

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Matematika – 2. stopnja

Urh Primožič

**ISKANJE MORFIZMOV Z GRADIENTNIM
SPUSTOM**

Magistrsko delo

Mentor: prof. dr. Ljupčo Todorovski

Somentor: doc. dr. Urban Jezernik

Ljubljana, 2025

Zahvala

Zahvaljujem se mentorjema za pomoč in vodenje ter prijateljem in družini za podporo čez študij.

Kazalo

Uvod	1
Oznake	1
1 Optimizacija z gradientnim spustom	2
1.1 Gradientni spust	2
1.2 Lastnosti gradientnega spusta	3
1.3 Moment	3
1.4 Adam	3
1.5 Preturbniran gradientni spust	3
2 Iskanje upodobitev	5
2.1 Upodobitve končnih grup	5
2.2 Iskanje upodobitev z gradientnim spustom	6
2.2.1 Funkcija izgube nad relacijami	7
2.2.2 Funkcija izgube za nerazcepnost	8
2.2.3 Funkcija izgube za unitarnost	8
2.2.4 Gradientni tok	8
2.3 Obstoj rešitve	9
2.4 Ciklične grupe	12
2.5 Redukcija funkcij izgube	13
2.6 Diedrske grupe	14
2.6.1 Enodimenzionalne upodobitve	14
2.6.2 Dvodimenzionalne upodobitve	15
2.7 Splošna dinamika	18
3 Iskanje delovanj	19
3.1 Parametrizacija stohastičnih matrik	20
3.2 Funkcija izgube nad relacijami	20
3.3 Pretvorba modela v delovanje	21
3.4 Delovanje ciklične grupe	24
3.5 Delovanje diedrske grupe	24
4 Iskanje izomorfizmov grafov	27
4.1 Funkcija izgube za bijektivnost	27
4.2 Funkcija izgube za povezanost	28
4.3 Prehod na izomorfizem	29
4.4 Rezultati	30
4.5 Uporaba tabele inverzij	31
4.5.1 Porazdelitve nad tabelo inverzij	32
4.5.2 Funkcija izgube	33
4.5.3 Hiter izračun funkcije izgube	33
4.5.4 Preparametrizacija z nevronskimi mrežami	35

A Nevronske mreže	38
A.1 Nevronske mreže dosežajo globalni minimum funckije izgube	38
A.2 Preparametrizacija z grafovskimi nevronskimi mrežami in algoritmom Sinkhorn	39
B Rezultati simulacij	40
B.1 Upodobitve	40
B.1.1 Diedrska grupa	40
B.2 Delovanja	48
B.2.1 Ciklične grupe	48
B.2.2 Diedrske grupe	50
B.3 Grafi	50
B.3.1 Avtomorfizmi cayleyevih grafov	50
Literatura	57

Iskanje morfizmov z gradientnim spustom

POVZETEK

V delu predstavimo metodo iskanja nerazcepnih upodobitev končnih grup z gradientnim spustom. Preslikave med končnimi grupami in matrikami predstavimo z vektorji v evklidskem prostoru in definiramo nenegativno funkcijo izgube, ki doseže nič natanko v nerazcepnih unitarnih upodobitvah.

Predstavimo povezavo med gradientnim spustom in reševanjem diferencialnih enačb. Vsako nerazcepno upodobitev izrazimo kot limito gradientnega toka gladke funkcije z ugodnimi začetnimi parametri in preučujemo numerične rezultate za ciklične in diedrske grupe.

Metodo razširimo na iskanje delovanj in izomorfizmov grafov. Definiramo gladko družino distribucij nad preslikavami končnih množic in maksimaliziramo verjetnost, da je slučajna preslikava morfizem. Definiramo gladko družino distribucij nad tabelo inverzij, ki ponuja alternativo Sinkhornovemu algoritmu.

Finding morphisms with gradient descent

ABSTRACT

A method for finding irreducible representations of finite groups using gradient descent is presented. Mappings between finite groups and matrices are represented as vectors in Euclidean space and a non-negative loss function is defined, which is zero exactly at irreducible unitary representations.

The connection between gradient descent and solving differential equations is presented. Each irreducible representation is expressed as a limit of the gradient flow of a smooth function with suitable initial parameters. Numerical results for cyclic and dihedral groups are studied.

The method is extended to finding actions and graph isomorphisms via a smooth family of distributions over mappings of finite sets. A smooth family of distributions over the inversion table is defined, offering an alternative to the Sinkhorn algorithm.

Math. Subj. Class. (2020): 65K10, 68T05, 20C99, 68T09

Ključne besede: gradientni spust, upodobitve, izomorfizmi grafov, delovanja, tabela inverzij

Keywords: gradient descent, representations, graph isomorphisms, actions, inversion table

Uvod

Nerazcepne upodobitve, delovanja in izomorfizme grafov lahko enačimo z množico ničel gladke funkcije $\mathcal{L}: \mathbb{R}^n \rightarrow [0, \infty)$ nad evklidskim prostorom. S tem iskanje diskretnih struktur prevedemo na optimizacijo gladkih funkcij.

Optimizacije se lotimo z gradientnim spustom, temeljno optimizacijsko metodo v strojenem učenju, ki je ekvivalentna numeričnemu reševanju diferencialne enačbe $\frac{d\phi}{dt} = -\nabla \mathcal{L}(\phi)$. Študiramo dinamiko rešitev in odvisnost konvergencije od izbire začetnih parametrov.

V poglavju 1 predstavimo gradientne optimizacijske metode in povezavo med gradientnim spustom in reševanjem diferencialnih enačb. V poglavju 2 predstavimo potrebno teoretično ozadje iz teorije upodobitev končnih grup. Uvedemo metodo iskanja nerazcepnih upodobitev z gradientnim spustom in predstavimo empirične rezultate na cikličnih in diedrskih grupah. V poglavju 3 definiramo metodo za iskanje delovanj, v poglavju 4 pa metodo za iskanje izomorfizmov grafov. Iskanje izomorfizmov grafov povežemo s problemom ujemanja grafov. V poglavju ?? metodo razširimo z uporabo tabele inverzij, preko katere kandidate za izomorfizem iščemo med distribucijami nad množico permutacij S_n in podamo empirične rezultate metode v kombinaciji s preparametrizacijo z nevronskimi mrežami.

Vse definirame metode teoretično obravnavamo in dokažemo, da ob dovolj majhnih vrednosti $\mathcal{L}(\phi)$ metoda najde iskano strukturo.

Metode v delu so originalni rezultati avtorja, mentorja in somentorja. Podoben pristop, kot ga uporabimo za iskanje nerazcepnih upodobitev končnih grup, v članku [4] izkoriščajo za učenje nevronske mreže, ki predstavlja grupo. Gladko distribucijo nad tabelo inverzij v strojnem učenju študirajo v [24] in v [6]. Preparametrizacijo z nevronskimi mrežami za iskanje izomorfizmov grafov predstavijo v [16].

Oznake

S_n	Množica permutacij števil $1, 2, \dots, n$.
$[n]$	Množica števil $\{1, 2, \dots, n\}$.
flatten	Sploščitev matrike flatten: $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n^2}$. Za $A = [a_{i,j}] \in \mathbb{R}^{n \times n}$ je flatten(A) $_i = a_{\lfloor \frac{i}{n} \rfloor, i \bmod n}$.
softmax	Funkcija softmax: $\mathbb{R}^n \rightarrow \mathbb{R}^n$. Glej definicijo 3.2.
Π	Množica permutacijskih matrik.
F_S	Prosta grupa nad množico S .
A^*	Prosti monoid nad množico A .
diag($\sigma_1, \sigma_2, \dots, \sigma_n$)	Diagonalna matrika z elementi $\sigma_1, \sigma_2, \dots, \sigma_n$ na diagonali.
$\ A\ $	Frobeniusova norma matrike A .
$B_r(a)$	Krogla okoli a s polmerom r . Velja $B_r(a) = \{x \mid x - a < r\}$.

Oznake matrik

Za matriko $A = [a_{i,j}] \in \mathbb{R}^{n \times n}$ označimo i, j -ti element z $A_{i,j} = a_{i,j}$, i -to vrstico z $a_i = [a_{i,1}, a_{i,2}, \dots, a_{i,n}]$ in j -ti stolpec z $a^j = [a_{1,j}, a_{2,j}, \dots, a_{n,j}]^T$.

1 Optimizacija z gradientnim spustom

Na različne probleme v matematiki lahko gledamo kot na iskanje minimuma $\hat{\phi} = \operatorname{argmin}(\mathcal{L}(\phi) \mid \phi)$ funkcionalna \mathcal{L} . Če je domena funkcionalna evklidski prostor, $\mathcal{L}: \mathbb{R}^n \rightarrow \mathbb{R}$ pa gladka, lahko minimum iščemo z gradientnimi metodami optimizacije.

V tem poglavju definiramo numerične gradientne optimizacijske metode. Definicije sledijo [21].

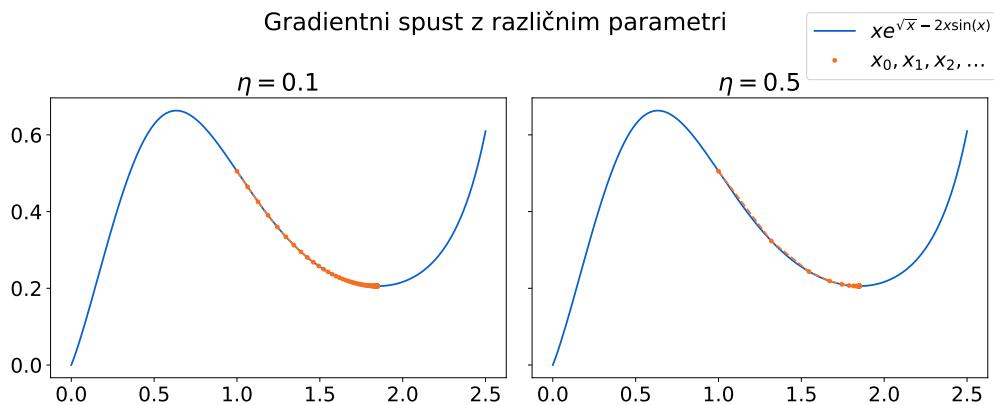
1.1 Gradientni spust

Definicija 1.1. Naj bo $\mathcal{L}: \mathbb{R}^n \rightarrow \mathbb{R}$ odvedljiva funkcija in $\phi_0 \in \mathbb{R}^n$ poljubna začetna vrednost parametrov. *Gradientni spust* je iterativna optimizacijska metoda, definirana s pravilom

$$\phi_{i+1} = \phi_i - \eta \nabla \mathcal{L}(\phi_i) \quad (1.1)$$

Parametru $\eta > 0$ rečemo stopnja ali hitrost učenja (*learning rate*) gradientnega spusta in vpliva na velikost spremembe parametrov v posameznem koraku.

Primer 1.2. Na sliki 1 sta prikazana grafa rezultatov gradientnega spusta na funkciji $f(x) = xe^{\sqrt{x}-2x\sin(x)}$ z začetnim parametrom $x_0 = 1$ in različnima hitrostima učenja. Spremembe parametrov sledijo smeri padanja krivulje in so večje, če krivulja pada hitreje. ◇



Slika 1: Sto korakov gradientnega spusta na funkciji $f(x) = xe^{\sqrt{x}-2x\sin(x)}$ z začetnim parametrom $x_0 = 1$ in hitrostjo učenja η . V obeh primerih metoda konvergira do lokalnega minimuma funkcije.

Definicija 1.3. Naj bosta \mathcal{L} in ϕ kot zgoraj. *Gradientni tok* je rešitev diferencialne enačbe

$$\frac{d\phi}{dt} = -\nabla \mathcal{L}(\phi), \quad \phi(0) = \phi_0. \quad (1.2)$$

Gradientni spust ustrezza eksplisitni Eulerjevi diskretizaciji diferencialne enačbe gradientnega toka *gradientnega toka* (glej [5]).

Opomba 1.4. V globokem učenju [13] se za optimizacijo uporablja metode, ki sicer temeljijo na gradientnem spustu, ampak niso numerični ekvivalent reševanja gradientnega toka. Med drugim lahko z uporabo šuma v izračunu gradienata dobimo metodo, ki z visoko verjetnostjo pobegne iz sedel [12]. Te lastnosti nima nobena rešitev enačb gradientnega toka.

Alternativno bi se lahko optimizacije lotili z boljšimi numeričnimi metodami za reševanje diferencialnih enačb (glej [5] za učenje z metodo Runge-Kutta).

1.2 Lastnosti gradientnega spusta

V primeru 1 metoda konvergira do lokalnega minimuma funkcije. V splošnem je konvergenca odvisna od izbire η in začetnih vrednosti parametrov ϕ_0 .

Rešitev $\phi(t)$ gradientnega toka $\frac{d\phi}{dt} = -\nabla \mathcal{L}(\phi)$, $\phi(0) = \phi_0$ vedno konvergira do kritične točke \mathcal{L} [10, stran 203]. Limita $\lim_{t \rightarrow \infty} \phi(t)$ je odvisna od izbire začetnih parametrov ϕ_0 .

1.3 Moment

Kljub gladkosti funkcije \mathcal{L} so lahko spremembe parametrov pri uporabi gradientnega spusta kaotične. lahko jih delno omejimo z uporabo *momenta* ([21, stran 86])

$$\begin{aligned} m_{i+1} &= \beta \cdot m_t + (1 - \beta) \nabla \mathcal{L}(\phi_i) \\ \phi_{i+1} &= \phi_i - \eta m_{i+1}, \end{aligned}$$

kjer sta $\beta \in [0, 1]$, $\eta > 0$ parametra metode. Sprememba parametra $\phi_{i+1} - \phi_i$ je s tem manj kaotična in bolj gladka.

1.4 Adam

Adam ([21, stran 88]) je iteracijski optimizacijski algoritem, podan s pravili

$$\begin{aligned} m_{i+1} &= \beta \cdot m_i + (1 - \beta) \nabla \mathcal{L}(\phi_i) \\ \bar{m}_{i+1} &= \frac{m_{i+1}}{1 - \beta^{i+1}} \\ v_{i+1} &= \gamma \cdot v_t + (1 - \gamma) (\nabla \mathcal{L}(\phi_i))^2 \\ \bar{v}_{i+1} &= \frac{v_{i+1}}{1 - \delta^{i+1}} \\ \phi_{i+1} &= \phi_i - \eta \frac{\bar{m}_{i+1}}{\bar{v}_{i+1} + \epsilon}, \end{aligned}$$

kjer so $\eta > 0$, $\beta \in [0, 1]$, $\delta \in [0, 1]$, $\gamma \in [0, 1]$ parametri modela.

Zaradi dobrih empiričnih rezultatov se v praksi v strojenem učenju namesto surovega gradientnega spusta za optimizacijo uporablja algoritem Adam.

1.5 Preturbiran gradientni spust

Gradientni spust lahko obtiči v sedlih. Iz sedla lahko pridemo z dodajanjem naključne perturacije parametrov, ko je gradienat ciljne funkcije majhen [12].

Za $\nabla \mathcal{L}(\phi) < \epsilon$ definiramo

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

kjer je $\xi_i \sim U[B_r(0)]$ enakomerno porazdeljen šum $|\xi_i| < r$ za izbrana hiperparametra r in ϵ .

2 Iskanje upodobitev

Obstaja očitna¹ $1 : 1$ korespondenca med preslikavami končnih grup v evklidske prostore $f: G \rightarrow \mathbb{R}^{n \times n}$ in prostorom $K = \mathbb{R}^{n \times n^{|G|}} \cong \mathbb{R}^{n^2 |G|}$.

Definiramo lahko gladko funkcijo $\mathcal{L}: K \rightarrow [0, \infty)$, za katero je $\mathcal{L}(x) = 0$ natanko tedaj, ko x predstavlja upodobitev končne grupe. Iskanje upodobitev lahko prevedemo na minimiziranje funkcije $\mathcal{L}(f)$.

V poglavju 2.1 predstavimo potrebno teoretično ozadje. V 2.2 natančno definiramo metodo za iskanje upodobitev. V poglavjih 2.4 in 2.6 študiramo dinamiko diferencialnih enačb gradientnega toka za ciklične in diedrske grupe.

2.1 Upodobitve končnih grup

Jezik teorije upodobitev je povzet po [11].

Definicija 2.1. Naj bo G grupa in V vektorski prostor nad poljem \mathbb{F} . *Upodobitev grupe G* je homomorfizem $\rho: G \rightarrow \mathrm{GL}(V)$ med grupo G in endomorfizmi prostora V .

Vsaka upodobitev $\rho: G \rightarrow \mathrm{GL}(V)$ naravno definira linearno delovanje $v^g = \rho(g)(v)$ grupe nad V . Ekvivalentno vsako linearno delovanje definira upodobitev $\rho(g) = (v \mapsto v^g)$.

Definicija 2.2. Upodobitvi $\rho: G \rightarrow \mathrm{GL}(V)$ in $\sigma: G \rightarrow \mathrm{GL}(W)$ sta *ekvivalentni*, če obstaja izomorfizem vektorskih prostorov $\Phi: V \rightarrow W$, da za vsak $g \in G$ in vsak $v \in V$ velja

$$\Phi(\rho(g)(v)) = \sigma(g)(\Phi(v)).$$

Preslikavi Φ rečemo *spletična* med V in W .

Definicija 2.3. Naj bo $\rho: G \rightarrow \mathrm{GL}(V)$ upodobitev in $W < V$ podprostor prostora V , invarianten za inducirano delovanje. Potem je upodobitev $\rho_W: G \rightarrow \mathrm{GL}(W)$, dana s predpisom $\rho_W(g) = \rho(g)|_W$ podupodobitev upodobitve ρ .

Definicija 2.4. Če sta edini podupodobitvi upodobitve $\rho: G \rightarrow \mathrm{GL}(V)$ $g \mapsto \mathrm{id}$ in ρ , je ρ *nerazčlena upodobitev*. Množico nerazčlenih upodobitev grupe G označimo z $\mathrm{Irr}(G)$.

Definicija 2.5. Upodobitev $\rho: G \rightarrow \mathrm{GL}(V)$ je *polenostavna*, če jo lahko izrazimo kot vsoto nerazčlenih upodobitev $\rho = \bigoplus_{i \in I} \rho_i$.

Če se omejimo le na končne grupe in dovolj lepa polja, so vse upodobitve sestavljene iz nerazčlenih.

Izrek 2.6. *Naj bo G končna grupa in \mathbb{F} polje. Vse upodobitve G nad poljem F so polenostavne če in samo če $\mathrm{char} F \nmid |G|$.*

Za grupo G množico njenih neizomorfnih nerazčlenih upodobitev označimo z $\mathrm{Irr}(G)$, za poljubno upodobitev ρ pa definiramo *normo karakterja*

$$|\chi_\rho| = \frac{1}{|G|} \sum_{g \in G} \mathrm{tr}(\rho(g)) \mathrm{tr}(\rho(g^{-1})). \quad (2.1)$$

¹Za $G = \{g_1, g_2, \dots, g_m\}$ lahko enačimo $f = (\mathrm{flatten}(f(g_1)), \mathrm{flatten}(f(g_2)), \dots, \mathrm{flatten}(f(g_m)))$.

Izrek 2.7. Za končno grupo G in polenostavno upodobitev $\rho: G \rightarrow \mathrm{GL}_n(\mathbb{C})$ velja

$$|\chi_\rho| = \sum_{\pi \in \mathrm{Irr}(G)} \mathrm{mult}_\rho(\pi)^2,$$

kjer je $\mathrm{mult}_\rho(\pi) \in \mathbb{N}$ število kopij nerazcepne upodobitve π v razcepu upodobitve $\rho = \bigoplus_{\pi \in \mathrm{Irr}(G)} \mathrm{mult}_\rho(\pi)\pi$.

Izrek 2.8. Naj bo G končna grupa in $\rho: G \rightarrow \mathrm{GL}(V)$ končno-razsežna upodobitev nad algebraično zaprtim poljem s karakteristiko 0. Potem velja

$$\rho \in \mathrm{Irr}(G) \iff |\chi_\rho| = \frac{1}{|G|} \sum_{g \in G} \mathrm{tr}(\rho(g)) \mathrm{tr}(\rho(g^{-1})) = 1. \quad (2.2)$$

Za študij upodobitev končnih grup nad ugodnimi polji je torej dovolj poznati le končno² množico nerazcepnih upodobitev $\mathrm{Irr}(G) = \{\rho: G \rightarrow \mathrm{GL}(V) \mid |\chi_\rho| = 1\}$.

2.2 Iskanje upodobitev z gradientnim spustom

Naj bo $G = \langle S | R \rangle$ končna grupa nad množico generatorjev S , definirana z relacijami R . Na S definiramo poljubno preslikavo

$$\hat{\rho}: S \rightarrow \mathbb{C}^{n \times n} \quad (2.3)$$

$$s \mapsto \begin{bmatrix} s_{1,1} & \cdots & s_{1,n} \\ \vdots & & \vdots \\ s_{n,1} & \cdots & s_{n,n} \end{bmatrix}, \quad (2.4)$$

ki vsak generator grupe preslika v poljubno matriko.

Definicija 2.9. Preslikavo $\hat{\rho}$ razširimo na S^* s predpisom

$$\hat{\rho}(s_1 s_2 \cdots s_m) = \hat{\rho}(s_1) \hat{\rho}(s_2) \cdots \hat{\rho}(s_m).$$

Opomba 2.10. Ker so vse grupe v tem delu končne³, lahko za vsak element grupe G izberemo predstavnika $w_g \in S^*$, da velja $w_g \equiv g \pmod{G}$. S tem lahko za vsak element $g \in G$ definiramo preslikavo $\hat{\rho}(g)$ kot $\hat{\rho}(w_g)$

Definicija 2.11. Preslikavo $\hat{\rho}$ definiramo na grapi G na slednji način:

- Za vsak $g \in G \setminus \{1\}$ izberemo kanoničnega predstavnika $w_g \in S^*$.
- Za $1 \in G$ definiramo w_1 kot prazno besedo.

Za $g \in G$ s kanoničnim predstavnikom $w_g \in S^*$ definiramo

$$\hat{\rho}(g) = \hat{\rho}(w_g).$$

Preslikava $\hat{\rho}$ ni nujno upodobitev, služi pa nam lahko kot model upodobitve, odvisen od parametrov $\phi = \{s_{i,j} \mid s \in S, 1 \leq i, j \leq n\}$. Parametre matrik $\hat{\rho}(s)$ bomo postopoma spreminali s gradientnim spustom, da bo model vse bližje upodobitvi.

²Glej [11, poglavje Dekompozicija regularne upodobitve].

³V splošnem je $w_g \in (S \cup S^{-1})^*$. V končnih grupah za $s \in S$ velja $s^{-1} = s^{\mathrm{red}(s)-1}$.

2.2.1 Funkcija izgube nad relacijami

Preslikava $\hat{\rho}$ je homomorfizem natanko tedaj, ko za vsako relacijo $r \in R$ velja $\hat{\rho}(r) = I$. Manjše, kot so norme $\|\hat{\rho}(r) - I\|$, bolj je model $\hat{\rho}$ podoben homomorfizmu.

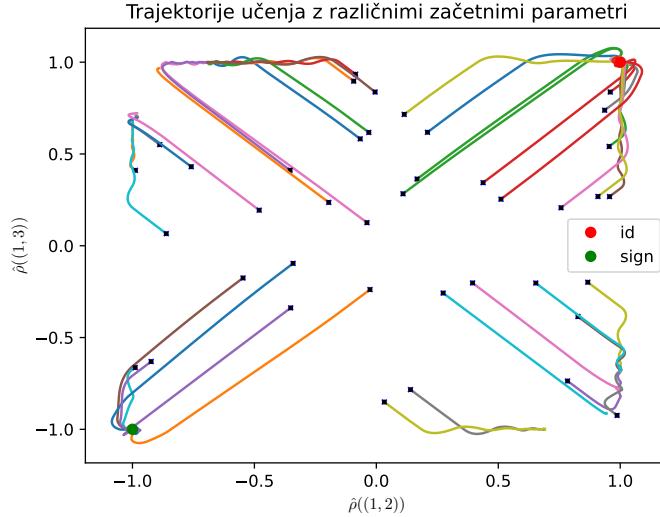
Definicija 2.12. *Funkcija izgube nad relacijami* je definirana kot

$$\mathcal{L}_{\text{rel}}(\hat{\rho}) = \frac{1}{|R|} \sum_{r \in R} \|\hat{\rho}(r) - I\|. \quad (2.5)$$

Očitno je model $\hat{\rho}$ upodobitev natanko tedaj, ko je $\mathcal{L}_{\text{rel}}(\hat{\rho}) = 0$.

