VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
MATEMATIKA
MAGISTRO STUDIJŲ PROGRAMA

Automatizuotas Retas Modelių Identifikavimas ir Parametrų Optimizavimas Reakcijos-Difuzijos Sistemoms

Automated Sparse Model Identification and Parameter Optimization for Reaction-Diffusion Systems

Baigiamasis magistro darbas

Atliko: Kasparas Valatkevičius
VU el. p.: kasparas.valatkevicius@mif.stud.vu.lt
Vadovas: Prof. Dr. Olga Štikonienė

Vilnius

2025

Turinys

Santrauka

Šiame darbe pateikiama dalinai automatizuotą, modulinę darbo eigą reakcijos–difuzijos sistemų valdančiųjų dalinių diferencialinių lygčių atradimui iš duomenų. Eiga jungia *Reto Netiesinės Dinamikos Identifikavimą* (SINDy), sukonstruotus išsamius hiperparametrų šlavimus ir dviejų etapų *dėmenų stabilumu pagrįstą genėjimą*, kuris pašalina retai pasirenkamus arba nuolat perteklinį priderinimą sukeliančius požymius prieš Pareto optimalią modelio atranką. Metodika išbandoma keturiuose kanoniniuose vienmačiuose DLDI (šilumos, Fišerio–KPP, Aleno–Kano ir Grėjaus–Skoto lygtyse), naudojant sintetinius duomenis tinkleliuose nuo 16×16 iki 256×256 taškų.

Be triukšmo visos lygtys atgautos su žemais paklaidų įverčiais ir mažomis koeficientų paklaidomis < 5 %. Adityvaus Gauso triukšmo bandymai atskleidė išvestinėmis grįstų metodų silpnąją vietą: identifikavimas ženkliai blogėja esant vos $\sim 1\%$ triukšmui difuzijos ir $\sim 0.1\%$ – reakcijos-difuzijos sistemose. Siūlomas kandidatinių funkcijų rinkinio genėjimas bibliotekoje sumažina dinamikai aprašyti naudojamų dėmenų skaičių, išlaikydamas visus tikruosius dėmenis, sumažina likučio paklaidas ir leidžia teisingai identifikuoti Šilumos, Fišerio-KPP, Aleno-Kano lygtis, Grėjaus–Skoto sistemą. Visas kodas pateikiamas kaip atvirojo kodo sprendimas, palaikantis kontroliuojamus skaitinės simuliacijos parametrus, optimizatorius ir jų hiperparametrus bei kandidatinių funkcijų rinkinius..

Tyrimas apibrėžia, kada klasikinis įprastosios formos SINDy metodas patikimas, bei kiekybiškai nustato jo triukšmo ribas, ir pateikia praktines rekomendacijas kandidatinio rinkinio sudarymui bei hiperparametrų derinimui. Tolimesni darbai turėtų integruoti silpnos formos ir Bajeso retumo metodus bei tirti automatinį rinkinio konstravimą, siekiant taikyti metodą triukšmingiems, didelės dimensijos eksperimentiniams duomenims.

1 Įvadas

1.1 Motyvacija

Pastaruoju metu vis dažniau domimasi duomenimis grindžiamu dinaminių sistemų modelių atradimu, kai dinamiką nusakančios lygtys atrandamos tiesiai iš eksperimentinių ar simuliacinių duomenų, o ne išvedamos tradiciniu analizės būdu. Ypač sudėtinga iš erdvės-laiko matavimų atrasti diferencialines lygtis dalinėmis išvestinėmis (DLDI): ieškant tikslių lygčių išraiškų reikia ieškoti kombinatoriškai galimų dėmenų rinkiniuose (polinominiai, erdvinių išvestinių, sandaugų dėmenys ir pan.) bei tvarkytis su potencialiai triukšmingais ar ribotais eksperimentiniais duomenimis. Retosios regresijos metodai dinamikos identifikavimui tapo ypač priimtinu sprendimu: darant prielaidą, kad tikroji DLDI yra *reta* didelėje funkcijų erdvėje, įmanoma atkurti dinamiką nusakančius modelius, kurie derina būtent išraiškos paprastumą ir retumą su dinamikos tikslumu.

