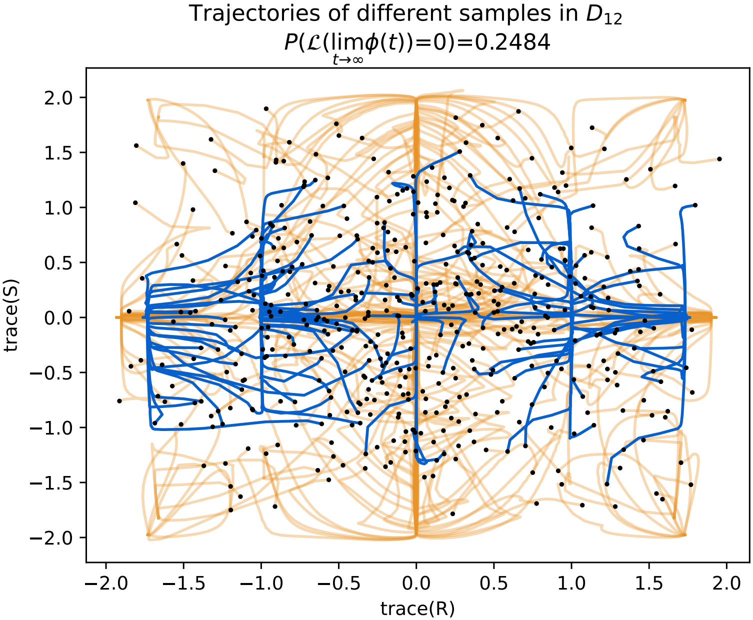
Each colormap encodes one limit point. Faster the convergence, lighter the color.



Sample size: 10^4. $\mathrm{tr}(R), \mathrm{tr}(S) \in (-2, 2)$.



Sample size: 10^4. $\mathrm{tr}(R), \mathrm{tr}(S) \in (-2, 2)$.



Sample size: 10^4. $\mathrm{tr}(R), \mathrm{tr}(S) \in (-2, 2)$.

Probability for success

Group	$P\left(\mathcal{L}(\lim_{t\to\infty}\phi(t)) = 0\right)$
D_3	0.274
D_4	0.158
D_5	0.306
D_6	0.21
D_7	0.311
D_8	0.235
D_9	0.33
D_{10}	0.241
D_{11}	0.327
D_{12}	0.248
D_{13}	0.319

Group actions

Map generators of $G =< S|R >$ to random maps on $[n]$.

$$\hat{\rho}\colon S \to \mathcal{P}(\mathrm{fun}([n],[n]))$$

$$s \mapsto P_s = \begin{bmatrix} P(s(1)=1) & P(s(1)=2) & \cdots & P(s(1)=n) \\ P(s(2)=1) & P(s(2)=2) & \cdots & P(s(2)=n) \\ \vdots & \vdots & & \vdots \\ P(s(n)=1) & P(s(n)=2) & \cdots & P(s(n)=n) \end{bmatrix}$$

For $f \in \mathrm{fun}([n],[n])$, define

$$P(s = f) = \prod_{i=1}^{n} P(s(i) = f(i)) = \prod_{i=1}^{n} s_{i,f(i)}$$

Construction details

For $g = s_1 s_2 \cdots s_m \in G$ define

$$P_g = P_{s_1} P_{s_2} \cdots P_{s_m} = \begin{bmatrix} P(g(1) = 1) & \cdots & P(g(1) = n) \\ \vdots & \vdots & \vdots \\ P(g(n) = 1) & \cdots & P(g(n) = n) \end{bmatrix}$$

Get stochastic matrices using $\operatorname{softmax}(v)_i = \frac{e^{v_i}}{\sum_{j=1}^{n} e^{v_j}}$

$$P \colon \mathbb{R}^{n \times n} \to \mathcal{S}_n$$
$$\begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \mapsto \begin{bmatrix} \operatorname{softmax}(a_1^T) \\ \vdots \\ \operatorname{softmax}(a_n^T) \end{bmatrix}$$