Primer 2.13. Poglejmo si model $\hat{\rho}: S_3 \rightarrow \mathbb{R}$ eno dimenzijske realne upodobitve permutacijske grupe na treh točkah za prezentacijo $S_3 = < (12), (13) \mid (12)^2 = (13)^2 = (13)^{(12)}(12)^{(13)} = 1 >$. Model je definiran z vrednostmima $\hat{\rho}((12)) = x$ in $\hat{\rho}((13)) = y$.

Edini enodimensionalni upodobitvi S_3 sta identiteta id in predznak permutacije sign. Na sliki 2 so trajektorije gradientnega spusta pri optimizaciji \mathcal{L}_{rel} za različne začetne parametre (x_0, y_0) . Opazimo, nekatere trajektorije konvergirajo do id nekatere do sign, nekatere do modela, ki ni upodobitev. V tem primeru je gradientni spust našel lokalni ekstrem funkcije izgube, ki ni globalni. \diamond



Slika 2: Različne trajektorije učenja modela $\hat{\rho}: S_3 \rightarrow \mathbb{R}$. Začetni parametri vplivajo na uspeh optimizacije in končno upodobitev.

Opomba 2.14. Model bi lahko definirali na celi grupi kot $\hat{\rho}(g) = [g_{i,j}]_{i,j=1,\dots,n}$, za parametre pa vzeli elemente matrik $\hat{\rho}(g)$. V tem primeru je model homomorfizem, če $\forall g, h \in G. \hat{\rho}(gh) = \hat{\rho}(g)\hat{\rho}(h)$. Za funkcijo izgube lahko vzamemo

$$\mathcal{L}_{\text{homo}} = \frac{1}{|G|^2} \sum_{i=1}^n \sum_{j=1}^n \|\hat{\rho}(gh) - \hat{\rho}(g)\hat{\rho}(h)\|.$$

in iskanje upodobitev spet prevedemo na optimizacijski problem.

Ta pristop ni praktičen, saj privede do velikega števila parametrov $N = |G|^n$. Model upodobitve permutacijske grupe S_n bi tako imel $(n!)^2$ parametrov, kar je neugodno za numerične simulacije. Če pa sledimo pristopu s prezentacijami in vza-memo minimalno množico generatorjev (recimo tranzpozicija in n -cikel), operiramo le s štirimi kompleksnimi parametri, neodvisno od velikosti grupe.

2.2.2 Funkcija izgube za nerazcepnot

Iščemo lahko le nerazcepne upodobitve. Vemo že (izrek 2.2), da je upodobitev ρ nerazcepna, če je norma njenega karakterja χ_ρ enaka 1.

Definicija 2.15. Definiramo funkcijo izgube za nerazcepnot

$$\mathcal{L}_{\text{irr}}(\rho) = (|\chi_\rho| - 1)^2. \quad (2.6)$$

Za poljuben model $\hat{\rho}$ lahko z gradientnim spustom minimaliziramo funkcijo izgube $\mathcal{L}_{\text{rel}} + \mathcal{L}_{\text{irr}}$ in iščemo *nerazcepne* upodobitve.

2.2.3 Funkcija izgube za unitarnost

Dodatno lahko zahtevamo, da naš model slika v unitarne matrike. Velja namreč⁴, da je vsaka upodobitev ekvivalentna upodobitvi v unitarne matrike. S tem zmanjšamo število različnih upodobitev, h katerimi lahko konvergiramo.

Definicija 2.16. Funkcija izgube za unitarnost je

$$\mathcal{L}_{\text{unitary}} = \frac{1}{|S|} \sum_{s \in S} \|\hat{\rho}(s)\hat{\rho}(s)^H - I\|. \quad (2.7)$$

Če ne piše drugače, od tu naprej vedno privzamemo, da minimiziramo funkcijo izgube $\mathcal{L} = \mathcal{L}_{\text{rel}} + \mathcal{L}_{\text{irr}} + \mathcal{L}_{\text{unitary}}$. Globalni minimumi te funkcije so natanko unitarne nerazcepne upodobitve.

2.2.4 Gradientni tok

V delu opazujemo dinamiko sistema $\frac{d\phi}{dt} = -\nabla \mathcal{L}$. Zanimajo nas limite trajektorij in meje med območji začetnih parametrov z različnimi limitami.

Diferencialne enačbe gradientnih tokov imajo nekatere lepe lastnosti. Limite različnih trajektorij so natanko ničle $\nabla \mathcal{L}(\phi) = 0$ gradienta ([10, poglavje 9.3]).

Opomba 2.17. Enačbe gradientnih tokov nimajo vedno enostavnih rešitev v zaprti obliki [18]. Običajno se jih računa numerično.

V delu enostavne⁵ gradientne tokove (ciklične in diedrske grupe) integriramo s Runge-Kutto (s pythonovo metodo `scipy.integrate.solve_ivp()`), enačbe zapletenih gradientnih tokov rešujemo z avtograd implementacijo Adama (pythonova knjižnica `torch`).

⁴Glej [11, trditve 3.34].

⁵Za uporabo Runge-Kutte moramo poznati gradient $\nabla \mathcal{L}$. Eksplicitni izračun $\nabla \mathcal{L}$ z naraščajočim številom parametrov postane težaven in nepraktičen. V tem primeru se v delu numerične integracije lotimo z avtogradom, ki sam izračuna gradient.

2.3 Obstoj rešitve

Če je napaka modela $\mathcal{L}_{\text{rel}}(\hat{\rho})$ velika, smo zagotovo stran od upodobitve. Zanima nas, če velja obrat: ali iz majhne napake modela $\hat{\rho}$ lahko sklepamo na obstoj upodobitve ρ , ki je blizu modelu $\hat{\rho}$?

Za smiselnost zgornje metode iskanja upodobitev mora biti odgovor pozitiven. Z numerično optimizacijo v splošnem najdemo le model z napako $\mathcal{L}_{\text{rel}}(\hat{\rho}) > 0$. Le tako bo model $\hat{\rho}$ z majhno napako zagotovo dober približek upodobitve. To nam zagotavlja spodnji izrek.

Izrek 2.18 (Obstoj rešitve). *Za vsak $\epsilon > 0$ obstaja $\delta > 0$, da iz $\mathcal{L}_{\text{rel}}(\hat{\rho}) < \delta$ sledi, da obstaja nerazcepna upodobitev $\rho: G \rightarrow GL_n(\mathbb{C})$, za katero velja*

$$\sup_{s \in S} \|\hat{\rho}(s) - \rho(s)\| \leq \epsilon.$$

Velja še $\delta = o(\epsilon)$.

Izrek dokažemo s pomočjo rezultata iz [8, Theorem II.5 na strani 7].

Izrek 2.19 (Kazhdan). *Naj bo G končna⁶ grupa. Za $\epsilon < \frac{1}{100}$ in preslikavo $\mu: G \rightarrow U(\mathcal{H})$ v unitarne operatorje na Hilbertovem prostoru \mathcal{H} , za katero velja*

$$\sup_{g,h \in G} \|\mu(gh) - \mu(g)\mu(h)\| < \epsilon,$$

obstaja upodobitev $\pi: G \rightarrow U(\mathcal{H})$, da velja

$$\sup_{g \in G} \|\pi(g) - \mu(g)\| \leq 2\epsilon.$$

Izrek 2.18 dokažemo po korakih. Za začetek pokažemo, da iz dovolj majhne vrednosti $\mathcal{L}_{\text{unitary}}(\hat{\rho})$ lahko sklepamo, da je $\hat{\rho}(s)$ zelo blizu unitarne matrike za vsak $s \in S$.

Lema 2.20. *Iz $\mathcal{L}_{\text{unit}}(\hat{\rho}) < |S|\epsilon$ sledi, da je $\|\hat{\rho}(s) - Q_s\| < \epsilon$ za neko unitarno matriko Q_s .*

Dokaz. Naj bo $s \in S$ poljuben. Matriko $\hat{\rho}(s)$ lahko s *polarno dekompozicijo* izrazimo kot $\hat{\rho}(s) = Q_s H_s$, kjer je Q_s unitarna matrika in $H_s = (\hat{\rho}(s)\hat{\rho}(s)^*)^{\frac{1}{2}} \geq 0$.

Iz $\hat{\rho}(s) - Q_s = Q_s(H_s - I)$ sledi

$$\|\hat{\rho}(s) - Q_s\| = \|H_s - I\|,$$

ker je matrika Q_s unitarna.

Z $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ označimo singularne vrednosti matrike $\hat{\rho}(s)$. Očitno so to natanko lastne vrednosti matrike H_s .

Ker je H_s pozitivno semidefinitna, velja $H_s = V \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) V^*$ za neko unitarno matriko V . Računamo

$$\|H_s - I\| = \|\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n) - I\| = \sum_{i=1}^n (\sigma_i - 1)^2.$$

⁶Izrek v splošnem velja za vse *amenabilne* grupe.

Podobno je

$$\|\hat{\rho}(s)\hat{\rho}(s)^* - I\| = \sum_{i=1}^n (\sigma_i^2 - 1)^2$$

in očitno velja $|\sigma_i - 1| \leq |\sigma_i^2 - 1|$. Iz tega sledi

$$\|\hat{\rho}(s) - Q_s\| = \|H_s - I\| \leq \|\hat{\rho}(s)\hat{\rho}(s)^* - I\| < \epsilon.$$

□

Če je $\mathcal{L}_{\text{unitary}}(\hat{\rho}) < \epsilon$, lahko zapišemo $\hat{\rho}(s) = Q_s + F_s$ za $\|F_s\| < \epsilon$. Iz tega sledi ocena

$$\|\hat{\rho}(s)\| \leq \|Q_s\| + \|F_s\| < \sqrt{n} + \epsilon.$$

Z njo dokažemo, da model z majhno napako različne zapise iz S^* istega elementa grupe preslika blizu. Iz tega bo sledilo, da je model skoraj homomorfizem.

Lema 2.21. *Naj bo $\hat{\rho}$ tak, da velja $\mathcal{L}_{\text{rel}}(\hat{\rho}) < |R|\epsilon$ in $\mathcal{L}_{\text{unitary}}(\hat{\rho}) < |S|\epsilon$. Potem za besedo $w \in S^*$, ki predstavlja $1 \in G$, velja*

$$\|\hat{\rho}(w) - I\| < \left(1 + (\sqrt{n} + \epsilon)^{|w|}|w|\epsilon\right)^L - 1$$

Dokaz. Vse besede w , za katere velja $w \equiv 1 \pmod{G}$, so generirane z besedami oblike vru , kjer je $r \in R$ in $vu \equiv 1 \pmod{G}$. Najprej za te generatorje pokažemo, da velja $\|\hat{\rho}(vru) - I\| < (\sqrt{n} + \epsilon)^{|vru|}|vru|\epsilon$.

Dokaz izvedemo z indukcijo po dolžini besede vru . Za $|vru| = |r|$ je po predpostavki $\|\hat{\rho}(r) - I\| < \epsilon < |r|\epsilon$.

Naj bo $|v|+|u| > 0$. Ker je $|r| \geq 1$ in $vu \equiv 1 \pmod{G}$, po indukcijski predpostavki velja $\hat{\rho}(vu) = I + E_{vu}$ za $\|E_{vu}\| < (\sqrt{n} + \epsilon)^{|vu|}|vu|\epsilon$.

Velja

$$\begin{aligned} \hat{\rho}(v)\hat{\rho}(r)\hat{\rho}(u) &= \hat{\rho}(v)(I + E_r)\hat{\rho}(u) \\ &= \hat{\rho}(v)\hat{\rho}(u) + \hat{\rho}(v)E_r\hat{\rho}(u) \\ &= I + E_{vu} + \hat{\rho}(v)E_r\hat{\rho}(u) \\ &= I + E_{vru} \end{aligned}$$

za $\|E_{vru}\| = \leq (\sqrt{n} + \epsilon)^{|vru|}|vru|\epsilon$. Splošna beseda $w \equiv 1 \pmod{G}$ je oblike $w = v_1 r_1 u_1 v_2 r_2 u_2 \cdots v_L r_L u_L$. Računamo

$$\begin{aligned} \hat{\rho}(w) &= \hat{\rho}(v_1 r_1 u_1)\hat{\rho}(v_2 r_2 u_2) \cdots \hat{\rho}(v_L r_L u_L) \\ &= (I + E_{v_1 r_1 u_1})(I + E_{v_2 r_2 u_2}) \cdots (I + E_{v_L r_L u_L}) \\ &= I + \sum_{\emptyset \neq T \subseteq [L]} \prod_{t \in T} E_{v_t r_t u_t} \end{aligned}$$

in ocenimo

$$\begin{aligned} \left\| \sum_{\emptyset \neq T \subseteq [L]} \prod_{t \in T} E_{v_t r_t u_t} \right\|_F &\leq \sum_{\emptyset \neq T \subseteq [L]} \prod_{t \in T} (\sqrt{n} + \epsilon)^{|v_t u_t|} |v_t r_t u_t| \epsilon \\ &\leq \sum_{\emptyset \neq T \subseteq [L]} \prod_{t \in T} (\sqrt{n} + \epsilon)^{|w|} |w| \epsilon \\ &\leq \left(1 + (\sqrt{n} + \epsilon)^{|w|} |w| \epsilon\right)^L - 1. \end{aligned}$$

□

Lema 2.22. *Naj bo $\mathcal{L}_{\text{rel}}(\hat{\rho}) < |R|\epsilon$ in $\mathcal{L}_{\text{unitary}}(\hat{\rho}) < |S|\epsilon$. Potem za taki besedi $w, v \in S^*$, da velja $w \equiv v^{-1} \pmod{G}$ velja*

$$\|\hat{\rho}(w) - \hat{\rho}(v)^{-1}\| \leq \frac{(1 + (\sqrt{n} + \epsilon)^{|w|}|w|\epsilon)^L - 1}{\|\hat{\rho}(v)^{-1}\|} =: o_{w,v}(\epsilon).$$

Dokaz. Opazimo, da za $A, B \in GL_n$ iz $\|AB - I\| < \delta$ sledi $\|B - A^{-1}\| < \frac{\delta}{\|A^{-1}\|}$. Za dovolj majhen ϵ so matrike $\hat{\rho}(v)$ obrnljive. Dokaz sledi iz prejšnje leme 2.21 in zgornje opazke. \square

Lema 2.23. *Naj bo $\hat{\rho}$ model, da velja $\mathcal{L}_{\text{rel}}(\hat{\rho}) < |R|\epsilon$ in $\mathcal{L}_{\text{unitary}}(\hat{\rho}) < |S|\epsilon$. Potem za vsak $g, h \in G$ velja*

$$\|\hat{\rho}(gh) - \hat{\rho}(g)\hat{\rho}(h)\| < 2 \frac{(1 + (\sqrt{n} + \epsilon)^{l_{g,h}} l_{g,h} \epsilon)^L - 1}{\|\hat{\rho}(w_{(gh)^{-1}})^{-1}\|},$$

kjer je $l_{g,h} = \max(|w_{g,h}|, |w_g w_h|)$.

Dokaz. Za kanonične predstavnike $w_g, w_h, w_{gh} \in S^*$ velja $w_g w_h \equiv w_{gh}^{-1} \pmod{G}$.

Ocenimo

$$\begin{aligned} \|\hat{\rho}(w_g w_h) - \hat{\rho}(w_{gh})\| &\leq \|\hat{\rho}(w_g w_h) - \hat{\rho}(w_{(gh)^{-1}})\| \\ &\quad + \|\hat{\rho}(w_{(gh)^{-1}}) - \hat{\rho}(w_{gh})\| \\ &\leq o_{w_g, w_{(gh)^{-1}}}(\epsilon) + o_{w_{gh}, w_{(gh)^{-1}}}(\epsilon) \\ &= 2 \frac{(1 + (\sqrt{n} + \epsilon)^{l_{g,h}} l_{g,h} \epsilon)^L - 1}{\|\hat{\rho}(w_{(gh)^{-1}})^{-1}\|}, \end{aligned}$$

\square

S tem smo dokazali, da je model $\hat{\rho}$ z dovolj majhno napako skoraj homomorfizem. Dokazali bomo še, da obstaja preslikava

$$\mu: G \rightarrow U(n)$$

v unitarne matrike, ki je blizu modelu $\hat{\rho}$. Nato bo iz izreka 2.19 sledilo, da obstaja prava upodobitev ρ blizu modelu $\hat{\rho}$.

Za model $\hat{\rho}$ z majhno napako definiramo preslikavo $\mu: S \rightarrow U(n)$ s predpisom

$$\mu: s \mapsto Q_s.$$

Preslikavo implicitno razširimo na S^* in na G enako kot $\hat{\rho}$.

Lema 2.24. *Za preslikavo $\mu: G \rightarrow U(n)$, inducirana iz modela $\hat{\rho}$ z $\mathcal{L}_{\text{unitary}}(\hat{\rho}) < |S|\epsilon$ in $\mathcal{L}_{\text{rel}}(\hat{\rho}) < |R|\epsilon$, velja*

$$\|\mu(gh) - \mu(g)\mu(h)\| < k_{g,h}(\epsilon)$$

za

$$k_{g,h}(\epsilon) = 2|w_{gh}|(\sqrt{n} - \epsilon)^{|w_{gh}|-1} \epsilon + 2 \frac{(1 + (\sqrt{n} + \epsilon)^{l_{g,h}} l_{g,h} \epsilon)^L - 1}{\|\hat{\rho}(w_{(gh)^{-1}})^{-1}\|} =$$

Dokaz. Iz leme 2.20 sledi, da je za $g \in G$ $\|\mu(g) - \hat{\rho}(g)\| < |w_g|(\sqrt{n} - \epsilon)^{|w_g|-1}\epsilon$. Ocenimo

$$\begin{aligned} \|\mu(gh) - \mu(g)\mu(h)\| &\leq \|\mu(gh) - \hat{\rho}(gh)\| + \|\hat{\rho}(gh) - \hat{\rho}(g)\hat{\rho}(h)\| \\ &\quad + \|\hat{\rho}(g)\hat{\rho}(h) - \mu(g)\mu(h)\| \\ &\leq 2|w_{gh}|(\sqrt{n} - \epsilon)^{|w_{gh}|-1}\epsilon + 2 \frac{(\sqrt{n} + \epsilon)^{l_{g,h}} l_{g,h} \epsilon^L - 1}{\|\hat{\rho}(w_{(gh)^{-1}})^{-1}\|} \\ &:= k_{g,h}(\epsilon) \end{aligned}$$

□

Iz definicije preslikave μ za modele z majhno napako očitno sledi $\|\mu(g) - \rho(g)\| \leq o(\epsilon)$. Iskana upodobitev ρ sledi iz izreka 2.19. Treba je preveriti še, da je upodobitev nerazcepna. Povzemo posledice konstrukcije μ in izreka 2.19 v spodnji lemi.

Lema 2.25. *Naj bo $\hat{\rho}$ model z $\mathcal{L}_{\text{rel}}(\hat{\rho}) < |R|\epsilon$ in $\mathcal{L}_{\text{unitary}}(\hat{\rho}) < |S|\epsilon$. Za inducirano preslikavo μ in upodobitev ρ , definirano kot zgoraj, velja*

$$\|\hat{\rho}(s) - \rho(s)\| < o(\epsilon)$$

in

$$\|\hat{\rho}(s) - \mu(s)\| < o(\epsilon).$$

Dokaz. Opazujmo normo karakterja preslikave ρ . Iz pogojev zgornje leme sledi

$$|\chi_\rho| = |\chi_{\hat{\rho}}| + o(\epsilon).$$

Za dovolj majhno ϵ je $|\chi_{\hat{\rho}}| < 2$, iz česar po izreku 2.7 sledi, da je upodobitev ρ nerazcepna. S tem je izrek 2.18 dokazan. □

2.4 Ciklične grupe

Oglejmo si naš pristop iskanja nerazcepnih upodobitev na cikličnih grupah $C_n = \langle z | z^n = 1 \rangle$. Edine nerazcepne upodobitve ciklične grupe so

$$\rho_k(z) = e^{\frac{2\pi k}{n}},$$

kjer je $k \in \{0, 1, \dots, n\}$. Pri iskanju teh upodobitev z gradientnim spustom se bomo omejili na enodimensionalne modele.

Model $\hat{\rho}: G \rightarrow \mathbb{C}^*$ je določen z vrednostjo $\hat{\rho}(z) = x + iy$ in je odvisen od dveh realnih parametrov x, y .

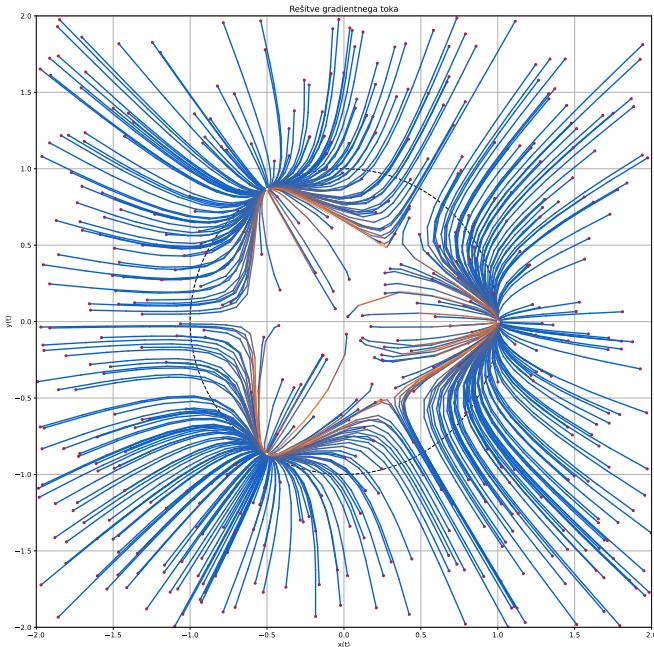
Funkcije izgube so

$$\mathcal{L}_{\text{rel}} = |(x + iy)^n - 1|^2 \tag{2.8}$$

$$\mathcal{L}_{\text{irr}} = \left(\frac{1}{n} \sum_{j=0}^{n-1} |(x + iy)^j|^2 - 1 \right)^2 \tag{2.9}$$

$$\mathcal{L}_{\text{unitary}} = |x^2 + y^2 - 1|^2. \tag{2.10}$$

Na grafu 3 so trajektorije rešitev diferencialne enačbe $\frac{d(x,y)}{dt} = -\nabla \mathcal{L}$ za različne začetne parametre, izračunane s pomočjo gradientnega spusta. Opazimo, da vse rešitve konvergirajo do nerazcepnih upodobitev. Ta rezultat teoretično podpremo v izreku 2.26.



Slika 3: Rešitve $\frac{d(x,y)}{dt} = -\nabla \mathcal{L}$ pri različnih začetnih parametrih (x_0, y_0) za $n = 3$.

2.5 Redukcija funkcij izgube

Nerazcepne upodobitve ρ_k so edine enodimensionalne upodobitve cikličnih grup. Za poljubno upodobitev $\rho: C_n \rightarrow \mathbb{C}^*$ mora namreč veljati $1 = \rho(z^n) = \rho(z)^n$, iz česar sledi, da je $\rho = \rho_k$ za nek k . Zato lahko pri iskanju nerazcepnih upodobitev z gradientnim spustom izpustimo \mathcal{L}_{irr} .

Ker so vse upodobitve ρ_k unitarne, lahko iz funkcije izgube izpustimo tudi $\mathcal{L}_{\text{unitary}}$ in za funkcijo izgube \mathcal{L} vzamemo le $\mathcal{L} = \mathcal{L}_{\text{rel}}$.

V splošnem lahko pri optimizaciji funkcije izgube pričakujemo lokalne minimum, v katerih je vrednost \mathcal{L} neničelna. Edine privlačne točke enačbe $\frac{d(x,y)}{dt} = -\nabla \| (x + iy)^n - 1 \|^2$ pa so korenji enote $z^n = 1$, torej natanko unitarne nerazcepne upodobitve.

Izrek 2.26. *Edine kritične točke preslikave $(x, y) \mapsto \| (x + iy)^n - 1 \|^2$ so $e^{\frac{2\pi i k}{n}}$ za $k \in \{0, 1, \dots, n-1\}$ in $(0, 0)$.*

Dokaz. Lokalni ekstremi funkcije \mathcal{L}_{rel} so natanko v ničlah njenega gradijeta $\nabla \mathcal{L}_{\text{rel}} = (\frac{d\mathcal{L}_{\text{rel}}}{dx}, \frac{d\mathcal{L}_{\text{rel}}}{dy})$.