Vienas iš tokių retosios regresijos metodų - Brunton ir bendraautorių pristatyta Reto Netiesinės Dinamikos Identifikavimo (angl. Sparse Identification of Nonlinear Dynamics) (SINDy) sistema, kuri modelio atradimą formuluoja kaip retos regresijos uždavinį ir iteratyviai naikinant nereikšmingus dėmenis, padeda atrasti glaudžius tyrinėjamą dinamiką aprašančius dėsnius [1]. Netrukus po to Rudy ir kt. šias idėjas pritaikė DLDI atveju, pasiūlę *PDE–FIND* algoritmą, kuris konstruoja erdvinių išvestinių ir netiesinių funkcijų rinkinį ir taiko retąją regresiją valdančioms DLDI atkurti, naudodamas Pareto analizę, kad subalansuotų modelio sudėtingumą ir simuliacinę paklaidą [2]. Mangan ir kt. dar labiau išplėtė SINDy iki neišreikštinio uždavinio formulavimo, leidžiančio atrasti racionaliąsias funkcijas turinčią dinamiką biologiniuose tinkluose [3]. Bendrai, šie darbai rodo, jog retumo prielaida leidžia gauti gerai interpretuojamus ir tikslią dinamiką prognozuojančius modelius tiek paprastųjų diferencialinių lygčių (PDL), tiek DLDI lygčių atvejais tuo pačiu atskleisdami esminius praktinius iššūkius tokio tipo uždaviniuose: tinkamo kandidatinių funkcijų rinkinio parinkimą, triukšmo valdymą išvestinių skaičiavimuose bei tikrinimą, ar atrastas modelis apibendrina dinamiką už naudojamų duomenų ribų.

1.2 Uždavinys ir Tikslai

Šiame susitelkiama į klasikinių reakcijos–difuzijos sistemų, tokių kaip šilumos, Fišerio–Kolmogorovo, Grėjaus–Skoto ir Aleno–Kano lygtys, analizei. Pasitelkę skaitinių skaičiavimų duomenų bazę, taikysime retųjų regresijų metodus šių lygčių formų identifikavimui ir, naudodami Pareto analizę, lyginsime modelius, subalansuodami likučio paklaidą ir lygties retumą bei kitus tikslumo matus. Pagrindinis tikslas – nustatyti optimalius parametrų intervalus ir geriausiai bendruosius reakcijos–difuzijos procesus atspindinčias lygtis, tuo pačiu pateikiant rekomenduojamą darbo eigą tolimesnėms praktinėms užduotims šioje srityje.

Nepaisant pažangos dinaminių lygčių identifikavimo srity, šiuo metu dar nėra vieningos, išplėstinės programinės sistemos, kuri – nuo DLDI skaitinio sprendinio iki retosios identifikacijos ir Pareto analize pagrįsto įverčio – leistų atlikti atkartojamus, masinius eksperimentus įvairioms lygtims su įvairiais retos regresijos optimizatoriais, jų hiperparametrais ir kandidatinių funkcijų rinkiniais. Didžioji dalis darbų remiasi atskirai vienai lygčiai ar metodui pritaikytais kodais, todėl dirbant su retosios regresijos sistemomis dažnai sunku atkartoti gautus rezultatus ar tiesiog optimaliu laike būdu atlikti išplėstinius dinaminių modelių tyrimus.

Šio darbo tikslas – sukurti ir pademonstruoti tokią modulinę retųjų DLDI identifikavimo sistemą, siekiant šių uždavinių:

1. **Reakcijos–difuzijos sistemų analizė.** Analizuoti šilumos, Fišerio–Kolmogorovo, Grėjaus–Skoto ir Aleno–Kano modelius, pritaikant retosios regresijos identifikavimą ir Pareto analizę; parengti rekomenduojamą darbo eigą optimaliems parametrų rėžiams, šlavimams ir kandidatinių funkcijų rinkinių konfigūracijoms nustatyti bei išbandyti ją baziniais bei triukšmingų duomenų atvejais.

2. **Sistemos kūrimas.** Sukurti lanksčią struktūrą, kuri automatizuoja DLDI sistemų skaitinę simuliaciją ir integruoja SINDy pagrįstą retąją regresiją, leidžiančią atlikti parametrų šlavimus (angl. parameter sweep) per optimizatorius, slenkstines jų vertes, hiperparametrus ir kintamas kandidatinių funkcijų rinkinių konfigūracijas.

3. **Optimizatorių ir rinkinių vertinimas.** Kiekybiškai palyginti kelis retosios regresijos optimizavimo metodus (pvz. STLSQ, SR3) ir funkcijų rinkinius (polinominės, racionalieji, sandaugų dėmenys) naudojant tokius matavimus kaip likučio paklaida, lygties retumas, terminų parinkimo tikslumas/atkūrimas/F1, koeficientų paklaida.