Make relations probable

For $r \in R$, we want $P(r = \mathrm{id}) = 1$

$$0 = \log(P(r = \mathrm{id})) = \log(\prod_{i=1}^{n} P(r(i) = i)) = \mathrm{tr}(\log(P_r))$$

$$\mathcal{L}_{\text{rel}} = -\frac{1}{|R|} \sum_{r \in R} \mathrm{tr}(\log(P_r))$$

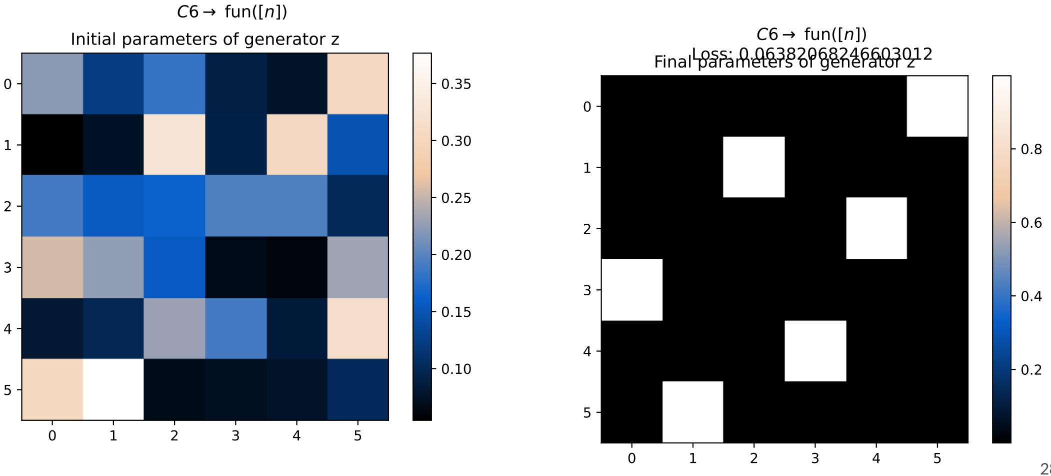
Make bijections probable

$$P(s \in S_n) = \sum_{\sigma \in S_n} P(s = \sigma) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} s_{i,\sigma(i)} = \text{Perm}(P_s).$$

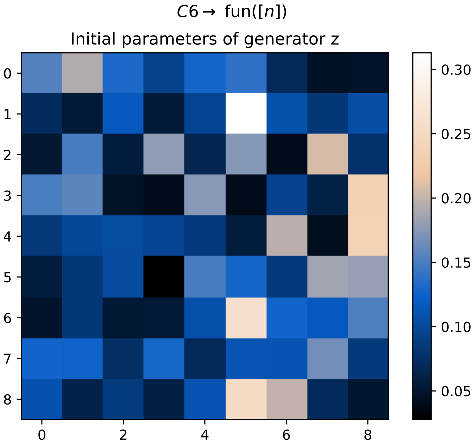
$$P(s \in S_n) = 1 \iff P_S \text{ is permutation matrix} \iff P_S \in U_n$$

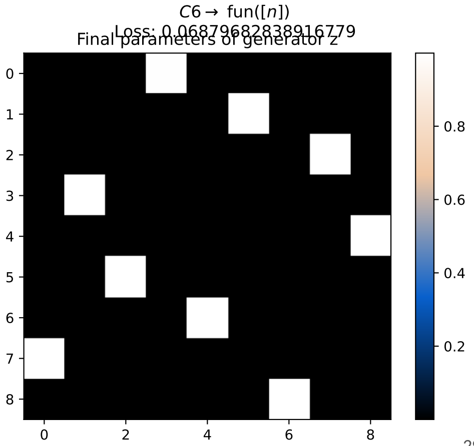
$$\mathcal{L}_{\text{bijective}} = \frac{1}{|S|} \sum_{s \in S} ||P_s P_s^T - I||_F^2 = \mathcal{L}_{\text{unitary}} \circ \mathrm{softmax}$$