Uvedemo $z = x + iy$ ⁷. Velja $\mathcal{L} = \mathcal{L}_{\text{rel}} = (z^n - 1)(\bar{z}^n - 1)$,

$$\frac{d\mathcal{L}_{\text{rel}}}{dx} = nz^{n-1}(\bar{z}^n - 1) + (z^n - 1)n\bar{z}^{n-1} = 2 \operatorname{Re}(nz^{n-1}(\bar{z}^n - 1))$$

⁷Imaginarno število i v tem primeru predstavlja index. Zanima nas gradien $\nabla \mathcal{L} \in \mathbb{R}^2$. Za lažje računanje \mathbb{R}^2 identificiramo z \mathbb{C} . Za $u \in \mathbb{R}$ in iu predstavlja točko $(0, u) \in \mathbb{R}^2$. Zato pri izpeljavi $\frac{d\mathcal{L}}{dy}$ izpustimo i .

in

$$\frac{d\mathcal{L}_{\text{rel}}}{dy} = 2 \operatorname{Im}(nz^{n-1}(\bar{z}^n - 1)).$$

Če velja $\nabla\mathcal{L} = 0$, je $z^{n-1}(\bar{z}^n - 1) = 0$. Take vrednosti z so natanko $x = y = 0$ in $x + iy = e^{\frac{2\pi ik}{n}}$ za $k \in \mathbb{N}$. \square

2.6 Diedrske grupe

Opazujmo diedrsko grupo $D_n = \langle r, s | r^n = s^2 = (rs)^2 = 1 \rangle$. Enorazsežne upodobitve diedrske grupe so

$$\begin{aligned}\chi_{\varepsilon, \delta}: \quad s &\mapsto \varepsilon, \\ r &\mapsto \delta,\end{aligned}$$

kjer je $\varepsilon \in \{-1, 1\}$, in $\delta \in \{1, -1^{n+1}\}$. Dvorazsežne nerazcepne upodobitve so

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

kjer je $0 \leq k < \frac{n}{2}$. Vse nerazcepne upodobitve so realne, zato jih bomo iskali z realnimi modeli.

2.6.1 Enodimensionalne upodobitve

Model enorazsežne upodobitve diedrske grupe je določen s slikama generatorjev⁸ $\hat{\rho}(s) = s \in \mathbb{R}, \hat{\rho}(r) = r \in \mathbb{R}$. Funkcije izgube so enake

$$\begin{aligned}\mathcal{L}_{\text{rel}} &= \frac{1}{3} (|s^2 - 1|^2 + |r^n - 1|^2 + |(rs)^2 - 1|^2) \\ \mathcal{L}_{\text{unit}} &= \frac{1}{2} (|r^2 - 1|^2 + |s^2 - 1|^2) \\ \mathcal{L}_{\text{irr}} &= \left(\frac{1}{2n} \sum_{i=0}^n (|r^i|^2 + |r^i s|^2) - 1 \right)^2.\end{aligned}$$

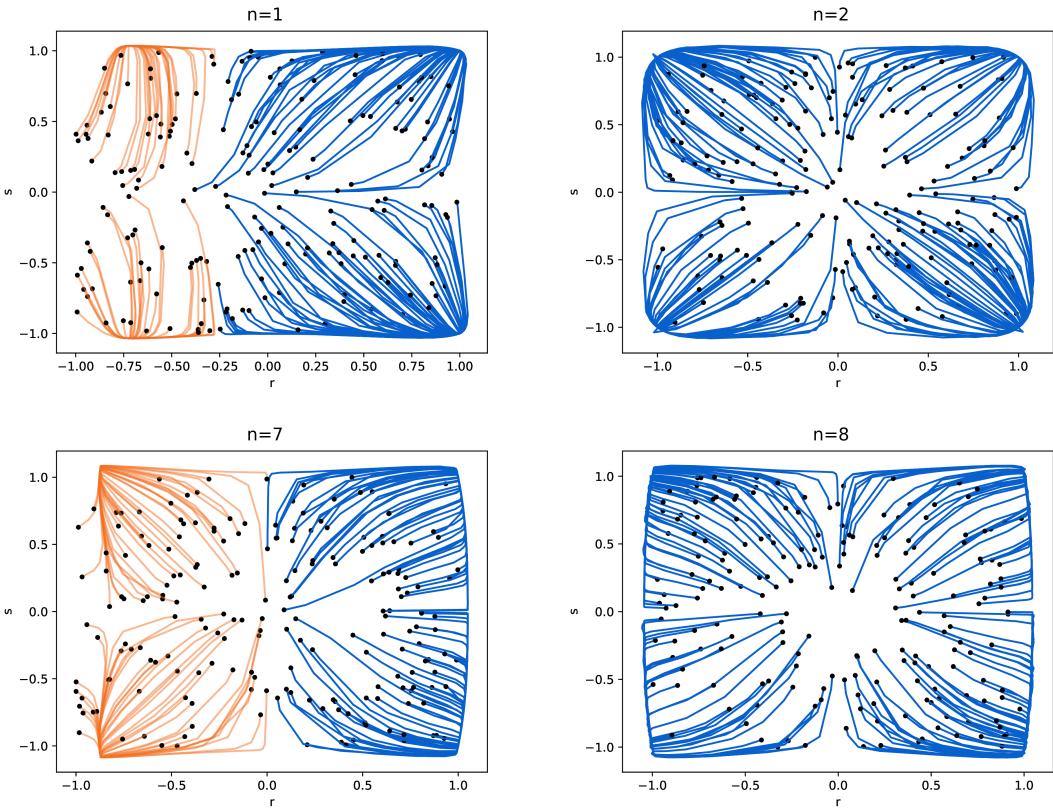
Numerične rešitve Numerično rešujemo enačbo gradientnega toka

$$\frac{d(r, s)}{dt} = -\nabla(\mathcal{L}_{\text{rel}} + \mathcal{L}_{\text{unit}} + \mathcal{L}_{\text{irr}}).$$

Eračbo rešujemo z metodo Dormand-Prince [9], implementirano v knjižnici SciPy [27].

Na sliki 13 so različne rešitve enačb gradientnega toka za $n \in \{1, 2, 7, 8\}$. Modre v limiti dosežejo ničelno funkcijo izgube. Za sod n vse narisane trajektorije konvergirajo do nerazcepne upodobitve, za lih n pa gradientni spust iz $[-1, 0] \times [-1, 1]$ nikoli ne doseže ničelne napake. Enake lastnosti opazimo na širši množici diedrskih grup - glej dodatek B.1.1. Opazimo, da vse numerične rešitve z začetnimi parametri (r_0, s_0) konvergirajo k $\operatorname{sign}(r_0), \operatorname{sign}(s_0)$.

⁸Izrabljamo notacijo in z r označimo abstraktni element grupe D_n in realno število $r \in \mathbb{R}$, ki predstavlja sliko generatorja r .



Slika 4: Rešitve enačb gradientnega toka za različne n in različne začetne parameterje⁹. Črne pike predstavljajo začetne vrednosti rešitev. Modre rešitve kovergirajo k nerazcepni upodobitvi, oranžne pa v lokalni minimum \mathcal{L} , ki ni globalni. Več grafov je v dodatku B.1.1.

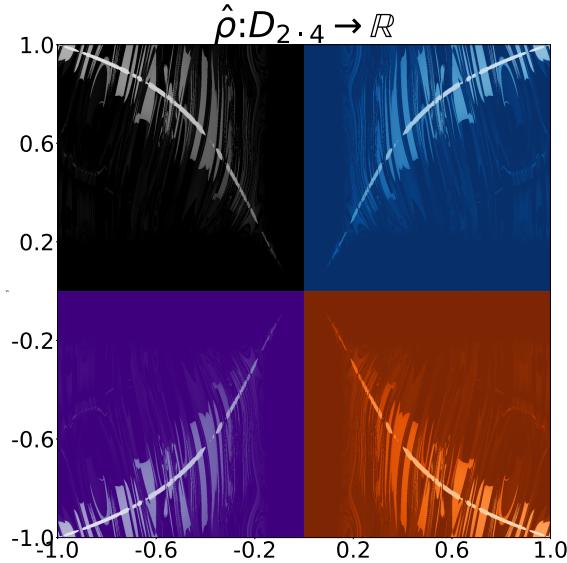
Dinamika Za vsak par začetnih parametrov $(r_0, s_0) \in \mathbb{R}^2$ enodimensionalnega modela lahko študiramo hitrost konvergencije k lokalnim ekstremom. Slika 5 prikazuje čas t_{r_0, s_0} , ob katerem je rešitev $(r(t), s(t))$ z začetnimi parametri (r_0, s_0) prvič v $\epsilon = 0.01$ pasu njene limite v neskončnosti. Različni grafi dinamike so v dodatku B.1.1.

2.6.2 Dvodimensionalne upodobitve

Model dvorazsežne upodobitve D_n je določen z matrikama $\hat{\rho}(s) = S \in \mathbb{R}^{2 \times 2}$ in $\hat{\rho}(r) = R \in \mathbb{R}^{n \times n}$. Funkcije izgube so enake

$$\begin{aligned}\mathcal{L}_{\text{rel}} &= \frac{1}{3} \left(\|S^2 - 1\| + \|R^n - 1\| + \|(RS)^2 - 1\| \right) \\ \mathcal{L}_{\text{unit}} &= \frac{1}{2} \left(\|R^2 - 1\| + \|S^2 - 1\| \right) \\ \mathcal{L}_{\text{irr}} &= \left(\frac{1}{2n} \sum_{i=0}^n (\|R^i\| + \|R^i S\|) - 1 \right)^2.\end{aligned}$$

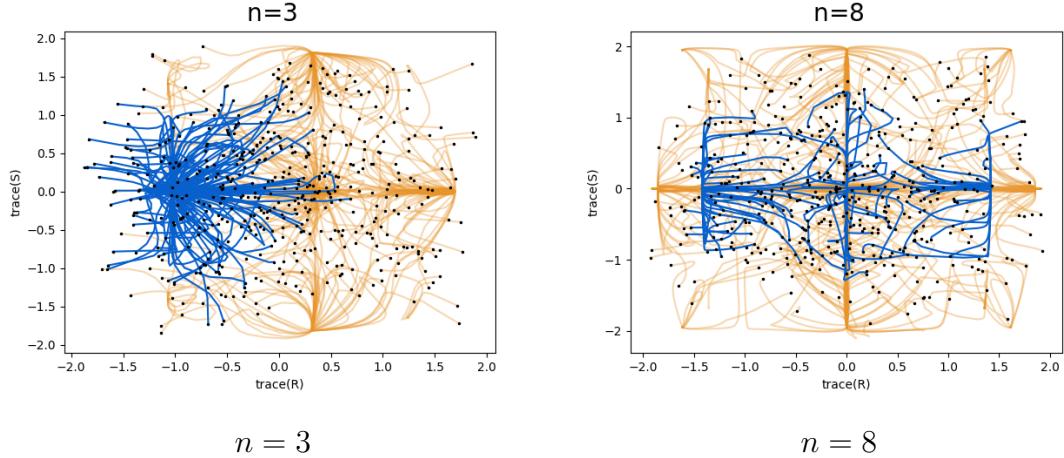
Hitrost konvergenc različnih začetnih parametrov



Slika 5: Hitrosti konvergenc različnih začetnih parametrov. Barva točke (r_0, s_0) določa limito - modre točke konvergirajo k $(1, 1)$, vijolične k $(-1, -1)$, oranžne k $(-1, 1)$ in sive k $(1, -1)$. Odtotenek točke določa hitrost konvergenc rešitve $(r(t), s(t))$. Svetlejša, kot je barva, manjši je prvi t_1 , v katerem se $(r(t_1), s(t_1))$ razlikuje od limite za manj kot $\epsilon = 0.01$.

Numerične rešitve Slika 6 prikazuje numerične rešitve za različne začetne parametre. Trajektorije rešitev $(R(t), S(t))$ vizualiziramo v \mathbb{R}^2 kot $(\text{tr}(R(t)), \text{tr}(S(t)))$. Nekateri začetni parametri vodijo do nerazcepne upodobitve, drugi ne. Računamo lahko verjetnost izbire začetnih parametrov, ki konvergirajo do ničel funkcije izgube. Verjetnosti v tabeli 1 so izračunane iz iz vzorcov v dodatku B.1.1.

¹⁰Začetni parametri R_0, S_0 so oblike $\begin{bmatrix} x & y \\ z & \text{tr} - x \end{bmatrix}$, kjer so $x, y, z, \text{tr} \sim U[-2, 2]$



Slika 6: Rešitve za različne n . Modre krivulje konvergirajo do nerazcepnih upodobitev. Črne pike predstavljajo začetne vrednosti parametrov¹⁰.

Group	$P \left(\mathcal{L} \left(\lim_{t \rightarrow \infty} \phi(t) \right) = 0 \right)$
D_3	0.274
D_4	0.158
D_5	0.306
D_6	0.210
D_7	0.311
D_8	0.235
D_9	0.330
D_{10}	0.241
D_{11}	0.327
D_{12}	0.248
D_{13}	0.319

Tabela 1: Verjetnost izbire začetnih parametrov, ki konvergirajo. Začetni parametri so vzorčeni enako kot v sliki 6. Velikost vzorca je 10^4 .

2.7 Splošna dinamika

Za model ρ s parametri iz \mathbb{R}^N in funkcijo izgube $\mathcal{L}: \mathbb{R}^N \rightarrow \mathbb{R}$ rešujemo enačbo gradientnega toka

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

kjer so ϕ_0 poljubni začetni parametri.

Za vsako izbiro začetnih parametrov ϕ_0 obstaja limita $\lim_{t \rightarrow \infty} \phi(t) = \phi_\infty$, h kateri konvergira rešitev enačbe gradientnega toka. Zanimajo nas meje med območji začetnih parametrov z istimi limitami. Limite so natanko kritične točke funkcije izgube \mathcal{L} [10, Trditev na strani 203]. Oglejmo si posplošitev modela diedrske grupe. Opazujemo model $\rho: G \rightarrow \mathbb{R}$ grupe $G = \langle s_G, z_G \mid R \rangle$ z dvema generatorjema.

Brez škode za splošnost predpostavimo, da ima vsaka relacija v R najmanj dve pojavivti generatorja s_G , ali pa je oblike $z_G^{\text{red}(z_G)}$. Simetrično za z_G . Označimo $R = \{r_1, r_2, \dots, r_l\}$ in za model s parametrom $\rho(s_G) = s \in \mathbb{R}$ in $\rho(z_G) = z \in \mathbb{R}$ označimo

$$\rho(r_i) = s^{a_i} z^{b_i} \in \mathbb{R}.$$

Opazujmo enačbo gradientnega toka

$$\frac{d(s, z)}{dt} = -\nabla \mathcal{L}_{\text{rel}}(s, z).$$

Glejmo točke, v katerih sta parcialna odvoda ničelna.

$$\frac{d}{ds} \left(\frac{1}{|R|} \sum_{i=1}^l (s^{a_i} z^{b_i} - 1)^2 \right) = \frac{2}{|R|} \sum_{i=1}^l a_i s^{a_i-1} z^{b_i} (s^{a_i} z^{b_i} - 1).$$

Očitno velja $\frac{d}{ds} \mathcal{L}_{\text{rel}}(0, z) = \frac{d}{dz} \mathcal{L}_{\text{rel}}(s, 0) = 0$. Rešitev z začetnima parametroma $s_0 = 0, z_0 \in \mathbb{R}$ je

$$s(t) = 0, \quad z(t) = z_0 + \frac{2}{|R|} z^{\text{red}(z)-1},$$

simetrično za $z_0 = 0$, kar ustreza rezultatu za diedrske grupe.

Preučimo še dogajanje znotraj kvadratnov. Recimo, da je $s > 0, z > 0$.

TODO: Uvedemo lahko $s = e^x, z = e^y, F(x, y) = \sum_i (e^{a_i x + b_i y} - 1)^2$. Kritične točke so iste kot v L . Ali obstajajo kritične točke izven $\{\pm 1\}^2$?

3 Iskanje delovanj

Na podoben način lahko iščemo delovanja grup. Naj bo G končna grupa in $[n] = \{1, 2, \dots, n\}$ končna množica. Iščemo homomorfizem $\rho: G \rightarrow S_n$.

Vsako preslikavo $f \in \text{fun}([n], [n])$ se da predstaviti z vektorjem $\begin{bmatrix} f(1) \\ \vdots \\ f(n) \end{bmatrix} \in \mathbb{R}^n$.

Lahko bi začeli s poljubno preslikavo $\hat{\rho}: S \rightarrow \mathbb{R}^n$ in sistematično spremojali elemente vektorjem $\hat{\rho}(s)$, dokler ne pridemo do homomorfizma. To se da, a za optimizacijo parametrov ne moremo uporabiti gradientnega spusta, saj je prostor vektorjev $\{[f(i)]_{i=1,\dots,n} | f \in \text{fun}([n], [n])\}$ diskreten.

Nad prostorom funkcij $\text{fun}([n], [n])$ definiramo gladko družino porazdelitev. Generatorje grupe sprva slikamo v poljubne porazdelitve, nato pa spremojamo porazdelitve, da je verjetnost za homomorfizem čim večja.¹¹

Model $\hat{\rho}$ delovanja $\rho: G \rightarrow S_n$ bo preslikava $\hat{\rho}: G \rightarrow \mathcal{P}(\text{fun}([n], [n]))$, definirana na generatorjih $s \in S$ s predpisom

$$s \mapsto \begin{bmatrix} P(s(1) = 1) & P(s(1) = 2) & \cdots & P(s(1) = n) \\ P(s(2) = 1) & P(s(2) = 2) & \cdots & P(s(2) = n) \\ \vdots & \vdots & & \vdots \\ P(s(n) = 1) & P(s(n) = 2) & \cdots & P(s(n) = n) \end{bmatrix}. \quad (3.1)$$

Matrika $P_s = \hat{\rho}(s)$ definira slučajno preslikavo¹² $s \in \text{fun}([n], [n])$ s porazdelitvijo

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

za $f \in \text{fun}([n], [n])$.

Definicija 3.1. Preslikavo $\hat{\rho}$ lahko podobno kot v definiciji 2.11 raširimo na celotno grupo:

- Za vsak $g \in G$ izberemo takega predstavnika $v_g \in F_S$, ki je sestavljen le iz generatorjev $s \in S$, ne pa iz njihovih inverzov.
- Za g s predstavnikom $v_g = s_1 s_2 \cdots s_m$ definiramo $\hat{\rho}(g) := P_{s_1} P_{s_2} \cdots P_{s_m}$.

V delu bomo za $g \in G$ s predstavnikom $v_g = s_1 s_2 \cdots s_m$ pisali kar $g = s_1 s_2 \cdots s_m$. Lahko je preveriti, da za $g = s_1 s_2 \cdots s_m \in G$ velja

$$P_{s_1} P_{s_2} \cdots P_{s_m} = \begin{bmatrix} P(g(1) = 1) & P(g(1) = 2) & \cdots & P(g(1) = n) \\ P(g(2) = 1) & P(g(2) = 2) & \cdots & P(g(2) = n) \\ \vdots & \vdots & & \vdots \\ P(g(n) = 1) & P(g(n) = 2) & \cdots & P(g(n) = n) \end{bmatrix}.$$

¹¹V strojnem učenju nad diskreten prostor izidov D običajno definiramo gladko družino porazdelitev $\{\mathcal{P}_\phi(D)\}$ in minimiziramo $-\log(P_\phi(\text{opaženi izidi}))$.

¹²Izraba notacije. S s označimo tako generator kot slučajno funkcijo nad $[n]$, ki je porazdeljena s porazdelitvijo $P(s = f) = \prod_{i=1}^n P(s(i) = f(i))$. Prav tako za poljuben element $g = s_1 s_2 \cdots s_m \in G$ z g označimo tudi slučajno spremenljivko $s_1 s_2 \cdots s_n$.

Naš model elemente grupe slika v stohastične matrike, ki definirajo verjetnostne porazdelitve slučajnih funkcij nad n . Parametre modela bomo optimizirali tako, da bo homomorfizem najbolj verjeten.

3.1 Parametrizacija stohastičnih matrik

Definicija modela $\hat{\rho}$ zahteva, da so matrike P_s stohastične¹³. Iteracije gradientnega spusta stohastičnosti ne ohranjajo. Če hočemo za optimizacijo modela $\hat{\rho}$ uporabiti gradientni spust, moramo metodo prilagoditi tako, da novi parametri ostanejo v družini stohastičnih matrik.

Družino stohastičnih matrik lahko parametriziramo s poljubno parametrizacijo $p: \mathbb{R}^{n \times n} \rightarrow \{P \in \mathbb{R}^{n \times n} \mid P \text{ je stohastična}\}$, stohastične matrike P_s pa definiramo kot $p(\phi_s)$, kjer je $\phi = \{\phi_s \mid s \in S\} \subset \mathbb{R}^{n \times n}$ množica poljubnih matrik. Za parametre modela vzamemo elemente matrik ϕ_s . Na ta način bodo matrike $P_s = P_s(\phi_s)$ vedno stohastične, ne glede na spremembo parametrov $-\eta \nabla_\phi \mathcal{L}(P)$.

Za parametrizacijo p lahko vzamemo preslikavo softmax, ki vsako vrstico spremeni v stohastični vektor.

Definicija 3.2. Preslikava softmax: $\mathbb{R}^n \rightarrow \mathbb{R}^n$ slika vektor $v = [v_1, v_2, \dots, v_n]^T$ v vektor z elementi $\text{softmax}(v)_i = \frac{e^{v_i}}{\sum_{j=1}^n e^{v_j}}$.

Za matriko $A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix}$ definiramo softmax po vrsticah kot

$$\text{softmax}(A) = \begin{bmatrix} \text{softmax}(a_1)^T \\ \text{softmax}(a_2)^T \\ \vdots \\ \text{softmax}(a_n)^T \end{bmatrix}.$$

Preslikava softmax očitno poljubne matrike preslika v stohastične. Z njo lahko za generator $s \in S$ matriko P_s definiramo kot

$$P_s = \text{softmax}(\phi_s).$$

Model za delovanja $\hat{\rho} = \hat{\rho}_\phi$ definiramo v odvisnosti od parametrov $\phi = \{\phi_s \in \mathbb{R}^{n \times n} \mid s \in S\}$ kot

$$\phi \mapsto (s \mapsto \text{softmax}(\phi_s))$$

in spremenjamo elemente matrik ϕ_s , da model $\hat{\rho}_\phi$ skoraj zagotovo predstavlja delovanje.

3.2 Funkcija izgube nad relacijami

Naj bo $\langle S | R \rangle$ prezentacija grupe G . Za vsako relacijo $r \in R$ bi radi $P(r = \text{id}) = 1$ oziroma $0 = \log(P(r = \text{id})) = \log(\prod_{i=1}^n P(r(i) = i)) = \text{tr}(\log(P_r))$, kjer $\log(P_r)$ računamo po elementih.

¹³ P je stohastična, če je vsota elementov v vsaki vrstici enaka 1, vsi elementi matrike pa so med 0 in 1.

Definicija 3.3. Funkcijo izgube nad relacijami za delovanja definiramo kot

$$\mathcal{L}_{\text{rel}}(\hat{\rho}) = -\frac{1}{|R|} \sum_{r \in R} \text{tr}(\log(P_r)). \quad (3.2)$$

3.3 Pretvorba modela v delovanje

Rezultati na sliki 7 namigujejo, da ob konvergenci $\mathcal{L}_{\text{rel}} \rightarrow 0$ matrika P_z konvergira do permutacijske matrike \overline{P}_z . To inducira delovanje $z \mapsto P_z$.

Za splošno grupo $G = \langle S | R \rangle$ lahko definiramo sledeč postopek pretvorbe modela $\hat{\rho}: s \rightarrow P_s$ v preslikavo $\rho: G \rightarrow \Pi$ med grupo G in grupo permutacijskih matrik Π^{14} .

- Za matriko P definiramo \overline{P} kot matriko, ki ima v vsaki vrstici 1 na mestu največjega elementa v vrstici matrike P in 0 drugje.
- Definiramo $\rho: G \rightarrow \mathbb{R}^{n \times n}$ s predpisom $s \rightarrow \overline{P}_s$.

V tem razdelku dokažemo, da je za model $\hat{\rho}$ z dovolj majhno funkcijo izgube nad relacijami $\mathcal{L}_{\text{rel}}(\hat{\rho})$ inducirana preslikava ρ homomorfizem med G in Π . Naj bo $r \in R$ in $s \in S$ poljuben generator. Najprej pokažemo, da je za dovolj majhno vrednost izgube nad relacijami matrika P_r blizu identitete, matrika P_s pa obrnljiva.

Lema 3.4. Iz $\mathcal{L}_{\text{rel}}(\hat{\rho}) \leq \frac{1}{|R|} - \log(1 - \frac{\epsilon}{n})$ sledi $P_r = I + E_r$ za $\|E_r\| < \epsilon$.