1.3 Darbo aprėptis

Šiame darbe nagrinėjami *deterministiniai*, sintetiniai duomenys, sugeneruoti iš gerai žinomų reakcijos–difuzijos modelių (pvz. Grėjaus–Skoto) taikant baigtinių skirtumų schemas. Apribojus tyrimą reprezentatyvia sistemų klase, galima atlikti išsamų ir kiekybiškai tikslų retųjų identifikavimo metodų vertinimą. Pagrindiniai indėliai yra šie:

- *Eksperimentinis tyrimas* su keliomis reakcijos–difuzijos sistemomis: sistemingai nagrinėjamas optimizatorių našumas, kandidatinių funkcijų rinkinių dizainas.

- *Integruota sistema* (pipeline) nuo pradžios iki pabaigos retajam DLDI atradimui, užtikrinanti pakartojamas darbo eigas ir lengvai plečiama naujais optimizatoriais ar kandidatinių funkcijų rinkiniais.

- *Skirtingų metrikų ir Pareto analizės pagrindu paremta modelio parinkimo* strategija, leidžianti griežtai lyginti atrastus modelius.

1.4 Darbo Struktūra

- **2 skyrius: Teorinės prielaidos ir literatūros apžvalga.** Apžvelgiami retieji regresijos metodai PDL/DLDI identifikavimui ir pristatomi reakcijos–difuzijos modeliai, naudojami šiame darbe.

- **3 skyrius: Metodologija.** Aprašoma skaitinio modeliavimo aplinka, kandidatinių funkcijų rinkinio konstravimas, retosios regresijos technikos ir vertinimo metrikos.
- **4 skyrius: Eksperimentiniai rezultatai.** Pateikiami identifikavimo rezultatai idealiomis ir triukšmingomis sąlygomis bei lyginamosios optimizatorių parametrų ir funkcijų rinkinių analizės.
- **5 skyrius: Diskusija ir validacija.** Aptariami rezultatai, identifikuotų modelių apribojimai ir rezultatų reikšmė platesniame mokslo kontekste.
- **6 skyrius: Išvados ir būsimi darbai.** Apibendrinamas indėlis, aptariamas darbo poveikis ir numatomos tolesnių potencialių tyrimų kryptys.

2 Teorinės prielaidos ir literatūros apžvalga

2.1 Duomenimis grindžiamas dinaminių sistemų atradimas

Duomenimis grindžiamas atradimas siekia valdančiąsias lygtis išvesti tiesiai iš matavimų, o ne iš fizikinio samprotavimo ar teorinės analizės. Ankstyvieji darbai naudojo simbolinę regresiją – ypač Šmito ir Lipsono genetinį programavimo metodą, kuris išvardija kandidatines išraiškas ir derina tikslumą su lygties sudėtingumu, kad iš duomenų išryškintų kompaktiškas diferencialines lygtis [5]. Nors tokia paieška gali būti galinga, visiškai simbolinės metodikos skaičiavimams imlios ir linkusios į perteklinį prisitaikymą prie duomenų, todėl kilo natūralus poreikis labiau struktūrizuotiems modeliams. Retoji regresija šį poreikį tenkina: sukuriamas didelis, bet fiksuotas kandidatinių funkcijų rinkinys ir daroma prielaida, kad tikroji dinamika naudoja tik keletą dėmenų iš kandidatinių funkcijų rinkinio; esant šiai retumo hipotezei, sprendžiama linijinė regresija su ℓ_1 apribojimu arba taikomas iteracinis slenkstinis dėmenų panaikinimas, kad nereikšmingi dėmenys būtų pašalinti [1].

Pagrindinis *Sparse Identification of Nonlinear Dynamics* (SINDy) algoritmas šią idėją pritaiko paprastosioms diferencialinėms lygtys: turint laiko eilučių duomenis suformuojama matrica, kurios stulpeliai – rinkinio funkcijų įverčiai sprendinio duomenims, o tuomet sprendžiama retumą skatinanti regresija $\Theta(X)\Xi \approx \dot{X}$, iš kurios koeficientų matricos nenulinės vertės nurodo aktyvius dėmenis [1]. Kadangi po iteratyvaus slenkstinio dėmenų šalinimo išlieka tik keli koeficientai, gautas modelis yra ir taupus, ir lengvai interpretuojamas. Parodyta, kad SINDy gali iš naujo atkurti, be kita ko, Lorenco, plėšrūno–grobio ir sūkurių nuslinkimo dinamikas net ir esant vidutiniam triukšmui, jei biblioteka pakankamai tiksli bei apima tikruosius dėmenis.