Converges to (162354)

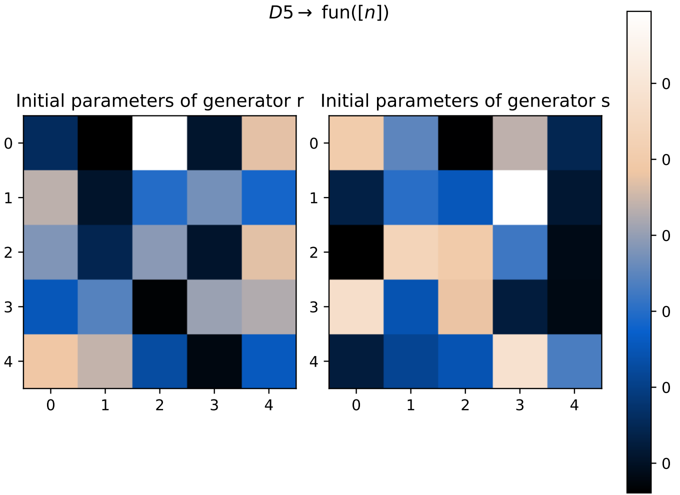


Converges to $(142638)(597)$

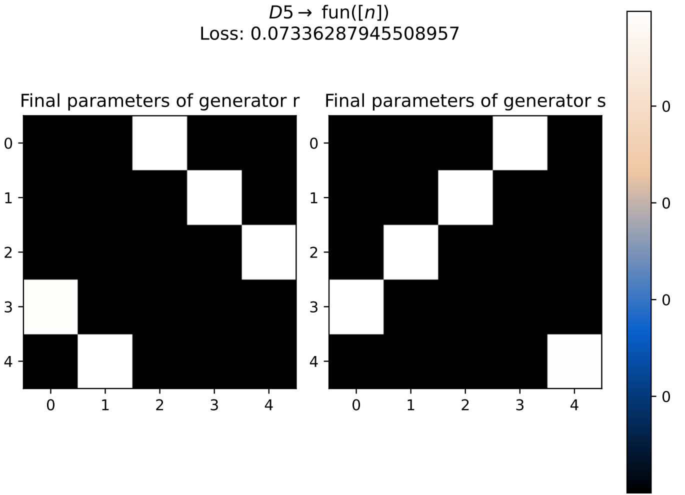




Converges to $r \xrightarrow{t\to\infty} (13524), s \xrightarrow{t\to\infty} (14)(23)$



Converges to $r \xrightarrow{t\to\infty} (13524),\, s \xrightarrow{t\to\infty} (14)(23)$



Graph isomorphisms

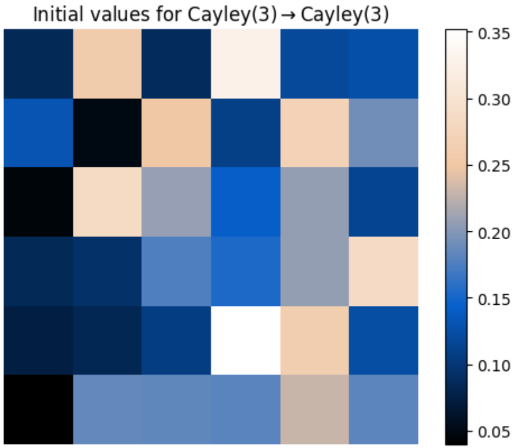
- G_1, G_2 graphs with adjacency matrices $M_1 = \left[m_{i,j}^{(1)}\right]$ and $M_2 = \left[m_{i,j}^{(1)}\right]$
- Random mapping $f : [n] \to [n] \sim P_f = \left[P(f(i) = j)\right]$
- For every $i \sim_1 j$, we want $P(f(i) \sim_2 f(j)) = 1$