Dokaz. Najprej računamo

$$\begin{aligned} -\frac{1}{|R|} \sum_{r \in R} \text{tr}(\log(P_r)) &\leq \delta \\ -\log(\prod_{i=1}^n (P_r)_{i,i}) &\leq \delta|R| \\ \log(P(r = \text{id})) &\geq -\delta|R| \\ 1 - P(r = \text{id}) &\leq 1 - e^{-\delta|R|} =: \gamma. \end{aligned}$$

Ocenimo

$$1 - (P_r)_{i,i} \leq 1 - (P_r)_{1,1}(P_r)_{2,2} \cdots (P_r)_{n,n} = 1 - P(r = \text{id}) \leq \gamma.$$

P_r je stohastična, torej velja $\sum_{j=1}^n (P_r)_{i,j} = 1$. Iz tega sledi $\sum_{j \neq i} (P_r)_{i,j} \leq \gamma$. Sledi

$$P_{i,j} \leq \sum_{j \neq i} P_{i,j} \leq \gamma,$$

za $i \neq j$. Iz tega sledi $\|E_r\| \leq n\gamma = n(1 - e^{-\delta|R|}) = \epsilon$. □

Lema 3.5 (Za dovolj majhen ϵ je P_s obrnljiva). Če je $\mathcal{L}_{\text{rel}}(\hat{\rho}) \leq \frac{-1}{|R|} \log(1 - \frac{1}{2n})$, je matrika P_s obrnjiva za vsak generator $s \in S$.

¹⁴To je ekvivalentno preslikavi $G \rightarrow S_n$, saj je $S_n \cong \Pi$.

Dokaz. Če je $\|E\| < 1$, je matrika $I + E$ obrnljiva. Velja namreč $\lambda_i(E) \leq \|E\|_2 \leq \|E\|$. E izrazimo z Jordanovo formo $E = QJQ^{-1}$ in računamo

$$\det(I + E) = \det(I + J) = \prod_{i=1}^k (1 + \lambda_i(E))^{d_i} > 0,$$

kjer so $\lambda_i(E)$ lastne vrednosti E in d_i njihove algebraične večkratnosti.

Če je $\mathcal{L}_{\text{rel}}(\hat{\rho}) \leq \frac{-1}{|R|} \log(1 - \frac{1}{2n})$, po prejšnji lemi velja $P_r = I + E_r$ za $\|E_r\| \leq \frac{1}{2}$. Torej je matrika $P_r = I + E_r = P_{s_1}P_{s_2} \cdots P_{s_n}$ obrnljiva, posledično pa so obrnljive vse matrike P_{s_i} . \square

Zdaj bomo dokazali, da je za dovolj majhen \mathcal{L}_{rel} in poljuben generator $s \in S$ matrika P_s blizu permutacijske matrike. Za to potrebujemo spodnji izrek 3.6 in lemo 3.7.

Izrek 3.6. *Obrnljiva stohastična matrika je permutacijska natanko tedaj, ko je njen inverz stohastičen.*

Elementaren dokaz izreka je v [7].

Lema 3.7. *Naj bo P obrnljiva vrstično stohastična matrika in R vrstično stohastična matrika, za katero velja*

$$\|P^{-1} - R\|_F \leq \frac{\epsilon^2}{2n^{3/2}}.$$

Potem velja $d(P, \Pi) \leq \epsilon$. Če je dodatno $\epsilon < \frac{1}{2}$, je $\bar{P} \in \Pi$ in velja $\|P - \bar{P}\|_F \leq \epsilon$.

Dokaz. Naj bo $X = [x_{i,j}]$ poljubna stohastična matrika, za katero velja $d(X, \Pi) > \epsilon$. Najprej bomo pokazali, da je tedaj $\|XY - I\|_F > \frac{\epsilon^2}{2n}$ za vsako stohastično matriko Y .

Ker je $d(X, \Pi) > \epsilon$, za vsako permutacijsko matriko $Q = [q_{i,j}]$ velja

$$\epsilon^2 < \|X - Q\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n (x_{i,j} - q_{i,j})^2.$$

Zato obstaja vrstica x_i , za katero je

$$\sum_{j=1}^n (x_{i,j} - q_{i,j})^2 > \frac{\epsilon^2}{n}.$$

Ker je v Q poljubna in ima v vsaki vrstici natanko en element enak 1, ostali pa 0, za vsak standardni bazni vektor e_k velja $\|x_i - e_k\|_2^2 \geq \frac{\epsilon^2}{n}$. Ocenimo

$$\frac{\epsilon^2}{n} < \|x_i - e_k\|_2^2 = (1 - x_{i,k})^2 + \sum_{j \neq k} x_{i,j}^2 \leq 2(1 - x_{i,k}).$$

Za vse k sledi $x_{i,k} \leq 1 - \frac{\epsilon^2}{2n}$. Naj bo Y poljubna stohastična matrika. Oglejmo si i -to vrstico produkta XY .

$$(XY)_{i,k} = \sum_{h=1}^n x_{i,h} y_{h,k} \leq 1 - \frac{\epsilon^2}{2n}$$

Posledično velja $\|(XY - I)_i\| \geq \frac{\epsilon^2}{2n}$, od koder sledi

$$\|XY - I\| \geq \frac{\epsilon^2}{2n}.$$

Iz $\|P^{-1} - R\|_F \leq \frac{\epsilon^2}{2n^{3/2}}$ sledi

$$\|PR - I\| \leq \|P\|\|P^{-1} - R\| \leq \frac{\epsilon^2}{2n}.$$

. Po zgornjem premisleku je

$$\|P - Q\| \leq \epsilon$$

za neko permutacijsko matriko Q .

Če je $\epsilon < \frac{1}{2}$, je očitno $\overline{P} = \overline{Q + (P - Q)} = Q$. \square

Zdaj lahko dokažemo glavni rezultat tega razdelka.

Izrek 3.8. Če je

$$\mathcal{L}_{\text{rel}}(\hat{\rho}) < n^{1-L}(1 + L^{\frac{3}{2}}(2^{L-1} - 1))^{-2},$$

kjer je $L = \max_{r \in R} |r|$, je inducirana preslikava $\rho: G \rightarrow \Pi$ homomorfizem.

Dokaz. Naj bo $r = s_1 s_2 \cdots s_m \in R$ in $s_1, s_2, \dots, s_m \in S$. Dovolj je pokazati, da velja

$$\overline{P_{s_1} P_{s_2} \cdots P_{s_m}} = I.$$

Če je $\mathcal{L}_{\text{rel}}(\hat{\rho}) \leq \frac{1}{|R|} - \log(1 - \frac{\epsilon}{n})$ po lemi 3.4 sledi, da je

$$P_r = P_{s_1} P_{s_2} \cdots P_{s_m} = I + E_r \quad (3.3)$$

za $\|E_r\| < \epsilon$. Če je dodatno $\epsilon < \frac{1}{2}$, je po lemi 3.5 matrika P_s obrnljiva za vsak generator $s \in S$.

Iz enačbe (3.3) sledi

$$P_{s_1}^{-1} = P_{s_2} P_{s_3} \cdots P_{s_m} - P_{s_1}^{-1} E_r. \quad (3.4)$$

Matrika $P_{s_2} P_{s_3} \cdots P_{s_m}$ je stohastična, saj je produkt stohastičnih matrik. Poleg tega je $\|P_{s_2} P_{s_3} \cdots P_{s_m} - P_{s_1}^{-1}\| = \|P_{s_1}^{-1}\| \|E_r\| < \|P_{s_1}^{-1}\| \epsilon$. Če je $\epsilon_1 = (\|P_s^{-1}\| \epsilon 2n^{\frac{3}{2}})^{\frac{1}{2}} < \frac{1}{2}$, je po lemi 3.7

$$P_{s_1} = \overline{P_{s_1}} + E_{s_1}$$

za $\|E_{s_1}\| < \epsilon_1$.

Enačbo (3.4) z desne pomnožimo z P_{s_1} , z leve pa z $P_{s_2}^{-1}$ in dobimo

$$P_{s_2}^{-1} = P_{s_3} \cdots P_{s_m} P_{s_1} - P_{s_2}^{-1} P_{s_1}^{-1} E_r P_{s_1}.$$

Če je $\epsilon_2 = (2n^{\frac{3}{2}} \|P_{s_2}^{-1}\| \|P_{s_1}^{-1}\| \|P_{s_1}\| \epsilon)^{\frac{1}{2}} < \frac{1}{2}$, je po lemi 3.7 $P_{s_2} = \overline{P_{s_2}} + E_{s_2}$. Induktivno nadaljujemo. Izberemo tak ϵ , da je za vsak $j \in [m]$ $\epsilon_j = (2n^{\frac{3}{2}} \|P_{s_j}^{-1}\| \prod_{i=1}^{j-1} \|P_{s_i}^{-1}\| \prod_{i=1}^{j-1} \|P_{s_i}\| \epsilon)^{\frac{1}{2}} < \frac{1}{2}$. Potem je

$$P_{s_j} = \overline{P_{s_j}} + E_{s_j}.$$

in $\|E_{s_j}\| < \epsilon_j$. Računamo

$$\begin{aligned} P_r &= P_{s_1} P_{s_2} \cdots P_{s_m} \\ &= (\overline{P_{s_1}} + E_{s_1})(\overline{P_{s_2}} + E_{s_2}) \cdots (\overline{P_{s_m}} + E_{s_m}) \\ &= \overline{P_{s_1} P_{s_2}} \cdots \overline{P_{s_m}} + \sum_{Z \subsetneq [m]} \prod_{i \in Z} \overline{P_{s_i}} \prod_{j \notin Z} E_{s_j} \\ &= I + E_r. \end{aligned}$$

Ocenimo

$$\begin{aligned} \left\| \prod_{i=1}^m \overline{P_{s_i}} - I \right\| &= \|E_r - \sum_{Z \subsetneq [n]} \prod_{i \in Z} \overline{P_{s_i}} \prod_{j \notin Z} E_{s_j}\| \\ &\leq \|E_r\| + \sum_{Z \subsetneq [m]} \prod_{i \in Z} \|\overline{P_{s_i}}\| \prod_{j \notin Z} \|E_{s_j}\| \\ &\leq \epsilon_m + \sum_{Z \subsetneq [m]} \prod_{i \in Z} \sqrt{m} \prod_{j \notin Z} \epsilon_m \\ &= \epsilon_m (1 + m^{\frac{3}{2}} (2^{m-1} - 1)) \xrightarrow{\epsilon_m \rightarrow 0} 0. \end{aligned}$$

Velja $\prod_{i=1}^m \overline{P_{s_i}} \in \Pi$. Razdalje med različnimi permutacijskimi matrikami ne morejo biti manjše od 1. Za dovolj majhen ϵ_m je $\prod_{i=1}^m \overline{P_{s_i}} = 1$.

Norma stohastične matrike je manjša ali enaka \sqrt{n} . S to oceno lahko ocenimo $\epsilon_j \leq (n^{j-1} \epsilon)^{\frac{1}{2}}$. Naj bo $L = \max_{r \in R} |r|$ največja dolžina relacije v R . Če je

$$\mathcal{L}_{\text{rel}}(\hat{\rho}) \leq \epsilon < n^{1-L} (1 + L^{\frac{3}{2}} (2^{L-1} - 1))^{-2},$$

je $\prod_{s \in r} \overline{P_s} = I$ za vsako relacijo $r \in R$ in je $\rho: G \rightarrow \Pi$ homomorfizem. \square

3.4 Delovanje ciklične grupe

Študiramo model za delovanje ciklične grupe $C_n = \langle z | z^n = 1 \rangle$ na končni množici $[m] = \{1, 2, \dots, m\}$. Minimaliziramo

$$\mathcal{L}(\phi) = -\text{tr}(\log(\text{softmax}(Z^n))),$$

kjer je $Z \in \mathbb{R}^{m \times m}$.

Rezultati Na sliki 7 so predstavljene začetne in končne vredosti matrike $P(Z)$ pri iskanju delovanj ciklične grupe C_n . Podatki so bili priodobljeni s `torch` implemenacijo algoritma *Adam* [20] z velikostjo koraka 10^{-2} . Parametri so bili inicializirani z vzorčenjem iz normalne porazdelitve z srednjo vrednostjo 0 in standardnim odklonom 1. Eksperiment smo prekinili po 2000 korakih gradientnega spusta ali pa ko je vrednost funkcije izgube padla pod 10^{-3} . Več rezultatov je v dodatku B.2.1.

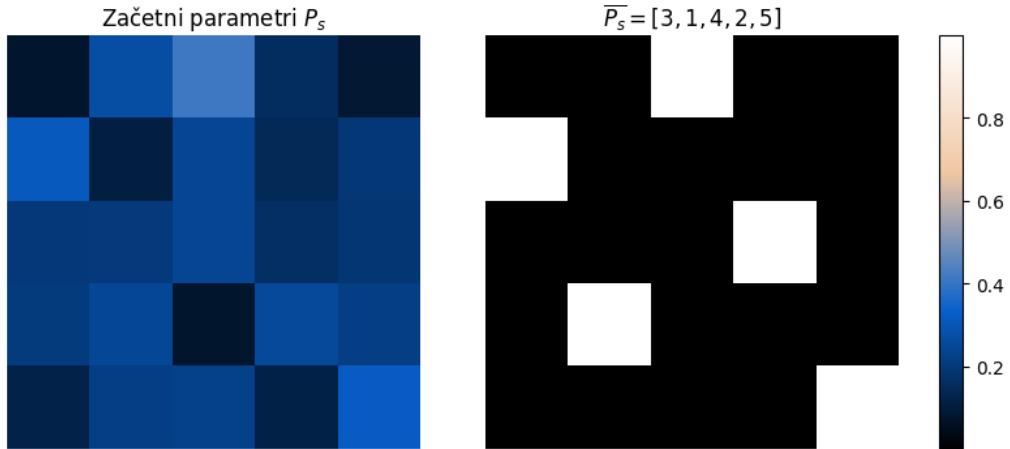
3.5 Delovanje diedrske grupe

Študiramo model za delovanje grupe $D_n = \langle r, s | r^n = s^2 = (rs)^2 = 1 \rangle$ na končni množici $[m] = \{1, 2, \dots, m\}$. Minimaliziramo

$$\mathcal{L}(R, S) = -\text{tr}(\log(P_r^n)) - \text{tr}(\log(P_s^n)) - \text{tr}(\log(P_r P_s)^2)),$$

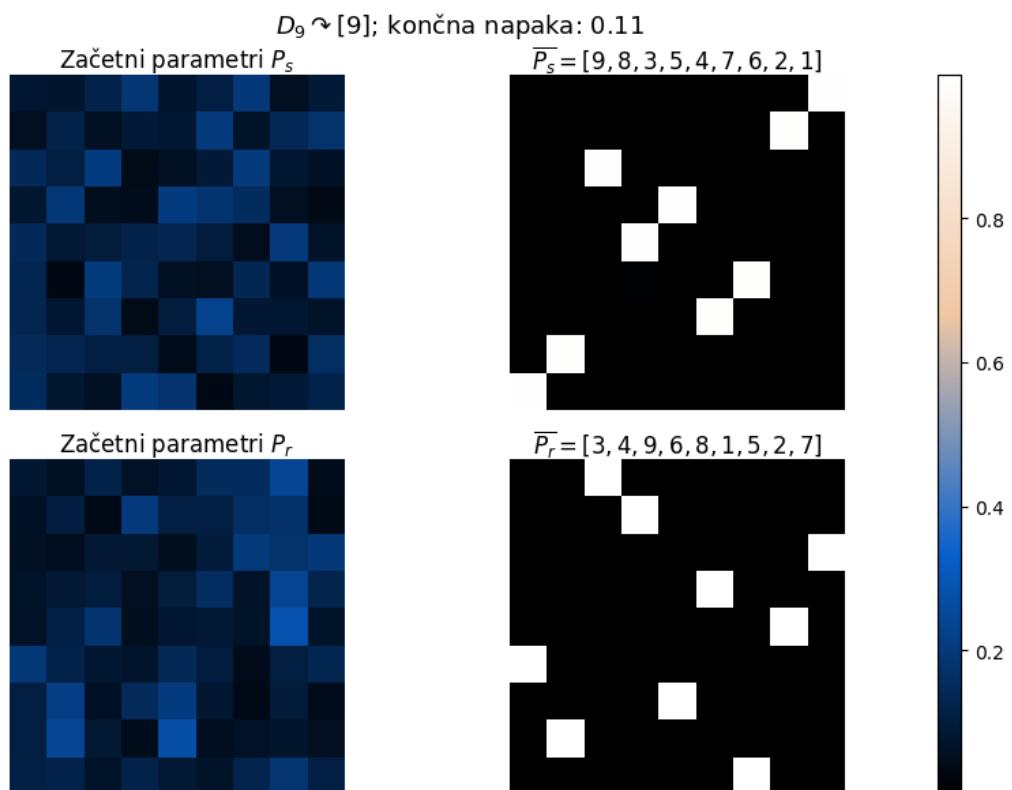
kjer je $P_r = \text{softmax}(R)$, $P_s = \text{softmax}(S)$ in $R, S \in \mathbb{R}^{m \times m}$.

$C_8 \cong [5]$; končna napaka: 0.06



Slika 7: Stohastična matrika P_s na začetku in po 2000 korakih gradientnega spusta med iskanjem delovanja ciklične grupe C_8 na množici [5]. Permutacijska matrika \bar{P}_s pripada permutaciji $(1342)(2)$ reda 4 in je ničla funkcije izgube.

Rezultati Na sliki 8 so matrike P_r in P_s na začetku in po 2000 korakih gradientnega spusta med iskanjem delovanja diedrske grupe D_9 na množici [9]. Več rezultatov je v dodatku B.2.2.



Slika 8: Stohastični matriki P_r in P_s na začetku in po 2000 korakih gradientnega spusta med iskanjem delovanja diedrske grupe D_9 na množici [9]. Permutacijski matriki \overline{P}_r in \overline{P}_s pripadatata permutacijama $\sigma_s = (123456789)$ in $\sigma_r = (19)(28)(37)(46)(5)$. Velja $\sigma_s^2 = \sigma_r^9 = (\sigma_r \sigma_s)^2 = 1$, torej je $s \mapsto \sigma_s, r \mapsto \sigma_r$ homomorfizem.

4 Iskanje izomorfizmov grafov

S podobnim pristopom kot v 2.3 lahko iščemo izomorfizme grafov. Iščemo model za bijektivno preslikavo $f: G_1 \rightarrow G_2$ med grafoma G_1 in G_2 , za katero velja $f(i) \sim_2 f(j) \iff i \sim_1 j$.

Za model lahko vzamemo slučajno preslikavo f , podano s stohastično matriko

$$P_f = \begin{bmatrix} P(f(1) = 1) & P(f(1) = 2) & \cdots & P(f(1) = n) \\ P(f(2) = 1) & P(f(2) = 2) & \cdots & P(f(2) = n) \\ \vdots & \vdots & & \vdots \\ P(f(n) = 1) & P(f(n) = 2) & \cdots & P(f(n) = n) \end{bmatrix}$$

in spremojamo elemente matrike tako, da je f vse bolj verjetno izomorfizem. Zato potrebujemo funkcijo izgube, ki je ničelna natanko v izomorfizmih - torej bijektivnih preslikavah, ki slikajo povezane vozlišča v povezane.

4.1 Funkcija izgube za bijektivnost

Naj bo f slučajna preslikava iz $\text{fun}([n], [n])$, definirana kot zgoraj. Radi bi, da je f skoraj zagotovo bijektivna, torej da velja

$$\begin{aligned} 1 &= P(f \in S_n) \\ &= \sum_{\sigma \in S_n} P(f = \sigma) \\ &= \sum_{\sigma \in S_n} \prod_{i=1}^n (P_f)_{i, \sigma(i)} \\ &= \text{Perm}(P_f), \end{aligned}$$

kjer $\text{Perm}(P_f)$ označuje *permanent*¹⁵ matrike P_f .

Za stohastične matrike velja spodnja lema.

Lema 4.1. Za stohastično matriko P velja $\text{Perm}(P) = 1$ natanko tedaj, ko je P permutacijska.

Dokaz. Če je P permutacijska, ima očitno enotski permanent. Dokažimo lemo še v obratno smer.

Naj bo $P = [p_{i,j}]_{i,j=1,\dots,n}$ poljubna stohastična matrika s $\text{Perm}(P) = 1$. Po definiciji so vsi elementi $p_{i,j}$ matrike med 0 in 1, vrstice se pa seštejejo v 1. Velja

$$\text{Perm}(P) = \sum_{\sigma \in S_n} \prod_{i=1}^n p_{i, \sigma(i)} \leq \sum_{\sigma \in \text{fun}([n], [n])} \prod_{i=1}^n p_{i, \sigma(i)},$$

saj so vsi elementi stohastične matrike nenegativni, in smo vsoti zgolj prišteli nene-gativne elemente.

¹⁵ $\text{Perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^n A_{i, \sigma(i)}$

Vsako funkcijo $\sigma \in \text{fun}([n], [n])$ lahko predstavimo z vektorjem $(\sigma_1, \sigma_1, \dots, \sigma_n)$, kjer je $\sigma_i = \sigma(i) \in \mathbb{N}$. Tako je

$$\sum_{\sigma \in \text{fun}([n], [n])} \prod_{i=1}^n p_{i, \sigma(i)} = \sum_{(\sigma_1, \sigma_1, \dots, \sigma_n) \in [n]^n} \prod_{i=1}^n p_{i, \sigma_i} = \prod_{i=1}^n (p_{i,1} + p_{i,2} + \dots + p_{i,n}).$$

Ker so vsote vrstic enake 1, velja

$$1 = \text{Perm}(P) \leq \sum_{\sigma \in \text{fun}([n], [n])} \prod_{i=1}^n p_{i, \sigma(i)} = \prod_{i=1}^n (p_{i,1} + p_{i,2} + \dots + p_{i,n}) = 1.$$

Sledi, da je $\sum_{\sigma \in S_n} \prod_{i=1}^n p_{i, \sigma(i)} = \sum_{\sigma \in \text{fun}([n], [n])} \prod_{i=1}^n p_{i, \sigma(i)}$ in posledično je vsota $\sum_{\sigma \in \text{fun}([n], [n]) \setminus S_n} \prod_{i=1}^n p_{i, \sigma(i)}$ enaka 0.

Zdaj lahko dokažemo, da je v poljubnem stolpcu matrike P lahko največ en neničelen element. Naj bo i indeks poljubnega stolpca in $r \neq s$ indeksa dveh vrstic. Recimo, da sta elementa $p_{r,i}$ in $p_{s,i}$ oba neničelna.

Definiramo lahko preslikavo $f \in \text{fun}([n], [n])$, da je $f(r) = f(s) = i$, za $j \notin \{r, s\}$ pa $f(j)$ izberemo tako, da $p_{j,f(j)}$ ni ničelen (to lahko vedno izberemo, saj je matrika stohastična, torej ima v vsaki vrstici vsaj en neničelen element).

Očitno je $\prod_{i=1}^n p_{i, f(i)} \neq 0$, kar pa je v protislovju z $\sum_{\sigma \in \text{fun}([n], [n]) \setminus S_n} \prod_{i=1}^n p_{i, \sigma(i)}$. V vsakem stolpcu matrike je torej natanko en neničelen element.

Velja še $1 = P(f \in S_n) = P(f \text{ je surjektivna}) = \prod_{i=1}^n P(\exists j \in \mathbb{N}. f(j) = i) = \prod_{i=1}^n (\sum_{j=1}^n p_{j,i})$, torej se vsote stolpcev zmnožijo v 1. Ker je v vsakem stolpcu natanko en neničelen element sledi, da je ta element enak 1. Posledično je tudi v vsaki vrstici natanko en neničelen element (sicer se vrstica ne bi seštela v 1), kar pomeni, da je P permutacijska. \square

Slučajna preslikava f je torej skoraj zagotovo bijektivna natanko tedaj, ko je P_f permutacijska. Enostavno je preveriti, da so stohastične matrike permutacijske natanko tedaj, ko so unitarne. Bolj, kot bo P_f unitarna, bolj verjetno bo f bijektivna.

Če P_f definiramo kot

$$P_f = \text{softmax}(\phi)$$

(glej poglavje 3.1), lahko za funkcijo izgube vzamemo $\phi \mapsto \mathcal{L}_{\text{unitary}}(P_f)$.

Definicija 4.2. Za poljubno realno matriko $\phi \in \mathbb{R}^{n \times n}$ funkcijo izgube za bijektivnost definiramo kot

$$\mathcal{L}_{\text{bijective}} = \mathcal{L}_{\text{unitary}} \circ \text{softmax}.$$

Z minimiziranjem iste funkcije izgube lahko rešujemo dva različna problema (problem iskanja unitarnih matrik in problem iskanja bijekcij) preslikava softmax pa pretvarja med problemoma.

4.2 Funkcija izgube za povezanost

Naj bosta $G_1 = ([n], E_1)$ in $G_2 = ([n], E_2)$ grafa in M_1, M_2 njuni matriki sosednosti. Za $i \sim_1 j$ bi radi, da je $P(f(i) \sim_2 f(j)) = 1$. Računamo

$$P(f(i) \sim_2 f(j)) = \sum_{k=1}^n \sum_{h=1}^n p_{i,k} p_{j,h} m_{k,h}^{(2)} = (PM_2P^T)_{i,j}.$$

Definicija 4.3. Žeeli bi si, da je $P(f(i) \sim_2 f(j)) = (M_1)_{i,j}$. Funkcijo izgube za povezanost lahko definiramo kot

$$\mathcal{L}_\sim = \|M_1 - PM_2P^T\|.$$

Naš cilj je poiskati permutacijsko matriko P , ki minimalizira $\mathcal{L}_\sim(P)$. Rešujemo problem *ujemanja grafov*.

Definicija 4.4. Naj bosta M_1 in M_2 matriki sosednosti dveh grafov na n vozljivih. Problem ujemanja graphov (*graph matching*) [17] je definiran kot iskanje minimuma

$$\operatorname{argmin}_{P \in \Pi} \|M_1 - PM_2P^T\|,$$

po množici vseh $n \times n$ permutacijskih matrik Π .

Če matriki M_1 in M_2 pripadata izomorfnima grafoma, potem iščemo izomorfizem.

Problem *ujemanja grafov* je NP-težek [23]. Klasični pristopi ga rešujejo z relaksacijami funkcije izgube $\|M_1 - PM_2P^T\|$ nad večje družine matrik P in prevedbo končne matrike nazaj na permutacijsko [17], [1]. Novejše metode se problema lotujejo z globokim učenjem.

V tem delu se s primerjavo različnih metod za ujemanje grafov ne ukvarjam. Osredotočeni smo na posplošitev metod iz poglavij 2.2 in 3 .

4.3 Prehod na izomorfizem

Kot v poglavju 3.3 lahko tudi model izomorfizma grafov pretvorimo v dejansko izomorfizem, če je le vrednost funkcije izgube dovolj majhna.

Definicija 4.5. Za majhno vrednost $\mathcal{L}_{\text{unitary}}(P) + \mathcal{L}_\sim(P)$ model P inducira izomorfizem med grafoma, podan s permutacijsko matriko \bar{P} , ki ima v vrstici i enico na mestu $\operatorname{argmax}_j(P_{i,j})$.

Treba je preveriti, da je \bar{P} res bijekcija in da predstavlja izomorfizem.

Lema 4.6. *Naj bo $P = QH, Q \in U_n, H = \sqrt{PP^*}$ polarni razcep matrike P . Če je $\mathcal{L}_{\text{unitary}}(P) = \|PP^* - I\| < \epsilon$, je $\|P - Q\| < \epsilon$. Če je $\epsilon < \frac{1}{2}$, je $\bar{P} = Q$.*

Dokaz. Glej dokaz leme 2.20. □

Lema 4.7. *Za preslikavo*

$$f: X \mapsto \|M_1 - XM_2X^T\|$$

velja $|f(X) - f(Y)| \leq \|M_2\|(\|X\| + \|Y\|)\|X - Y\|$.

Dokaz. Ocenimo

$$\begin{aligned} |f(X) - f(Y)| &\leq \|(YM_2Y^T) - (XM_2X^T)\| \\ &\leq \|(Y - X)M_2Y^T\| + \|XM_2(Y^T - X^T)\| \\ &= \|M_2\|(\|X\| + \|Y\|)\|X - Y\|. \end{aligned}$$

□

Lema 4.8. Če je

$$\mathcal{L}_\sim(P) \leq \epsilon < \min\left(\frac{1}{2}, \frac{\sqrt{2}}{1 + \|M_1\|\sqrt{n}}\right),$$

je permutacija, podana z \bar{P} izomorfizem med grafoma z matrikama sosednosti M_1 in M_2 .

Dokaz. Po lemi 4.6 je $\|P - \bar{P}\| < \epsilon$. Po lemi 4.7 pa velja

$$|f(P) - f(\bar{P})| \leq \|M_2\|(\|P\| + \|\bar{P}\|)\epsilon < \|M_2\|(1 + \sqrt{n})\epsilon$$

. Vrednost funkcije izgube v \bar{P} je omejena:

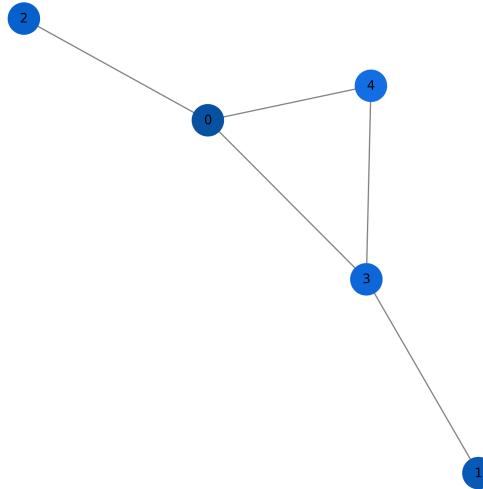
$$f(\bar{P}) < \epsilon(1 + \|M_2\|\sqrt{n})$$

Opazujmo funkcijo izgube $f: X \mapsto \|M_1 - XM_2X^T\|$. Če je X permutacijska matrika, je XM_2X^T matrika sosednosti grafa, ki ga dobimo, če vozljšča grafa M_2 preslikamo z permutacijo X . Vrednosti funkcije izgube nad množico permutacijskih matrik so torej diskretne. $\|M_1 - XM_2X^T\| = 0$ natanko tedaj, ko X predstavlja izomorfizem med grafoma M_1 in M_2 , sicer pa je $\|M_1 - XM_2X^T\| \geq \sqrt{2} > 1$.

Če je $\epsilon < \min\left(\frac{1}{2}, \frac{\sqrt{2}}{1 + \|M_1\|\sqrt{n}}\right)$, je $f(\bar{P}) = 0$. \square

4.4 Rezultati

Primer 4.9. Poglejmo si preprost graf 9 na petih točkah $\{0, 1, 2, 3, 4\}$, s povezavami $\{0 \sim 4, 4 \sim 3, 3 \sim 0, 3 \sim 1, 2 \sim 0\}$. Očitno je grupa avtomorfizmov tega grafa izomorfnna diedrski grupi trikotnika D_3 . Gradientni spust poleg trivialne rešitve



Slika 9: Graf na petih vozljščih.

na določenih začetnih parametrih uspe najti avtomorfizem $(03)(24)$, na drugih pa konvergira do limite, ki ni blizu avtomorfizma.

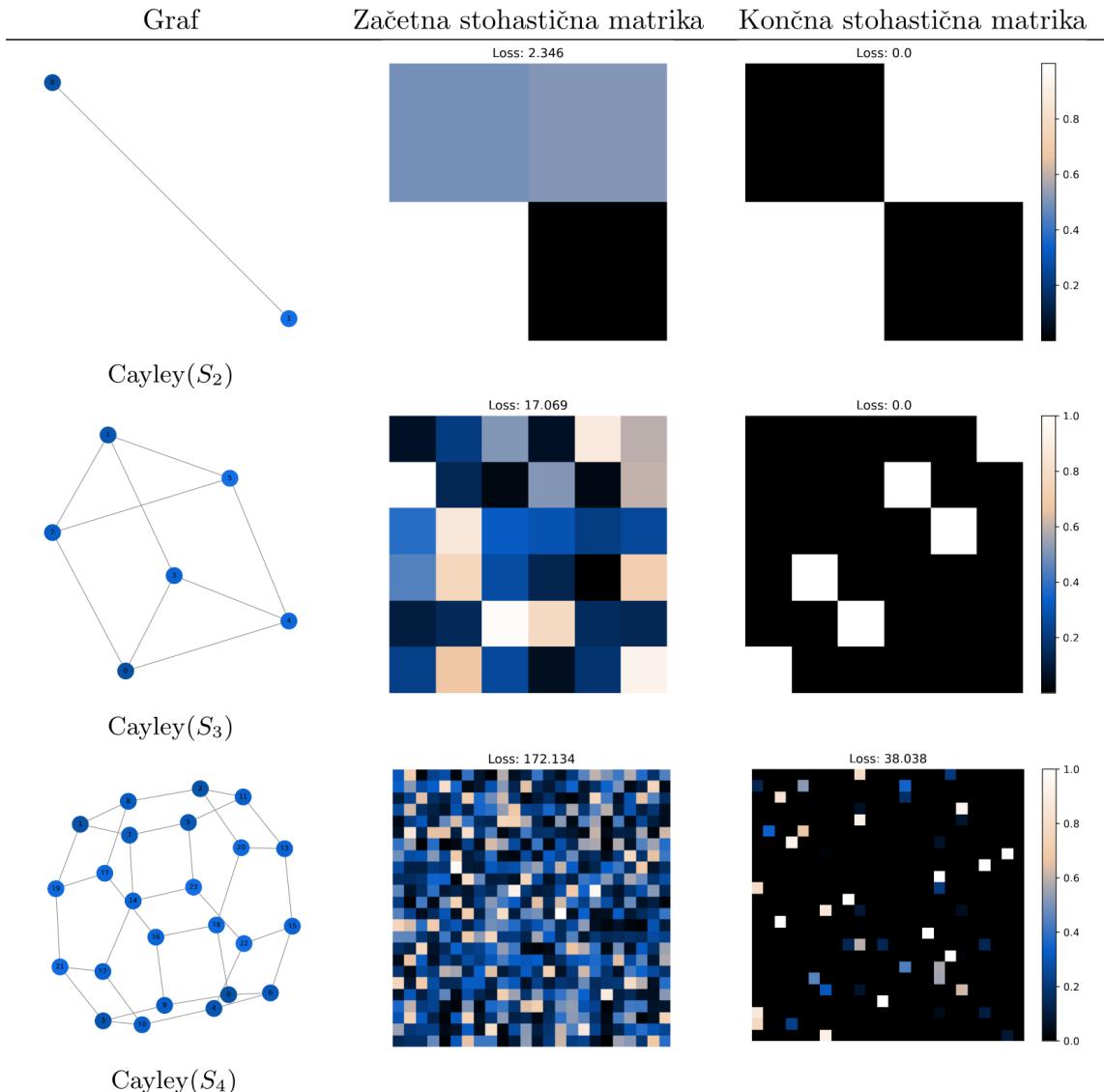
\diamond

Avtomorfizmi cayleyevih grafov Z metodo iščemo avtomorfizme Cayleyevih grafov simetrične grupe. Oštevilčenje vozljisč (elementov S_n) sledi metodi

```
list(itertools.permutations(range(1, n + 1))),
```

za množico generatorjev vzamemo transpozicijo (12) in cikel $(12 \dots, n)$, za $M_1 = M_2$ vzamemo matriko sosednosti dobljenega grafa.

Funkcijo izgube optimiziramo z algoritmom Adam s hitrostjo učenja 10^{-3} , dokler norma gradienta ne pade pod 10^{-12} . Na sliki 10 so prikazane matrike P na začetku in na koncu optimizacije za različne n . Več rezultatov je v dodatku B.3.1.



Slika 10: Stohastična matrika P na začetku in po končani optimizaciji med iskanjem avtomorfizmov Cayleyevih grafov simetrične grupe S_n za $n = 2, 3, 4$. Pri $n = 2$ smo našli avtomorfizem (01), pri $n = 3$ avtomorfizem (05)(13)(24).

4.5 Uporaba tabele inverzij

V poglavju 4.4 izomorfizem iščemo v podprostoru distribucij nad $\text{fun}([n], [n])$, ki ga parametriziramo s $(\mathbb{R}^{n \times n})^{|S|}$. V tem poglavju predstavimo alternativen pristop, ki

izomorfizem išče v podprostoru distribucij nad S_n . S tem se ognemo potrebi po uporabi funkcije izgube za bijektivnost.

Množico permutacij predstavimo s tabelami inverzij. Nad slednjimi lahko na trivialen način definiramo gladko družino porazdelitev z $O(n^2)$ parametri.

4.5.1 Porazdelitve nad tabelo inverzij

Definicija 4.10. (Tabela inverzij) Naj bo $\sigma \in S_n$ permutacija. Za vsak $k \in [n]$ definiramo $a_k = |\{j > k \mid \sigma(j) < \sigma(k)\}|$. Vektor

$$\text{TI}(\sigma) = (a_1, a_2, \dots, a_n)$$

se imenuje *tabela inverzij* permutacije σ .

Tabela inverzij $TI: S_n \rightarrow [0, n-1] \times [0, n-2] \times \dots \times [0]$ je bijekcija [26, Trditev 1. 3. 12]. Porazdelitev vektorjev iz $\prod_{i=n-1}^0 [0, i]$ trivialno inducira porazdelitev nad S_n .

Tabelo inverzij lahko razumemo kot proces tvorbe permutacij. Permutacijo σ s tabelo inverzij $\text{TI}(\sigma) = (a_1, a_2, \dots, a_n) \in \prod_{i=n-1}^0 [0, i]$ dobimo tako, da 1 postavimo na $(a_1 + 1)$ -to mesto v v tabeli $[1, 2, \dots, n]$, nato 2 postavimo na $(a_2 + 1)$ -to prsto mesto, in tako naprej, dokler ne postavimo n na $(a_n + 1)$ -to prsto mesto.

Primer 4.11. Naj bo $\text{TI}(\sigma) = (2, 1, 0)$. Začnemo s prazno tabelo

$$[_, _, _]$$

dolžine 3. Postavimo 1 na 3-to mesto

$$[_, _, 1]$$

nato 2 na 3-to prsto mesto

$$[_, 2, 1]$$

in na koncu 3 na 2-to prsto mesto

$$[3, 2, 1].$$

Velja $\sigma = (13)(2)$. ◇

Na podoben način lahko definiramo porazdelitev nad permutacijami. Naj bo

$$P_d = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{2,n-1} & p_{1,n} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n-1} & 0 \\ p_{3,1} & p_{3,2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{n-1,1} & p_{n-1,2} & \cdots & 0 & 0 \\ p_{n,1} & 0 & \cdots & 0 & 0 \end{bmatrix}$$

stohastična matrika. Potem lahko definiramo

$$P(a_k = i) = P(k \text{ postavimo na } i\text{-to prazno mesto}) = p_{k,i}$$

in

$$P(\text{TI}(\sigma) = (x_1, x_2, \dots, x_n)) = \prod_{i=1}^n P(a_k = x_k).$$

Primer 4.12. Naj bo

$$P = \begin{bmatrix} a & b & c \\ d & e & 0 \\ f & 0 & 0 \end{bmatrix}$$

stohastična matrika, ki podaja porazdelitev nad S_3 . Verjetnost za permutacijo $\sigma = (13)(2)$ iz primera 4.11 je

$$P(TI(\sigma) = (2, 2, 1)) = P(a_1 = 2)P(a_2 = 2)P(a_3 = 1) = cef$$

◊

4.5.2 Funkcija izgube

Naj bosta M_1 in M_2 matriki sosednosti dveh grafov na n vozljščih. Za povezani vozljšci $i \sim_1 j$ bi radi, da je $P(f(i) \sim_2 f(j)) = 1$, kjer je f slučajna permutacija.

Za funkcijo izgube lahko vzamemo

$$\mathcal{L} = \sum_{i,j} ((M_1)_{i,j} - P(f(i) \sim f(j)))^2.$$

Očitno velja

$$P(f(i) \sim f(j)) = \sum_{k=1}^n \sum_{h=1}^n P(f(i) = k, f(j) = h) (M_2)_{k,h}.$$

4.5.3 Hiter izračun funkcije izgube

Zanima nas $P(f(i) = k, f(j) = h)$. Najprej uvedemo pomožne funkcije:

$$\begin{aligned} S_m^<(i) &= \sum_{s=1}^i P(a_m = s) \\ S_m^{\text{mid}}(i, j) &= \sum_{s=i}^j P(a_m = s) \\ S_m^>(j) &= \sum_{s=j}^n P(a_m = s) \end{aligned}$$

Recimo, da smo že postavili $m - 1$ elementov in postavljam m -ti element. S $q(j, h, m)$ označimo verjetnost, da bomo med tvorjenjem slučajen premutacije h postavili na j -to trenutno prosto mesto. Podobno s $p(i, k, j, h, m)$ označimo verjetnost, da bomo med tvorjenjem slučajne permutacije k postavili na i -to in h na j -to trenutno prosto mesto. Očitno je $P(f(i) = k, f(j) = h) = p(i, k, j, h, 1)$.

Za $h \notin [m, n]$ ali $j \notin [1, n - m + 1]$ je $q(j, h, m) = 0$, sicer pa

$$q(j, h, m) = \begin{cases} P(a_m = j) & m = h \\ q(j - 1, h, m + 1)S_m^<(j - 1) + & \text{sicer} \\ q(j, h, m + 1)S_m^>(j + 1) & \end{cases}$$

Za p velja $p(i, k, j, h, m) = 0$, če velja ena od $k = h$, $k \notin [m, n]$, $h \notin [m, n]$, $i \notin [1, n - m + 1]$, $j \notin [1, n - m + 1]$. Sicer pa je

$$p(i, k, j, h, m) = \begin{cases} P(a_k = i)q(j - 1, h, k + 1) & m = k \\ P(a_h = j)q(i, k, h + 1) & m = h \\ p(i - 1, k, j - 1, h, m + 1)S_m^<(i - 1) + & \text{sicer} \\ p(i, k, j - 1, h, m + 1)S_m^{\text{mid}}(i + 1, j - 1) + \\ p(i, k, j, h, m + 1)S_m^>(j + 1) \end{cases}$$

Z uporabo zgornjih rekurzivnih enačb lahko funkcijo izgube izračunamo v času $O(n^5)$. S pametnim preoblikovanjem funkcije izgube lahko časovno zahtevnost znižamo na $O(n^3)$.

Izračun v $O(n^3)$ Računamo $\mathcal{L} = \sum_{i,j} ((M_1)_{i,j} - \sum_{k,h} p(i, k, j, h, 1)(M_2)_{k,h})^2$. Definiramo pomožne funkcije

$$\begin{aligned} U_{j,m,k} &= \sum_h q(j, h, m)(M_2)_{k,h} \\ V_{i,m,h} &= \sum_k q(i, k, m)(M_2)_{k,h} \\ T_{i,j,m} &= \sum_{k,h} p(i, k, j, h, m)(M_2)_{k,h} \end{aligned}$$

Očitno je $\mathcal{L} = \sum_{i,j} ((M_1)_{i,j} - T_{i,j,0})^2$. Iz rekurzivnih zvez za p in q sledi

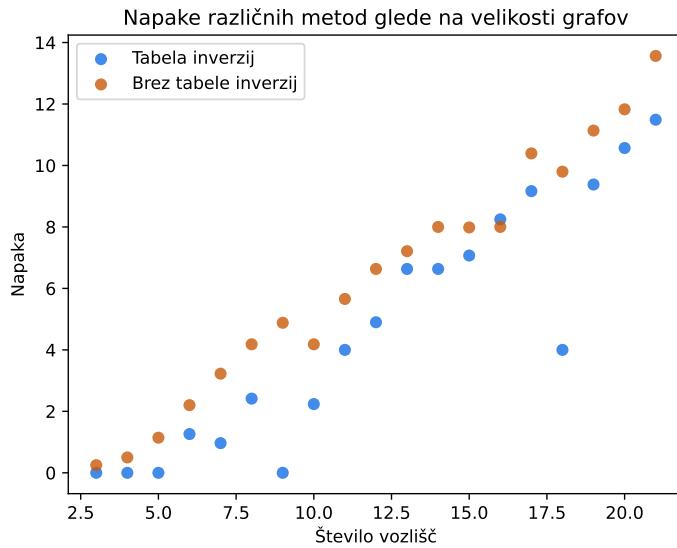
$$\begin{aligned} T_{i,j,m} &= + \sum_h P(a_m = i)q(j - 1, h, m + 1)(M_2)_{m,h} \\ &\quad + \sum_h P(a_m = j)q(i, k, m + 1)(M_2)_{k,m} \\ &\quad + \sum_{k,h} p(i - 1, k, j - 1, h, m + 1)S_m^<(i - 1)(M_2)_{k,h} \\ &\quad + \sum_{k,h} p(i, k, j - 1, h, m + 1)S_m^{\text{mid}}(i + 1, j - 1)(M_2)_{k,h} \\ &\quad + \sum_{k,h} p(i, k, j, h, m + 1)S_m^>(j + 1)(M_2)_{k,h} \\ &= P(a_m = i)U_{j-1,m+1,m} + P(a_m = j)V_{i,m+1,m} \\ &\quad + T_{i-1,j-1,m+1}S_m^<(i - 1) + T_{i,j-1,m+1}S_m^{\text{mid}}(i + 1, j - 1) + T_{i,j,m+1}S_m^>(j + 1) \end{aligned}$$

Z uporabo teh formul lahko funkcijo izgube izračunamo v času $O(n^3)$. Če želimo še hitrejši izračun, lahko naključno izberemo podmnožico indeksov (i, j) in izračunamo stohastični približek funkcije izgube na podmnožici.

Naj bo $S \subset [n] \times [n]$ naključno izbrana podmnožica indeksov. Za približek funkcije izgube lahko vzamemo

$$\mathcal{L}_{\approx} = \sum_{(i,j) \in S} ((M_1)_{i,j} - T_{i,j,1})^2.$$

Časovna zahtevnost izračuna približka je $O(|S|n^2)$.



Slika 11: Primerjava napak glede na velikost grafa. Za vsak n sta metodi preizkušeni na istih 10-tih naključno izbranih grafih. Vsaka točka je povprečje napak na teh grafih.

Primerjava Na sliki 11 je prikazana primerjava metode iz poglavja 4.4 in metode z uporabo tabele inverzij. Z uporabo tabele inverzij dosežemo boljše rezultate.

4.5.4 Preparametrizacija z nevronskimi mrežami

Verjetnost uspešne konvergencije lahko poizkusimo izboljšati s preparametrizacijo (*overparametrisation*) proplema. Matriko P definiramo kot izhod globoke nevronske mreže (glej dodatek A) in spremojmo njene parametre.

Preparametrizacija z nevronskimi mrežami deluje za *gradient dominante* funkcije izgube [2], glej A. Za splošne funkcije izgube obstajajo empirični rezultati, ki kažejo, da z rahlo overparametrizacijo in gradientnim spustom lahko dosežemo uspešno konvergenco [22, 25] do globalnega minimuma. Poleg tega empirični rezultati [19] kažejo, da so pri nekonveksni parametrizaciji funkcij izgube v velikih dimenzijah sredina bolj razširjena od lokalnih minimumov. Z večanjem števila parametrov se v reliefu funkcije izgube pojavlja več sedel, iz katerih lahko pobegnemo s preturbiranim gradientnim spustom, predstavljenim v poglavju 1.5.

Empirično testiramo dve izbiri preparametrizacije z nevronskimi mrežami.

Definicija 4.13. *Blaga preparametrizacija* modela izomorfizma grafov iz poglavja 4.5 velikosti n definiramo kot nevronsko mrežo z vhodom velikosti n , skrito plastjo velikosti n in izhodom velikosti $\frac{n^2-n-2}{2}$. Aktivacijska funkcija v skriti plasti je ReLu. Na zadnji plasti oblike $(x_{1,1}, \dots, x_{1,n}, x_{2,1}, \dots, x_{2,n-1}, \dots, x_{n-1,1}, x_{n-1,2})^T$ uporabimo softmax po blokih

$$(p_{i,1}, \dots, p_{i,n-i+1}) = \text{softmax}(x_{i,1}, \dots, x_{i,n-i+1})$$

, kjer je $p_{i,j}$ verjetnost, da i postavimo na j -to prosto mesto, če smo prej postavili

$i - 1$ elementov. Velja

$$M_{\text{blaga preparametrizacija}} = \text{Softmax}_{\text{po blokih}} \circ W_2 \circ \text{ReLU} \circ W_1,$$

kjer sta $W_1: \mathbb{R}^n \rightarrow \mathbb{R}^n$ in $W_2: \mathbb{R}^n \rightarrow \mathbb{R}^{\frac{n^2-n-2}{2}}$ afni preslikavi. Distribucijo P_d nad tabelami inverzij definiramo kot

$$\text{flatten}(P_d) = M_{\text{blaga preparametrizacija}}(x_0).$$

Vektor $x_0 \in \mathbb{R}^n$ in afni preslikavi W_1, W_2 so parametri modela.