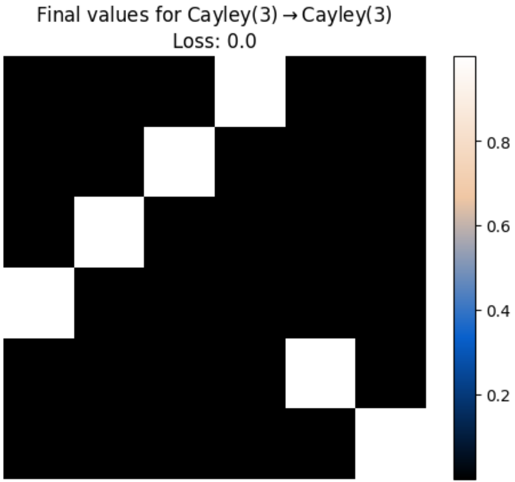
Relation loss

$$P(f(i) \sim_2 f(j)) = \sum_{k=1}^{n} \sum_{h=1}^{n} f_{i,k} f_{j,h} m_{k,h}^{(2)} = (P M_2 P^T)_{i,j}$$

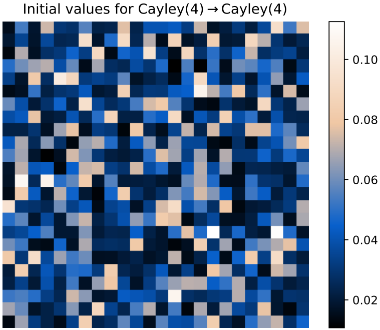
$$\begin{aligned} \mathcal{L}_{\text{rel}} &= - \sum_{i \sim_1 j} P(f(i) \sim_2 f(j)) \\ &= - \sum_{i=1}^{n} \sum_{j=1}^{n} log(f_j^T M_2 f_i) m_{i,j}^{(1)} \\ &= - \operatorname{tr}(\log(P_f M_2 P_f^T) M_1^T) \end{aligned}$$

P_f for f acting on $\mathrm{Cayley}(S_3)$

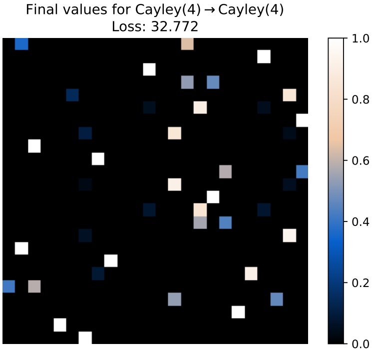




Automorphism of $\mathrm{Cayley}(S_4)$

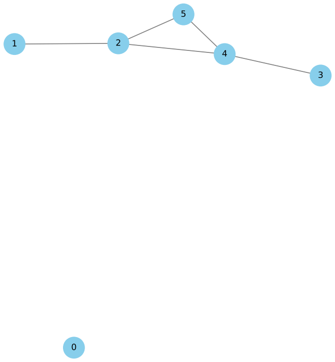


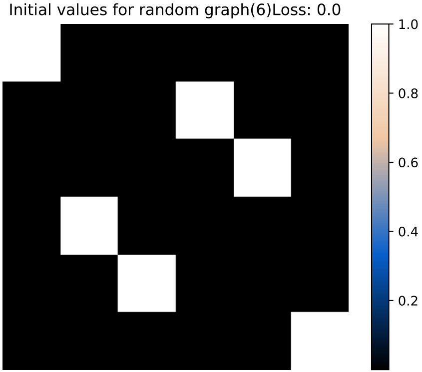
Caption



Caption

Automorphism of random graph





Further usage

- underparameterisation $x \mapsto \begin{bmatrix} \frac{x}{\sqrt{1-x^2}} & -\sqrt{1-x^2} \\ & x \end{bmatrix}$.
- overparameterisation
  $\phi_s = \mathrm{NeuralNetwork}(p), \dim(p) >> \dim(phi_s)$
- adaptive loss (Orthogonality of characters, adaptive dimension) - *bad results*
- paths in graphs, euler walks, ...

Implementation:

`github.com/urhprimozic/gofi`