Definicija 4.14. *Globoko preparametrizacijo* definiramo definiramo kot 5 plastno nevronske mrežo z vohdno in skritimi plastmi velikosti n^2 in izhodno plastjo velikosti $\frac{n^2-n-2}{2}$. Aktivacijske funkcije v skritih plasteh so ReLu, na zadnjih plasteh uporabimo softmax po blokih, kot v blagi preparametrizaciji. Distribucijo P_d nad tabelami inverzij izračunamo kot

$$\text{flatten}(P_d) = \text{softmax}_{\text{po blokih}} \circ W_4 \circ \text{ReLU} \circ w_3 \circ \text{ReLU} \circ W_2 \circ \text{ReLU} \circ W_1(y_0).$$

Parametri modela so $y_0 \in \mathbb{R}^{n^2}$ in affine preslikave W_i .

Primerjava metod Zgornje metode primerjamo s preparametrizacijo z grafovskimi konvolucijskimi mrežami (definicija A.7), predstavljeno v [3]. Gre za sodoben pristop k problemu ujemanja grafov z uporabo optimalnega transporta in nevronske mreže. Ta pristop namesto klasičnih nevronske mreže uporablja modele, prilagojene za delo z grafi (definicija A.4), namesto distribucije nad S_n pa uporablja *Sinkhornov* algoritom (definicija A.5), ki poljubne matrike zvezno preslika v *dvojno stohastične*¹⁶. Točna formulacija je v dodatku A.2.

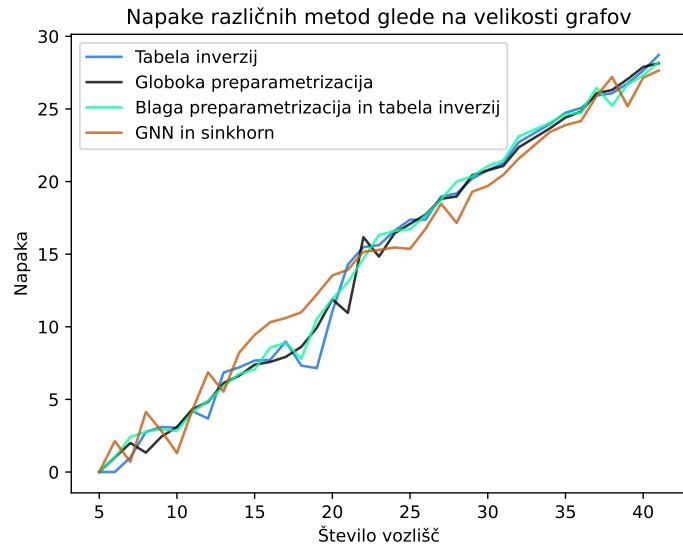
Slika 12 prikazuje primerjavo uporabe tabele inverzij brez preparametrizacije z blago in globoko preparametrizacijo. Za vsak n so metode preizkušene na istih 10-tih naključno izbranih grafih. Po koncu optimizacije izračunamo napako najbolj verjetne permutacije in jo povprečimo po 10-tih grafih.

Vsaka krivulja prikazuje rast povprečne napake najbolj verjetne preslikave glede na velikost grafa. Tabela 2 prikazuje povprečne napake metod na celotnem naboru grafov.

Za optimizacijo je bil uporabljen naključno preturbiran Adam z hitrostjo učenja 10^{-3} in z razporejevalnikom učne hitrosti `CosineAnnealingWarmRestarts`. Na posameznem grafu se je optimizacija izvajala 1000 korakov in se prekinila, če je norma gradienta padla pod $\epsilon_\nabla = 10^{-3}$.

TODO: Kaj več o graph matchingu? Uničevanje simetrij z naključnim šumom? Sinkhorn vs tabela inverzij?

¹⁶Stohastična matrika P je dvojno stohastična, če je P^T stohastična.



Slika 12: Primerjava napak različnih metod glede na velikost grafa.

Metoda	Povprečna napaka
Tabela inverzij	11.05
Blaga preparametrizacija	11.27
Globoka preparametrizacija	10.94
GNN in sinkhorn	11.67

Tabela 2: Povprečne napake metod.

A Nevronske mreže

Definicija A.1 (Nevronska mreža). Naj bo za vsak $i \in [l]$ $W_i: \mathbb{R}^{d_{i-1}} \rightarrow \mathbb{R}^{d_i}$ afina¹⁷ preslikava. Za vsak $i \in \{1, 2, \dots, l\}$ in $x \in \mathbb{R}^{d_{i-1}}$ velja $W_i(x) = \Omega_i x + b_i$ za $\Omega_i \in \mathbb{R}^{d_{i-1} \times d_i}$ in $b_i \in \mathbb{R}^{d_i}$.

Naj bodo $\sigma_1, \dots, \sigma_l$ realne funkcije v eni spremenljivki. Za poljuben $d \in \mathbb{N}$ definiramo $\sigma_i: \mathbb{R}^d \rightarrow \mathbb{R}^d$ po elementih s predpisom $\sigma(x)_i = \sigma(x_i)$. Preslikavam σ_i rečemo *aktivatorske funkcije*. Običajno aktivacijske funkcije izberemo izmed ReLU¹⁸, softmax, arctan, GeLU.

Nevronska mreža globine l je model

$$M_\phi = \sigma_l \circ W_l \circ \sigma_{l-1} \circ W_{l-1} \circ \dots \circ \sigma_1 \circ W_1: \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_l},$$

odvisen od parametrov $\phi = \Omega_0, b_0, \Omega_1, b_1, \dots, \Omega_l, b_l$. Če je $l \geq 1$ rečemo, da je mreža globoka.

Definicija A.2 (Regresija). Eden od klasičnih problemov, pri katerem se uporablajo nevronske mreže, je *regresija*.

Za slučajni spremenljivki $X \in \mathbb{R}^{d_1}$ in $Y \in \mathbb{R}^{d_2}$ iščemo model $\hat{Y}(X)$, ki čim bolje opiše $E[Y|X]$.

Predpostavimo, da imamo množico vzorcev $S_{\text{train}} = \{(x_i, y_i) \mid i \in [n]\}$ slučajne spremenljivke (X, Y) in nevronska mreža $M_{\phi_0}: \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_1}$ z naključno izbranimi začetnimi parametri ϕ in z gradientnim spustom minimaliziramo

$$\mathcal{L}_{\text{regresija}}(\phi) = \frac{1}{|S_{\text{train}}|} \sum_{(x,y) \in S_{\text{train}}} \|M_\phi(x) - y\|^2.$$

Uspešnost modela nato testiramo na množici vzorcev $S_{\text{test}} \subset S_{\text{train}}^c$.

V splošnem lahko za par (x, y) in nevronska mreža M_ϕ definiramo poljubno funkcijo izgube $f(z = M_\phi(x), y)$ in minimaliziramo

$$\mathcal{L}(\phi) = \frac{1}{|S_{\text{train}}|} \sum_{(x,y) \in S_{\text{train}}} f(M_\phi(x), y).$$

A.1 Nevronske mreže dosežejo globalni minimum funkcije izgube

Če za gladko funkcijo izgube $f(z, y)$ velja *Polyak-Lojasiewiczev pogoj*

$$\|\nabla f(z, y)\|^2 \geq \mu(f(z, y) - f(\min_z(f(z, y)), y)),$$

dovolj globoke nevronske mreže s gradientnim spustom vedno najdejo parametre, ki dosegajo poljubno natančen približek globalnega minimuma funkcije izgube. Spodnji izrek je povzet po [2, stran 6, izrek 1].

¹⁷ $W_i(x) = A_i x + b_i$ za $A_i \in \mathbb{R}^{d_{i-1} \times d_i}$ in $b_i \in \mathbb{R}^{d_i}$.

¹⁸ $\text{ReLU}(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases}$

Izrek A.3 (Allen-Zhu, Li, Song). Naj $f(z, y)$ izpolnjuje Polyak-Lojasiewiczev. Naj bo S_{train} taka množica vzorcev, da obstaja $\delta > 0$, da za dva vzorca $(x_i, y_i), (x_j, y_j) \in S_{train}$ z različnima indeksoma $i \neq j$ velja $\|x_i - x_j\| > \delta$.

Naj bo $M = \sigma \circ W_l \circ \sigma \circ W_{l-1} \circ \dots \circ \sigma \circ W_1$ taka l -plastna nevronskna mreža, da velja $d_1 = d_2 = \dots = d_{l-1} = m \in \mathbb{N} > d_0 \cdot \text{poly}(n, l, \delta^{-1}) \frac{d}{\mu}$.

Potem gradinetni spust s hitrostjo učenja $\eta = \Theta(d_0 \frac{\delta}{m \cdot \text{poly}(n, l)})$ z verjetnostjo vsaj $1 - e^{-\Omega(\log^2 m)}$ po $O(\frac{n^6 L^2}{\delta^2 \mu}) \log \epsilon^{-1}$ korakih najde parametre ϕ , za katere velja $\mathcal{L}_{regresija}(\phi) < \epsilon$.

A.2 Preparametrizacija z grafovskimi nevronskimi mrežami in algoritmom Sinkhorn

Definicija grafovke konvolucijske mreže sledi [14]. Definicija Sinkhornovega algoritma je povzeta po [15]. Implementacija ujemanja grafov z grafovskimi nevronskimi mrežami je povzeta po [3].

Definicija A.4 (Grafovka konvolucijska mreža). Naj bo M matrika sosednosti grafa G in D diagonalna matrika s stopnjami vozlišč grafa G . Konvolucijsko plast grafovke nevronskne mreže definiramo kot

$$H^{(l+1)} = \sigma \left(D^{-\frac{1}{2}} (M + I) D^{-\frac{1}{2}} H^{(l)} W^{(l)} \right),$$

kjer je $H^{(l)}$ matrika aktivacij v l -ti plasti, $W^{(l)}$ je matrika uteži v l -ti plasti in σ je aktivacijska funkcija.

Definicija A.5 (Sinkhornov algoritem). Naj bo $K \in \mathbb{R}^{n \times n}$ matrika z pozitivnimi elementi. Sinkhornov algoritem je iterativni postopek za izračun dvojno stohastične matrike $P \in \mathbb{R}^{n \times n}$ oblike $P = D_1 K D_2$, kjer sta D_1 in D_2 diagonalni matriki s pozitivnimi elementi na diagonali. Algoritem izračuna matriki D_1 in D_2 z naslednjimi koraki:

$$\begin{aligned} D_1^{(t+1)} &= \text{diag} \left(\frac{1}{K D_2^{(t)} \mathbf{1}} \right) \\ D_2^{(t+1)} &= \text{diag} \left(\frac{1}{K^T D_1^{(t+1)} \mathbf{1}} \right) \end{aligned}$$

kjer je $\mathbf{1} \in \mathbb{R}^n$ vektor enic.

Če nas zanima le P , je dovolj, da izmenično vse vrstice in vse stolpce matrike P normaliziramo na 1. Z $S_m(K)$ označimo matriko, dobljeno z uporabo Sinkhornovega algoritma na K po m iteracijah.

Za matriko K s strogo pozitivnimi členi velja, da Sinkhornov algoritem konvergira do dvojno stohastične matrike P [15, izrek 2.1].

Definicija A.6 (Stroškovna matrika). Naj bosta $X \in \mathbb{R}^{n \times d}$ in $Y \in \mathbb{R}^{n \times d}$ poljubni matriki in naj X_i označuje i -ti vrstico matrike X , Y_j pa j -to vrstico matrike Y . Stroškovno matriko $C(X, Y) \in \mathbb{R}^{n \times n}$ definiramo kot

$$C_{i,j} = \|X_i - Y_j\|^2.$$

Za graf velikosti N definiramo model M_{encoder} kot grafovsko nevronske mrežo z dvema konvolucijskima plastema in aktivacijsko funkcijo ReLU s slednjim predpisom:

$$M_{\text{encoder}}(M) = \hat{M} \text{ReLU}(\hat{M}W_1)W_2 \in \mathbb{R}^{n \times n^2},$$

kjer je $\hat{M} = D^{-\frac{1}{2}}(M + I)D^{-\frac{1}{2}}$ in sta $W_1 \in \mathbb{R}^{n \times n^2}$, $W_2 \in \mathbb{R}^{n^2 \times n^2}$ parametra modela. .

Definicija A.7. V poglavju 4.5.3 gravske nevronske mreže na sledeč način uporabimo za reševanje ujemanja grafov:

1. Za grafa z matrikama sosednosti M_1, M_2 računamo

$$P = S_{100}(C(M_{\text{encoder}}(M_1), M_{\text{encoder}}(M_2))).$$

2. Minimiziramo funkcijo izgube

$$\|M_1 - PM_2P^T\|.$$

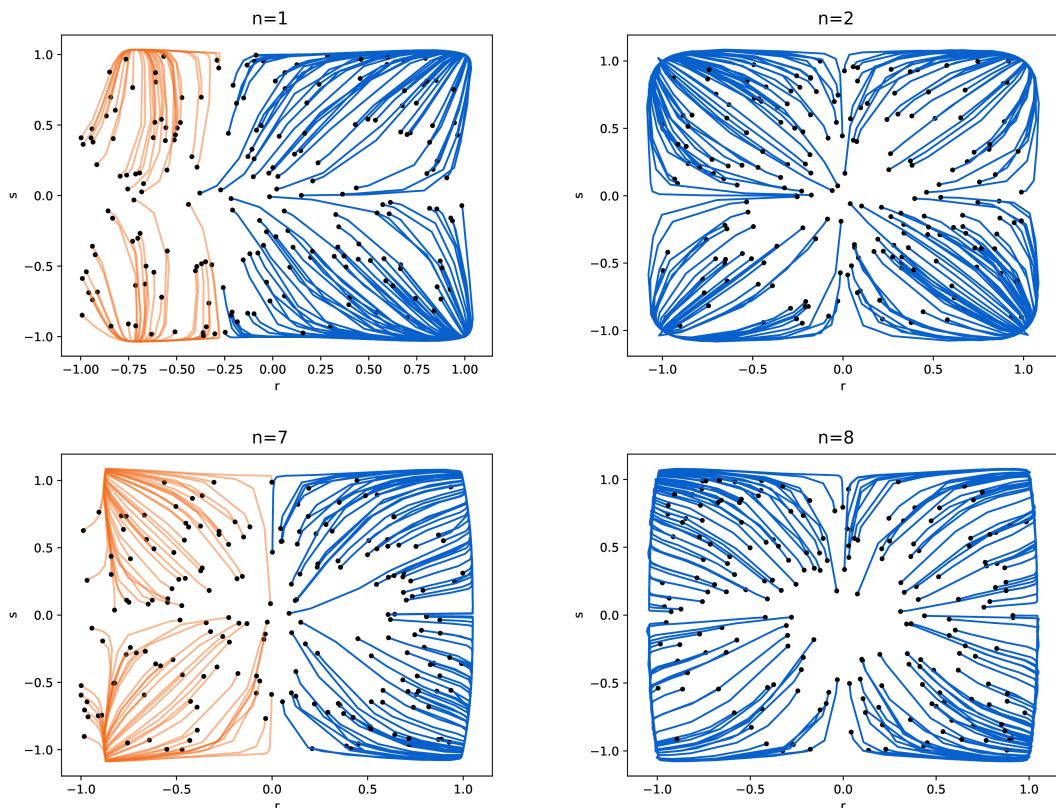
B Rezultati simulacij

B.1 Upodobitve

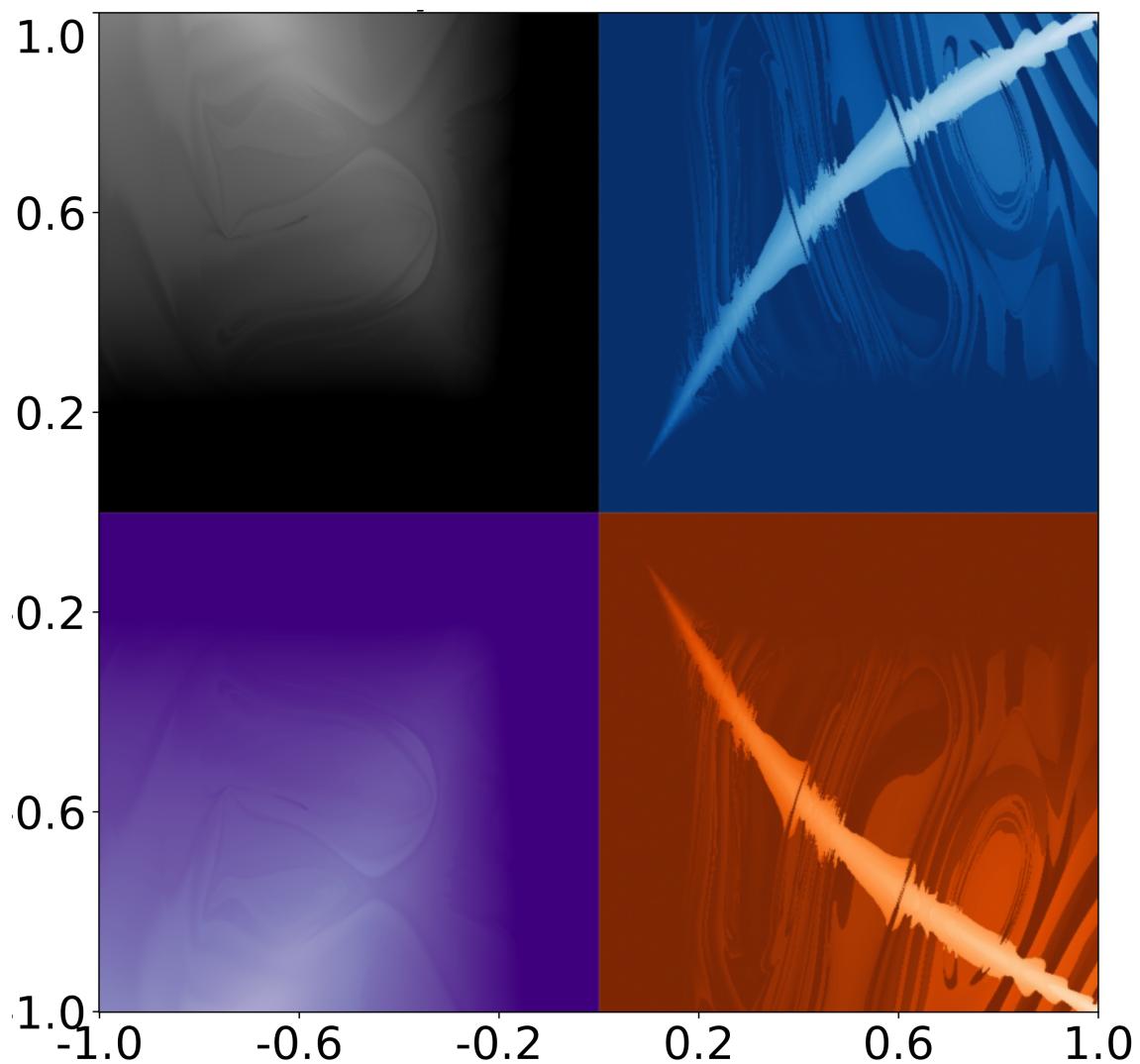
B.1.1 Diedrska grupa

Trajektorije Več trajektorij enodimensionalnega modela iz slike 4. Enačbo gradientnega toka rešujemo z metodo Dormand-Prince. Vsaka slika prikazuje 500 trajektorij z naključno izbranimi začetnimi parametri $r, s \in U[-1, 1]$.

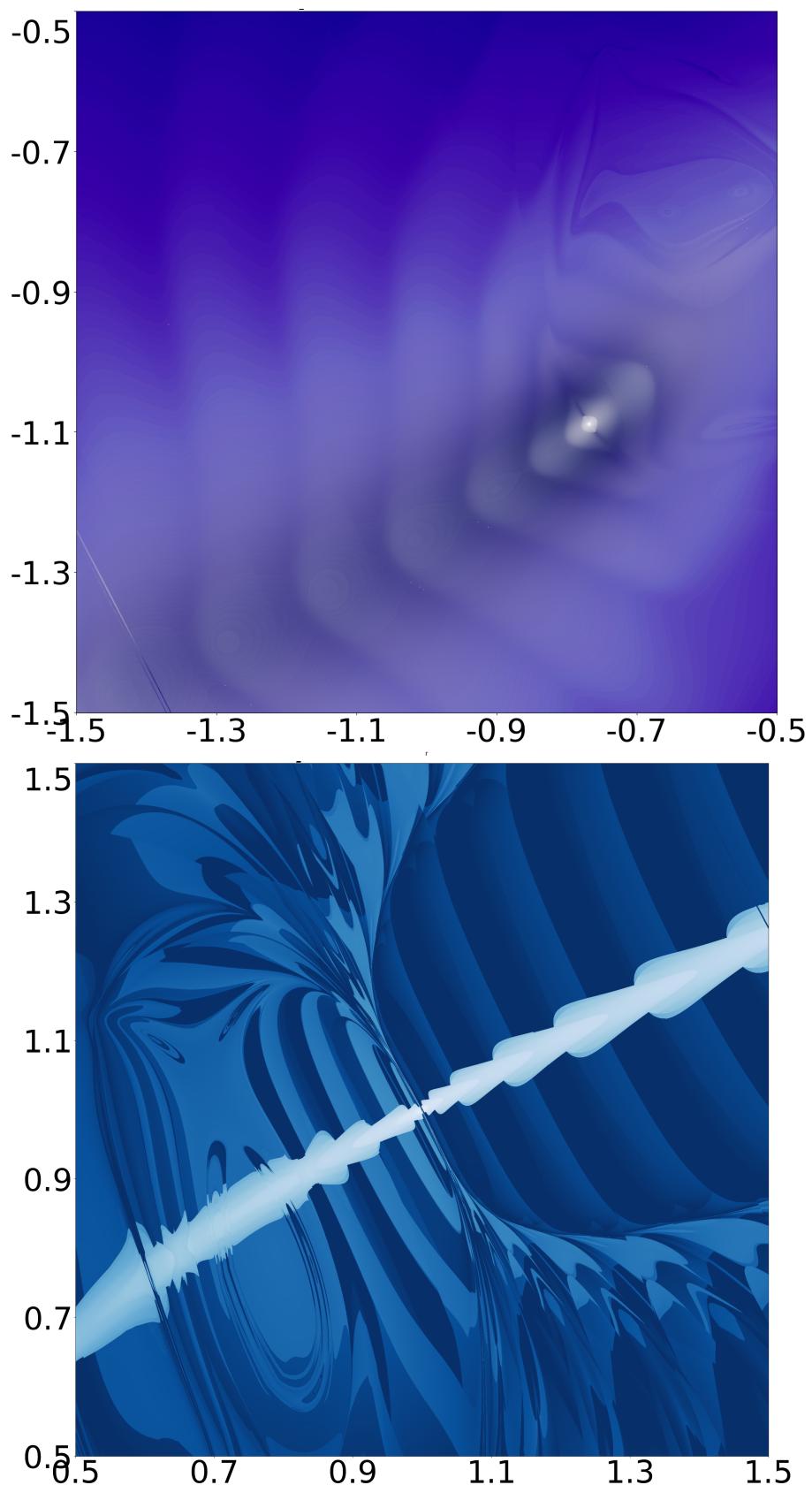
Dinamika enodimensionalnega modela Rešujemo diferencialno enačbo gradientnega toka za enodimensionalni model realne upodobitve nad diedrsko grupo in opazujemo čas t_{r_0, s_0} , v katerem model z začetnima parametromi $\rho(r)_0 = r_0, \rho(s)_0 = s_0$ prvič skonvergira do ϵ -okolice limite trajektorije. Slike 14, 15 in 16 so bile pridobljene z $\epsilon = 0.01$ in integracijskim intervalom $t \in [0, 1]$.



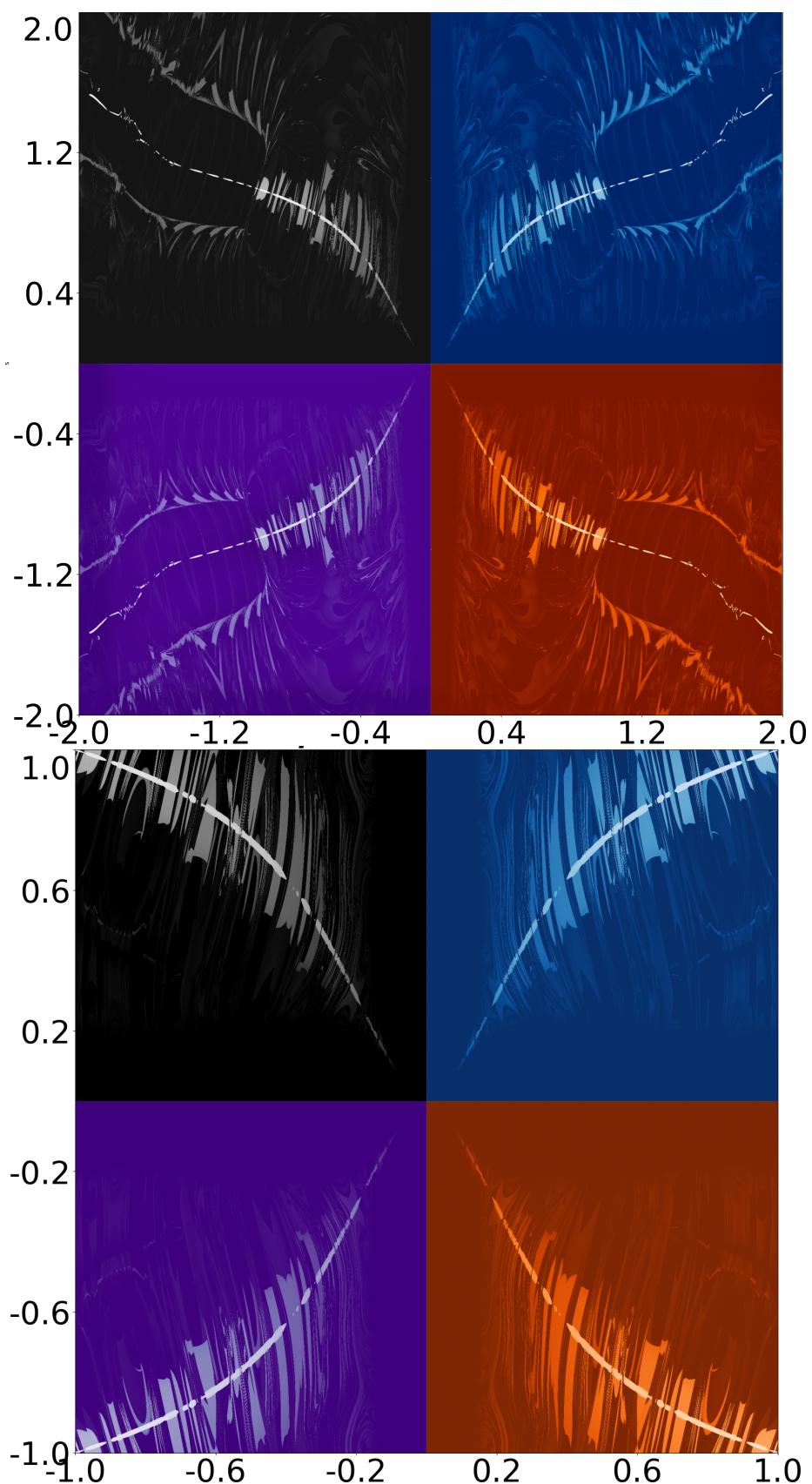
Slika 13: Trajektorije enodimenzionalnega modela realne upodobitve diedrske grupe za različne vrednosti n . Za vsako vrednost n je prikazanih 500 naključno izbranih trajektorij z začetnimi parametri iz $[-1, 1]^2$.



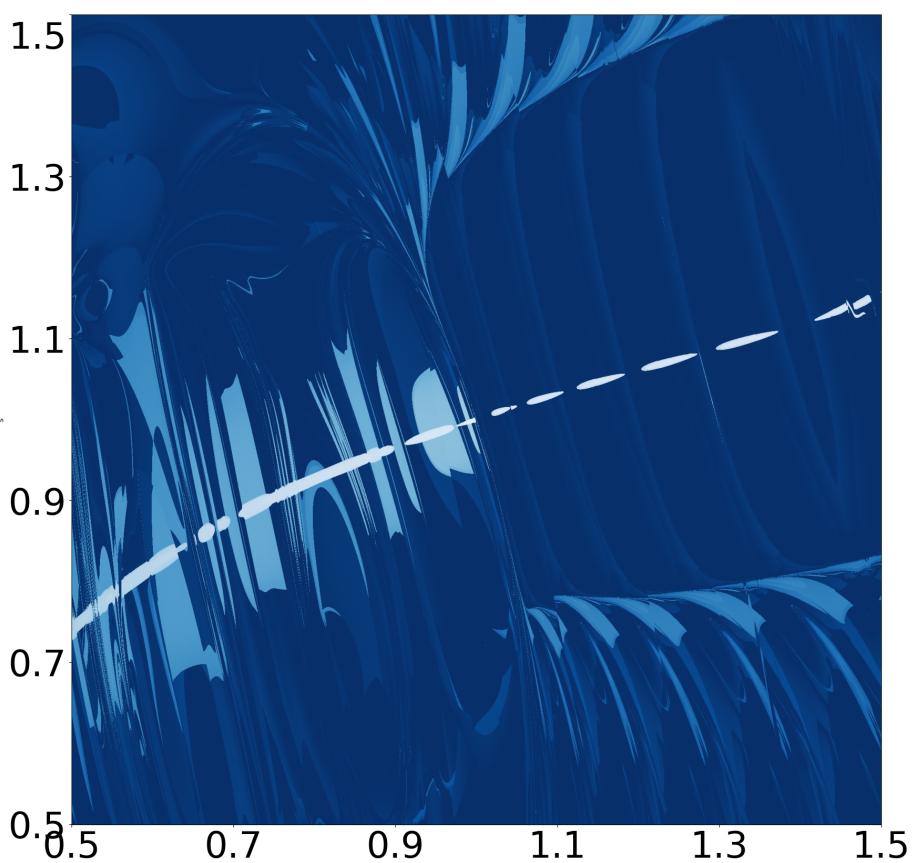
Slika 14: Hitrosti konvergencije enodimenzionalnega modela $D_3 \rightarrow \mathbb{R}$ z različnimi začetnimi parametri (r_0, s_0) . Na zgornji sliki so začetni parametri izbrani iz kvadrata $[-1, 1]^2$, na spodnji pa iz kvadrata $[0.5, 1.5]^2$.



Slika 15: Hitrosti konvergencije enodimenzionalnega modela $D_3 \rightarrow \mathbb{R}$ na $[-1.5, -0.5]^2$ in $[0.5, 1.5]^2$ v resoluciji 2000×2000 (zgoraj) in 3000×3000 (spodaj). Barvna shema je izbrana, da poudari kontraste med regijami z različnimi hitrostmi konvergenc.



Slika 16: Hitrosti konvergencije enodimenzionalnega modela $D_4 \rightarrow \mathbb{R}$ na $[-2, 2]^2$ in $[-1, 1]^2$. Resolucija slik je 2000×2000 .



Slika 17: Hitrosti konvergencije enodimenzionalnega modela $D_4 \rightarrow \mathbb{R}$ na $[0.5, 1.5]^2$. Resolucija slik je 2000×2000 .

Dvodimenzionalen model Opazujemo realen model $\rho: D_n \rightarrow \mathbb{R}^{2 \times 2}$ upodobitve diedrske grupe, podan z matrikama $R := \rho(r)$ in $S := \rho(s)$.

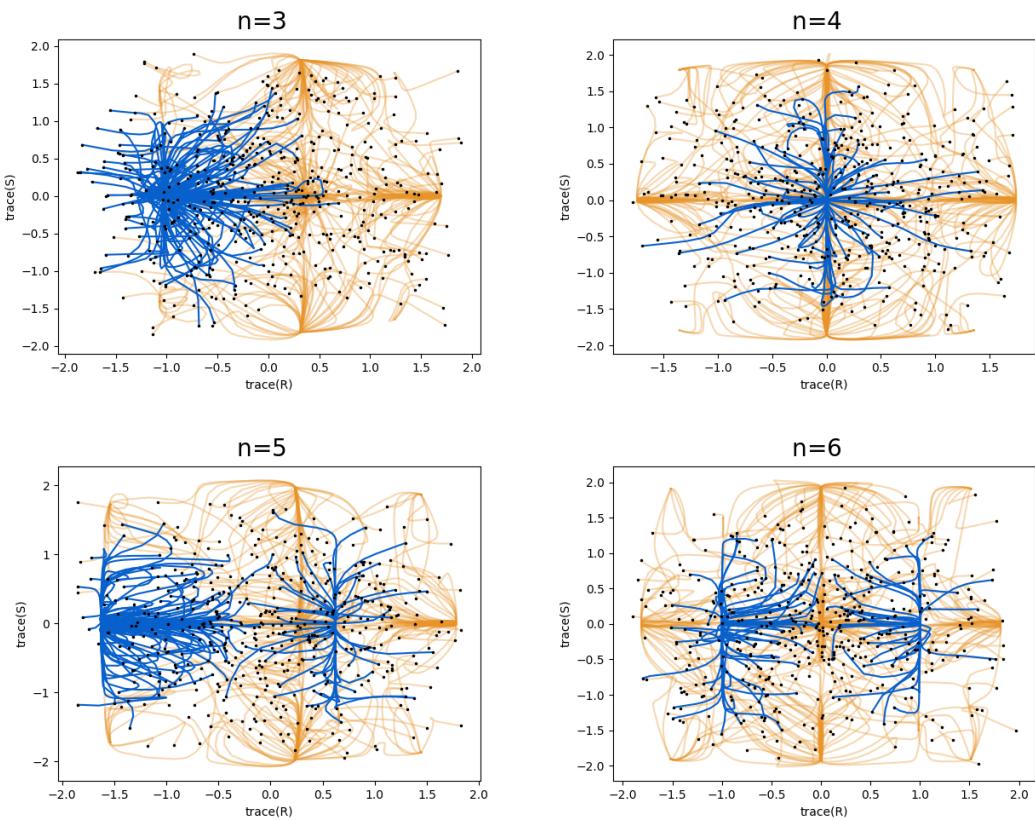
Za 500 naključno izbranih začetnih parametrov $(R_0, S_0) \sim U[(-1, 1)^8]$ rešujemo diferencialno enačbo

$$\frac{d(R, S)}{dt} = -\nabla \mathcal{L}(R, S)$$

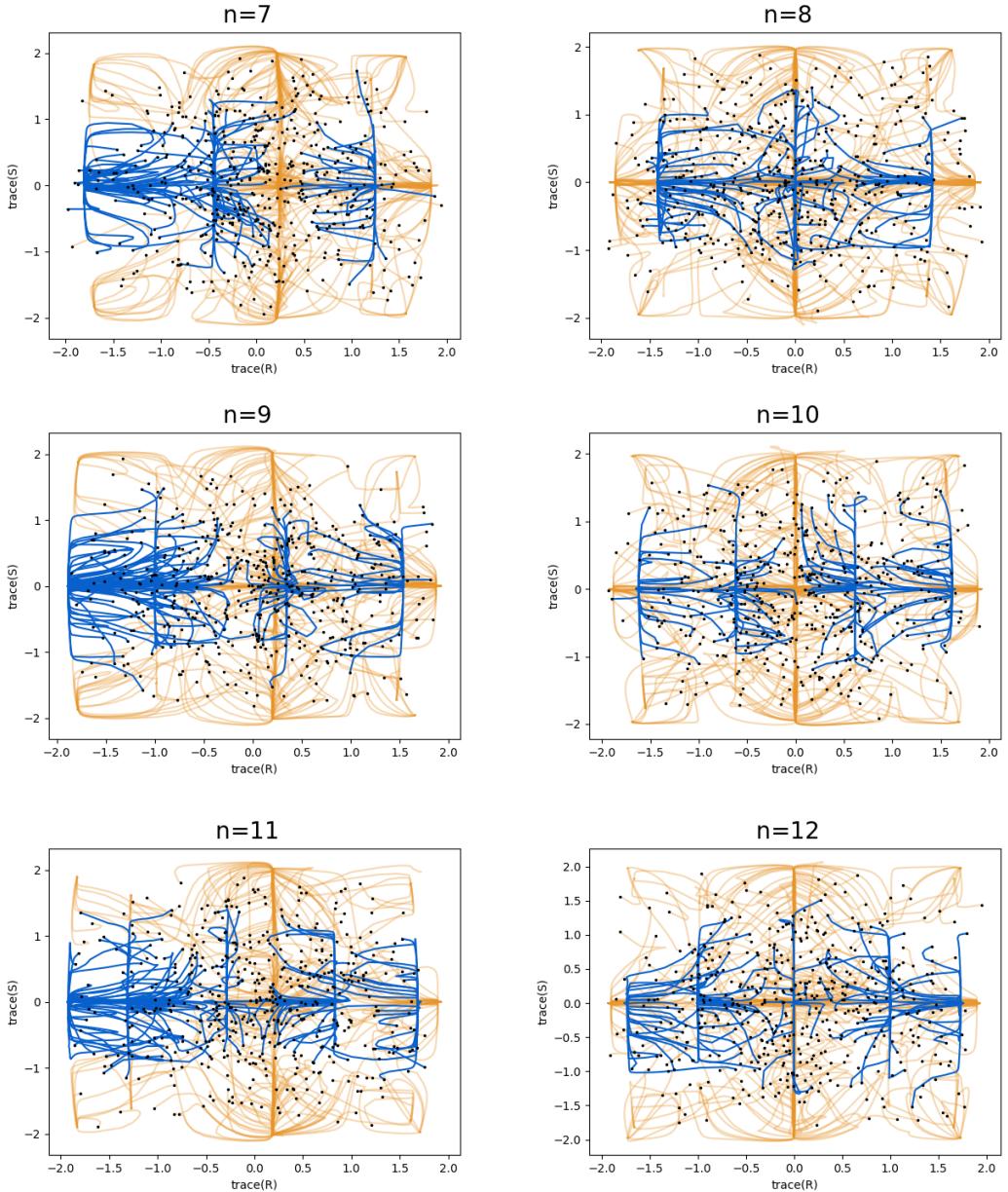
z začetnim pogojem $(R(0), S(0)) = (R_0, S_0)$.

Na sklikah 18 in 19 so prikazane trajektorije $\{(\text{tr}(R(t)), \text{tr}(S)(t)) \mid t \in [0, 1]\}$ za različne začetne parametre.

Enačbo rešujemo z metodo Dormand-Prince [9] (Runge-Kutta 5. reda s 4. redom ocene napake). Gradient $\nabla \mathcal{L}$ je izračunan analitično.



Slika 18: Trajektorije učenja modelov $D_n \rightarrow \mathbb{R}^{2 \times 2}$ z različnimi začetnimi parametri. Model $D_n \rightarrow \mathbb{R}^{2 \times 2}$ je definiran z vrednostima $R := \rho(r)$ in $S := \rho(s)$. Opazujemo trajektorije $(\text{tr}(R(t)), \text{tr}(S)(t))$. Črne pike predstavljajo začetne parametre. Modre trajektorije konvergirajo do upodobitve.

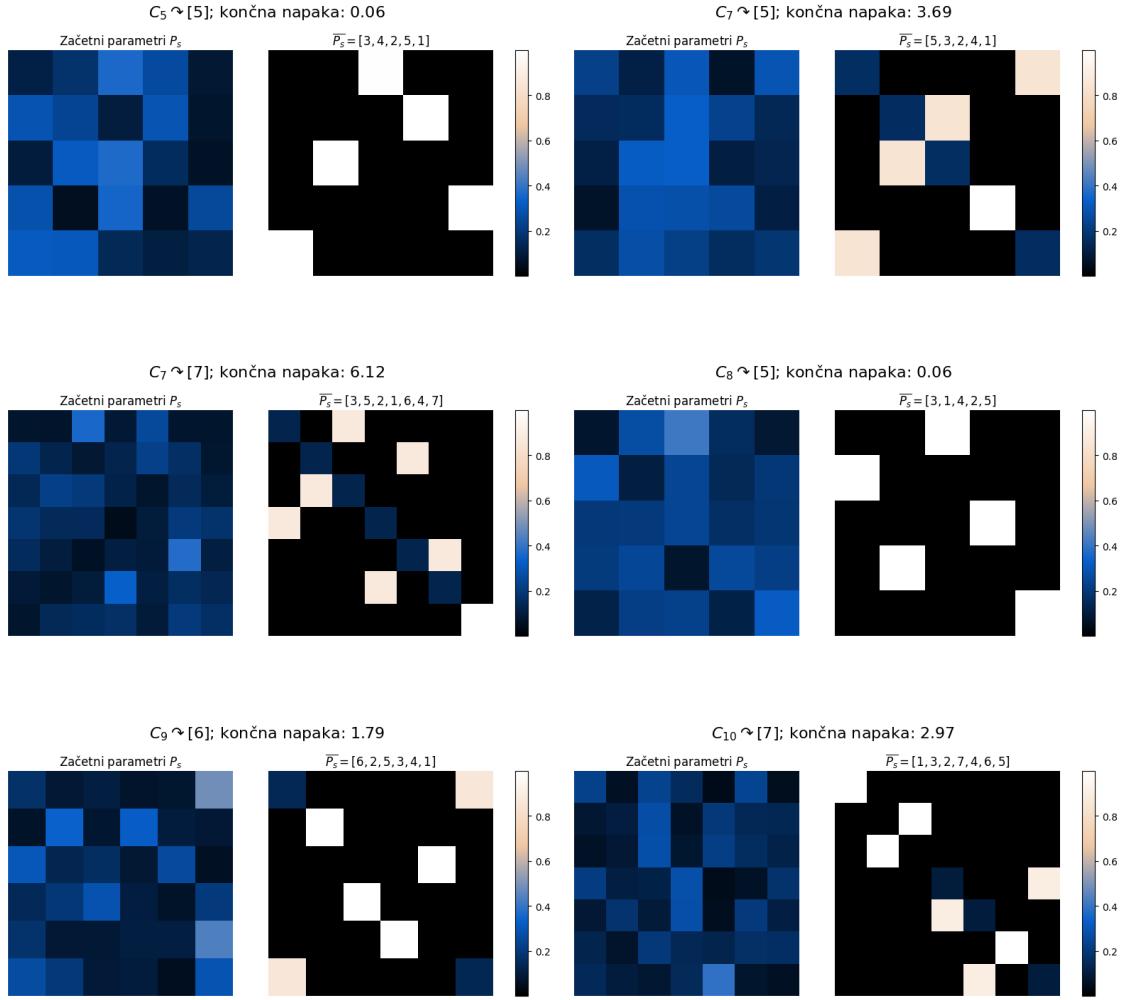


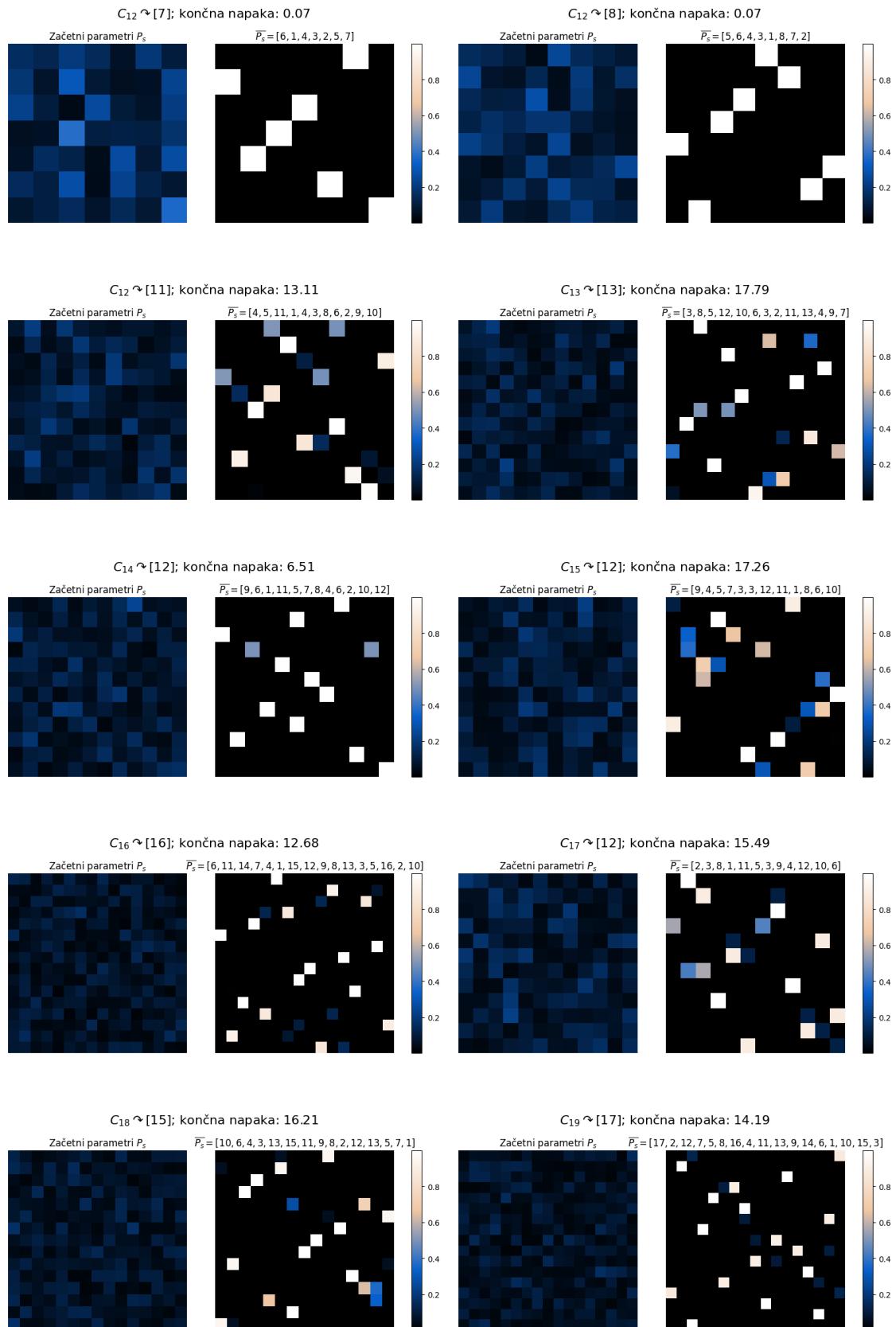
Slika 19: Trajektorije učenja (nadaljevanje).

B.2 Delovanja

B.2.1 Ciklične grupe

Opazujemo matrike P_s pred in po optimizaciji kompleksnih modelov $\rho: C_n \rightarrow \text{fun}([m], [m])$. Optimiziramo z algoritmom *adam* s hitrostjo učenja 0.01 in privzetimi parametri iz [20]. Vsak model učimo največ največ 2000 korakov. Če model doseže napako pod 10^{-3} , učenje ustavimo prej.

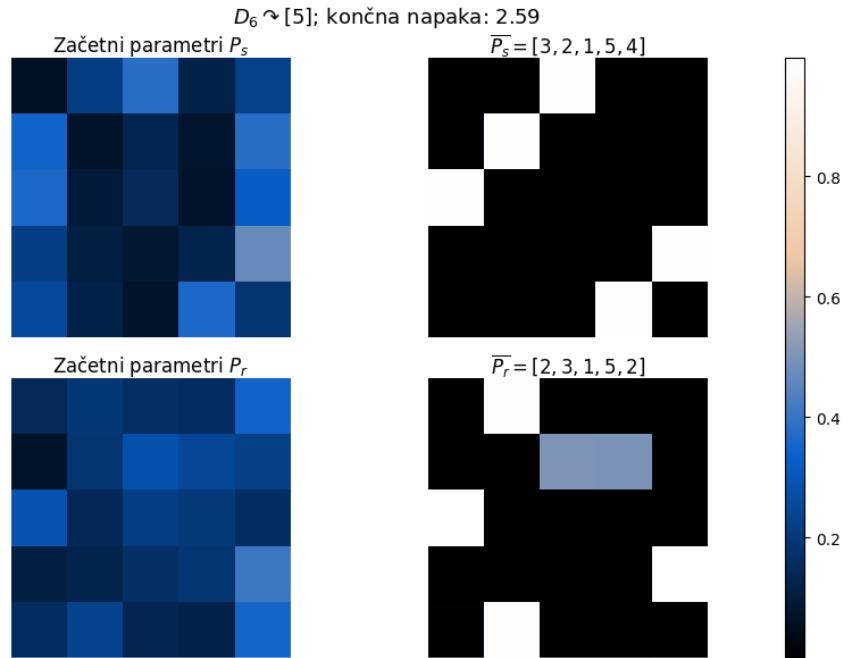




Slika 20: Matrike P_s pred in po optimizaciji modelov $\rho: C_n \rightarrow \text{fun}([m], [m])$ za različne vrednosti n in m .

B.2.2 Diedrske grupe

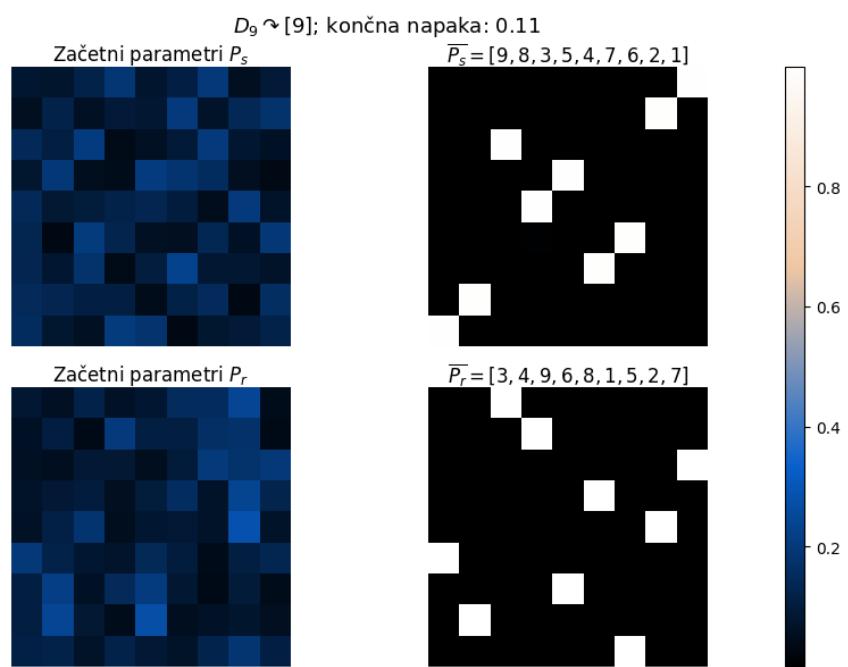
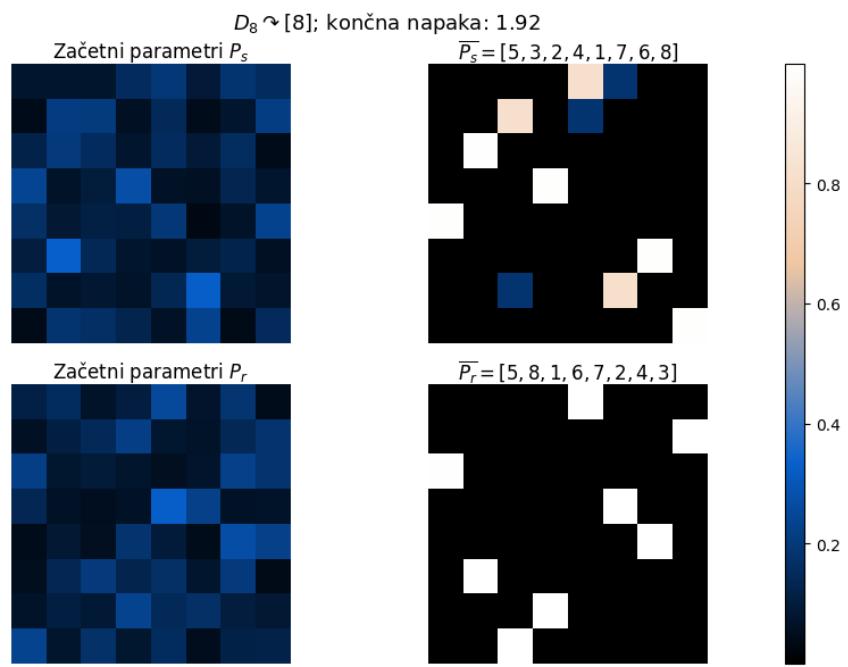
Opazujemo matrike P_s in P_r pred in po optimizaciji modelov $\rho: D_n \rightarrow \text{fun}([m], [m])$. Optimiziramo z algoritmom *adam* s hitrostjo učenja 0.01 in privzetimi parametri iz [20]. Vsak model učimo največ največ 2000 korakov. Če model doseže napako pod 10^{-3} , učenje ustavimo prej.

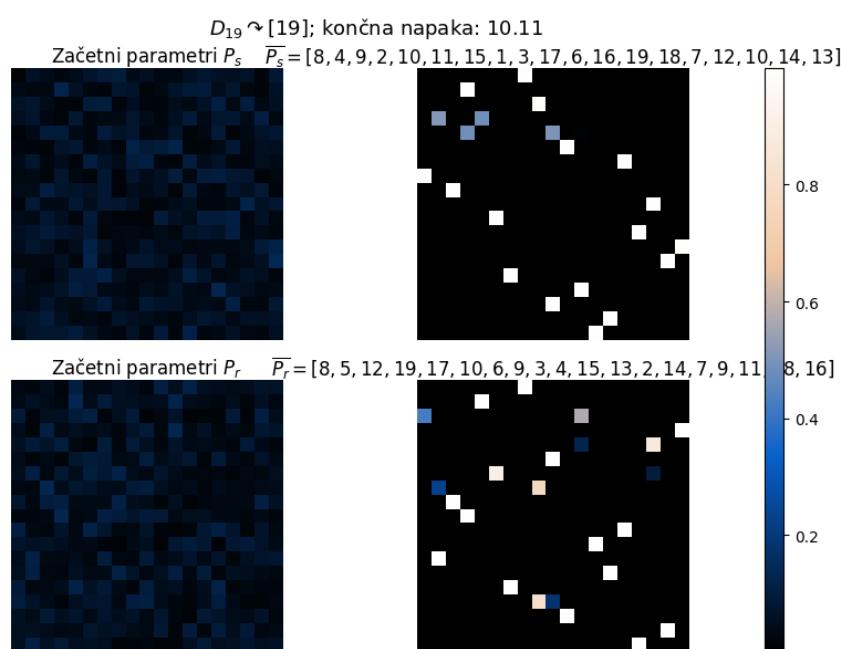
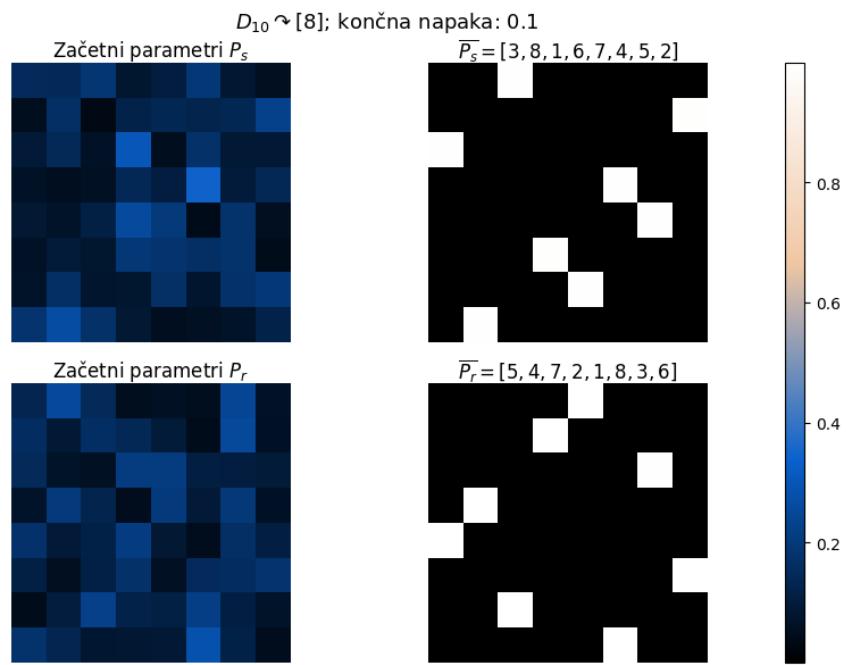


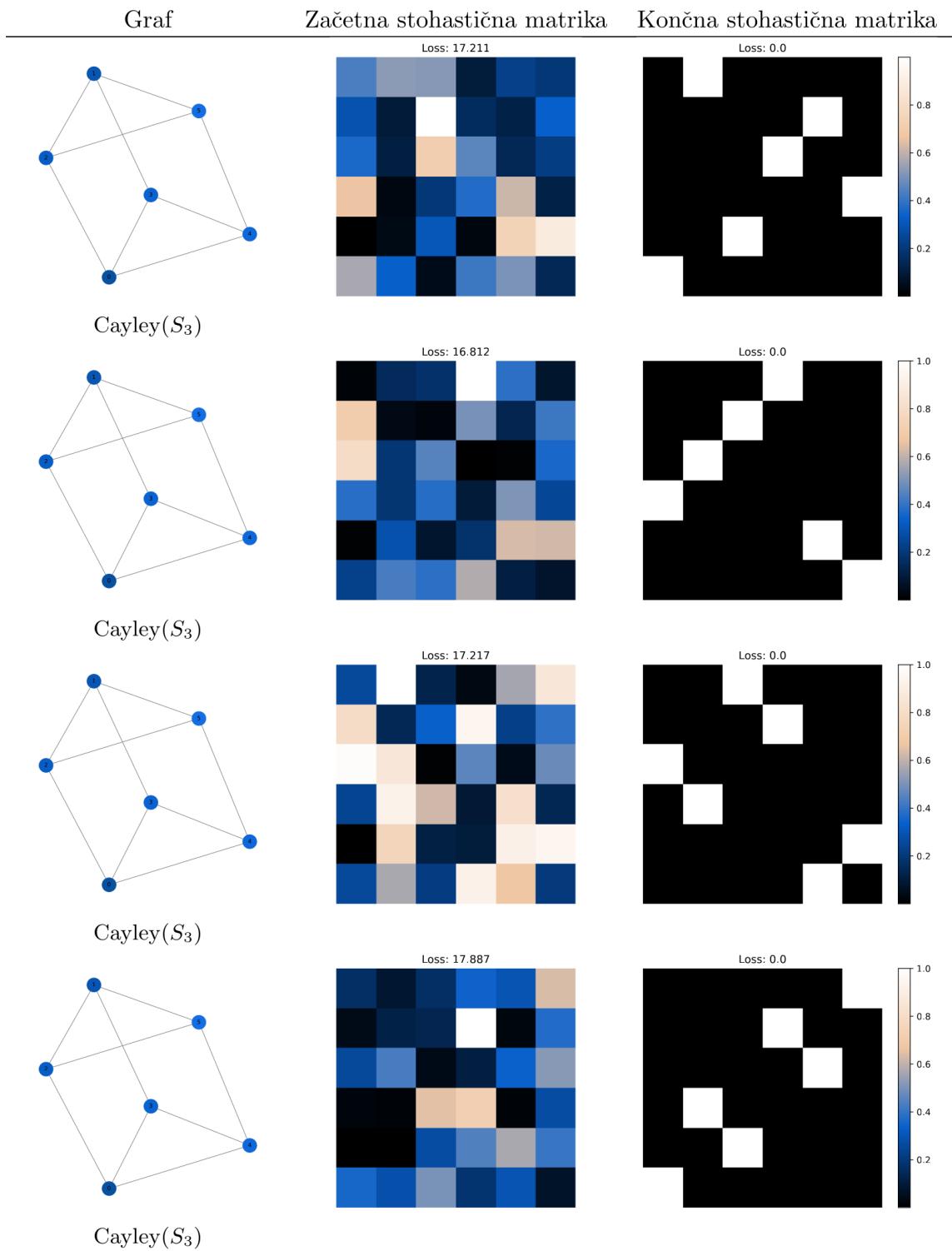
B.3 Grafi

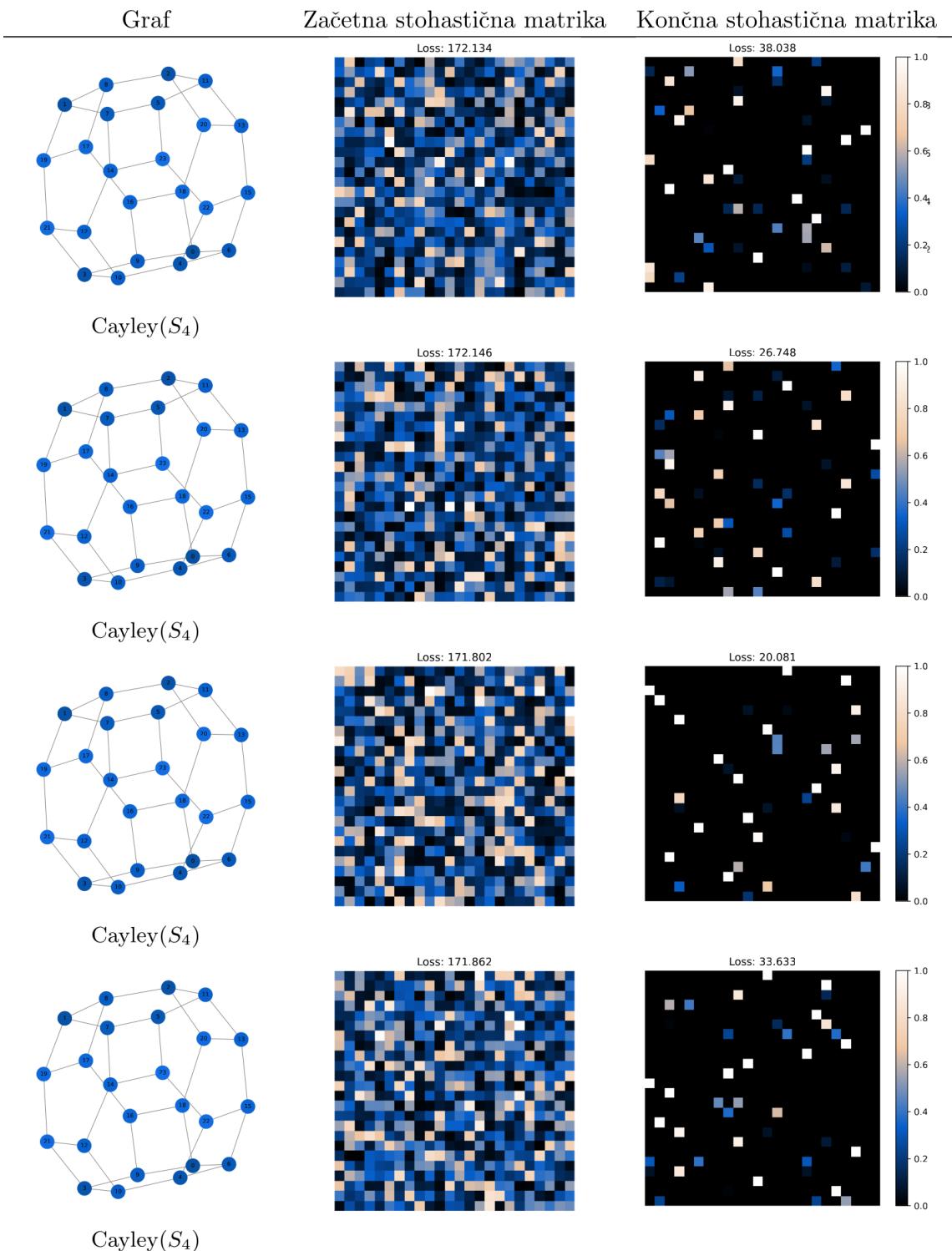
B.3.1 Avtomorfizmi cayleyevih grafov

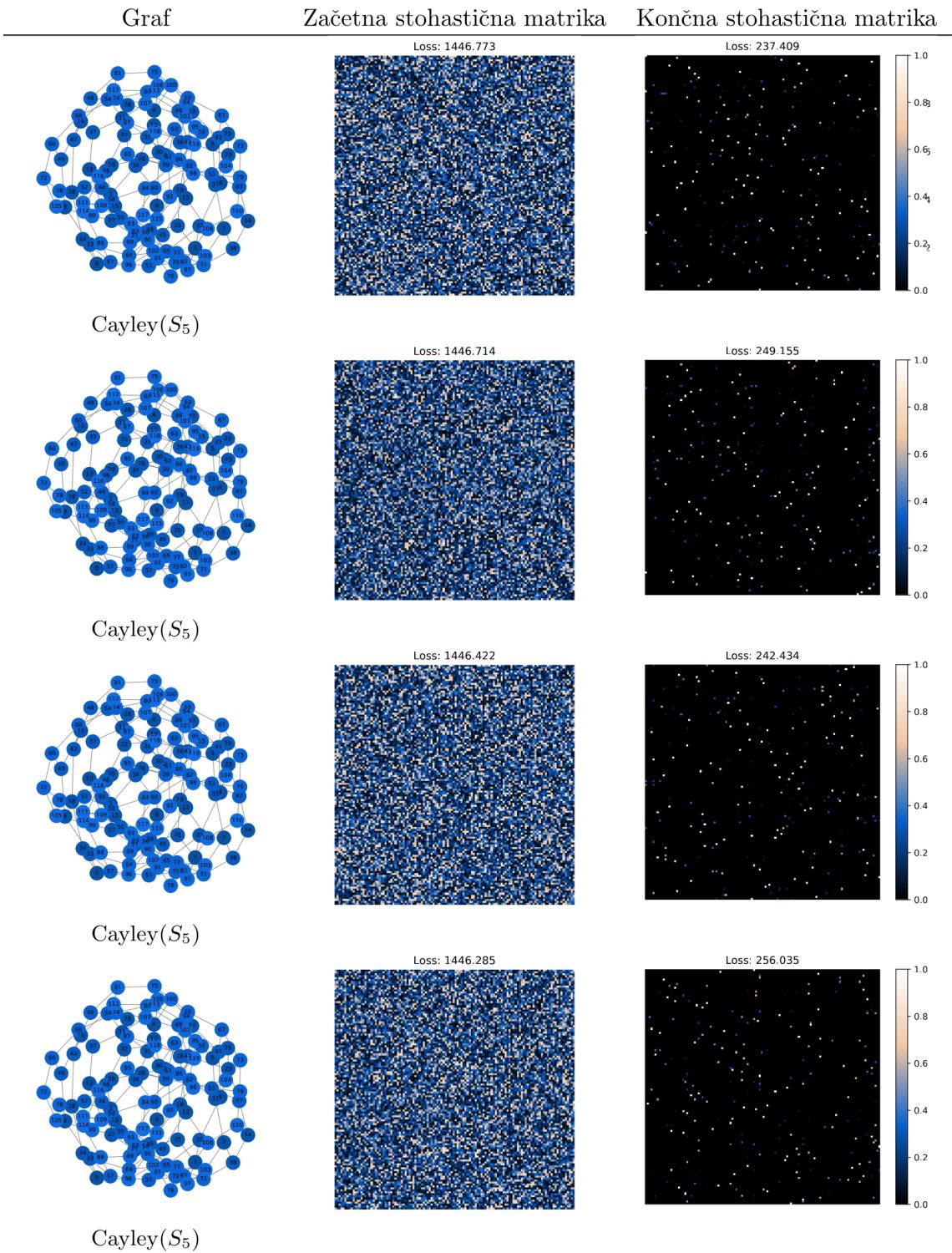
Rezultati iskanja avtomorfizmov Cayleyevih grafov brez uporabe tabele inverzij. Parametri optimizacije so opisani v poglavju 4.4.











Literatura

- [1] Y. Afalo, A. Bronstein in R. Kimmel, *On convex relaxation of graph isomorphism*, Proceedings of the National Academy of Sciences of the United States of America **112** (2015), DOI: 10.1073/pnas.1401651112.
- [2] Z. Allen-Zhu, Y. Li in Z. Song, *A Convergence Theory for Deep Learning via Over-Parameterization*, 2019, arXiv: 1811.03962 [cs.LG], dostopno na <https://arxiv.org/abs/1811.03962>.
- [3] L. Chen in dr. *Graph optimal transport for cross-domain alignment*, CoRR **abs/2006.14744** (2020), arXiv: 2006.14744, dostopno na <https://arxiv.org/abs/2006.14744>.
- [4] B. Chughtai, L. Chan in N. Nanda, *Neural networks learn representation theory: reverse engineering how networks perform group operations*, v: ICLR 2023 Workshop on Physics for Machine Learning, 2023, dostopno na https://openreview.net/forum?id=j4_YHiTAN63.
- [5] B. Dherin in dr. *Learning by solving differential equations*, 2025, arXiv: 2505.13397 [cs.LG], dostopno na <https://arxiv.org/abs/2505.13397>.
- [6] A. Diallo, Z. Markus in F. Johannes, *Permutation learning via lehmer codes*, v: 2020, DOI: 10.3233/FAIA200206.
- [7] J. Ding in N. H. Rhee, *Teaching tip: when a matrix and its inverse are stochastic*, The College Mathematics Journal **44**(2) (2013) 108–109, DOI: 10.4169/college.math.j.44.2.108, eprint: <https://doi.org/10.4169/college.math.j.44.2.108>, dostopno na <https://doi.org/10.4169/college.math.j.44.2.108>.
- [8] C. Gamm, *ε -representations of groups and Ulam stability*, doktorska disertacija, Master thesis, Georg-August-Universität Göttingen. 21, 2011.
- [9] E. Hairer, S. Norsett in G. Wanner, *Solving ordinary differential equations i: nonstiff problems*, 8, 1993, DOI: 10.1007/978-3-540-78862-1.
- [10] M. W. Hirsch, S. Smale in R. L. Devaney, *Differential equations, dynamical systems, and an introduction to chaos*, en, 3. izd., Academic Press, San Diego, CA, 2012.
- [11] U. Jezernik, *Teorija upodobitev*, <https://urbanjezernik.github.io/teorija-upodobitev/intro.html>, [Na spletu; dostopno 29. 8. 2025], 2025.
- [12] C. Jin in dr. *How to Escape Saddle Points Efficiently*, 2017, arXiv: 1703.00887 [cs.LG], dostopno na <https://arxiv.org/abs/1703.00887>.
- [13] M. Kastrati in M. Biba, *A state-of-the-art survey of advanced optimization methods in machine learning*, v: 2021.
- [14] T. N. Kipf in M. Welling, *Semi-supervised classification with graph convolutional networks*, CoRR **abs/1609.02907** (2016), arXiv: 1609.02907, dostopno na <http://arxiv.org/abs/1609.02907>.

- [15] P. Knight, *The sinkhornknopp algorithm: convergence and applications*, SIAM Journal on Matrix Analysis and Applications **30** (2008) 261–275, DOI: 10.1137/060659624.
- [16] Y. Li in dr. *Graph Matching Networks for Learning the Similarity of Graph Structured Objects*, 2019, arXiv: 1904.12787 [cs.LG], dostopno na <https://arxiv.org/abs/1904.12787>.
- [17] V. Lyzinski in dr. *Graph Matching: Relax at Your Own Risk*, 2015, arXiv: 1405.3133 [stat.ML], dostopno na <https://arxiv.org/abs/1405.3133>.
- [18] Y. Park, *Absence of Closed-Form Descriptions for Gradient Flow in Two-Layer Narrow Networks*, 2024, arXiv: 2408.08286 [cs.LG], dostopno na <https://arxiv.org/abs/2408.08286>.
- [19] R. Pascanu in dr. *On the saddle point problem for non-convex optimization*, 2014, arXiv: 1405.4604 [cs.LG], dostopno na <https://arxiv.org/abs/1405.4604>.
- [20] A. Paszke in dr. *Pytorch: an imperative style, high-performance deep learning library*, v: Advances in Neural Information Processing Systems, 2019.
- [21] S. J. Prince, *Understanding deep learning*, The MIT Press, 2023, dostopno na <http://udlbook.com>.
- [22] I. Safran, G. Yehudai in O. Shamir, *The Effects of Mild Over-parameterization on the Optimization Landscape of Shallow ReLU Neural Networks*, 2021, arXiv: 2006.01005 [cs.LG], dostopno na <https://arxiv.org/abs/2006.01005>.
- [23] S. Sahni in T. Gonzalez, *P-complete approximation problems*, Journal of the ACM **23**(3) (1976) 555–565, DOI: 10.1145/321958.321975.
- [24] D. Severo, B. Karrer in N. Nolte, *Learning Distributions over Permutations and Rankings with Factorized Representations*, 2025, arXiv: 2505.24664 [cs.LG], dostopno na <https://arxiv.org/abs/2505.24664>.
- [25] J. B. Simon in dr. *More is Better in Modern Machine Learning: when Infinite Overparameterization is Optimal and Overfitting is Obligatory*, 2024, arXiv: 2311.14646 [cs.LG], dostopno na <https://arxiv.org/abs/2311.14646>.
- [26] R. P. Stanley, *Enumerative combinatorics: volume 1*, 2nd, Cambridge University Press, USA, 2011.
- [27] P. Virtanen in dr. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nature Methods **17** (2020) 261–272, DOI: 10.1038/s41592-019-0686-2.

Stvarno kazalo

adam, 3

funkcija izgube nad relacijami, 7

(delovanje), 20

funkcija izgube za bijektivnost (grafi), 27

funkcija izgube za nerazcepnost, 8

funkcija izgube za povezanost (grafi), 29

funkcija izgube za unitarnost, 8

gradientni spust, 2

gradientni tok, 2

model delovanja, 19

model upodobitve, 6

nerazcepna upodobitev, 5

norma karakterja upodobitve, 5

tabela inverzij, 32

ujemanje grafov, 29

upodobitev, 5