Šis požiūris greitai buvo apibendrintas: Šeiferis parodė, jog tas pats retumo principas leidžia identifikuoti kanonines dalines diferencialines lygtis, į biblioteką įtraukiant erdvines išvestines ir taikant ℓ_1-reguliarizuotą optimizavimą [4]; Rudy ir kt. pasiūlė PDE-FIND metodą, kuris sparčiąją regresiją susieja su Pareto analize ir aptinka įvairias DLDI – nuo difuzijos iki Navjė–Stokso lygčių [2]; o Manganas ir kt. išplėtė SINDy į neišreikštinį formulavimą, galintį atkurti racionaliuosius netiesiškumus, dažnai pasitaikančius biocheminėje kinetikoje [3]. Tolesnės patobulintos versijos, tokios kaip silpnosios integralinės formos SINDy, mažina triukšmo jautrumą retumą taikydamos integralinėms dinamikos išraiškoms [6], o ansamblinės ar valdymo variacijos išplečia taikomumą sistemoms su neapibrėžtais duomenimis. Šie pasiekimai patvirtina, kad retumo taikymas apdairiai pasirinktoje funkcijų erdvėje suteikia skaičiavimo požiūriu efektyvų ir patikimą kelią atrasti glaustus, fiziškai prasmingus modelius iš sudėtingų duomenų.

2.2 Retasis DLDI identifikavimas (PDE–FIND)

Rudy ir bendraautoriai pasiūlė *PDE–FIND* kaip tiesioginę SINDy metodo plėtrą dalinėms diferencialinėms lygtims [2]. Turint erdvės–laiko matavimus $u(x,t)$ diskretizuotame tinklely, laiko išvestinių ėminiai sudedami į stulpelinį vektorių $\mathbf{U}_t$, suformuojama *rinkinio matrica* $\Theta(U)$, kurios stulpelius sudaro kandidatinių dėmenų rinkinys – netiesinės u kombinacijos, erdvinės išvestinės u_x, u_{xx}, $u_{xxx}, \ldots$.

Sprendžiant retąją regresiją

$$\mathbf{U}_t \;\approx\; \Theta(U)\,\boldsymbol{\xi},$$

taikant nuoseklų slenksčiuotą ridge metodą, gaunamas koeficientų vektorius $\boldsymbol{\xi}$, kuriame išlieka tik kelios nenulinės vertės; būtent tie nenuliniai koeficientai apibrėžia identifikuotos DLDI dešiniąją pusę.

Lyginant su PDL atveju, DLDI identifikavimas yra sudėtingesnis: erdvines išvestines reikia tiksliai įvertinti, o bet koks matavimo triukšmas diferenciavimo metu sustiprinamas, todėl būtinas kruopštus duomenų lyginimas ir slenkstinių verčių derinimas. Nepaisant šių iššūkių, PDE–FIND sėkmingai iš naujo atrado daugelį kanoninių lygčių – difuzijos lygtį, klampiąją Burgerso lygtį, Kortevego–de Friso lygtį ir net 2D Navjė–Stokso advekcijos–difuzijos dėmenis – vien iš triukšmingų sintetinių duomenų, pakanka tik vidutinio tankio duomenų [2]. Šie rezultatai patvirtina, kad SINDy retumo paradigma natūraliai tinka ir DLDI, todėl stipriai motyvuoja ją taikyti šiame darbe nagrinėjamoms reakcijos–difuzijos sistemoms.

2.3 Difuzijos lygtys ir reakcijos–difuzijos sistemos

Reakcijos–difuzijos lygtys sudaro standartinį fizinių, cheminių ir biologinių procesų modelių struktūrą: tiesinis difuzijos dėmuo aprašo erdvinį medžiagos sklidimą, o netiesiniai reakcijos dėmenys užfiksuoja lokalias sąveikas. Toks derinys leidžia atkurti šilumos sklaidą, judančius cheminių reakcijų frontus, bistabilių fazių sąsajas ar ilgalaikes laiko osciliacijas. Kadangi kiekviena iš šių elgsenų skirtingai išbando identifikavimo algoritmą, tolesniems eksperimento etapams šiame darbe pasiliekamos penkios kanoninės lygtys – reakcijos-difuzijos lygtys yra kontroliuojama, pakankamai varijuota lygčių klasė algoritminiam dinamikos identifikavimo tikslumui įvertinti.

Šilumos lygtis (grynoji difuzija). Kanoninis difuzijos modelio pavidalas yra

$$u_t \;=\; D\,\Delta u, \tag{1}$$

kur $u(t,x)$ žymi skaliarinį lauką (pvz. temperatūrą), o $D > 0$ – difuzijos koeficientas. Kadangi lygtis yra tiesinė, jos fundamentalusis sprendinys yra Gauso funkcija; bet koks pradinis netolygumas d matmenų erdvėje monotoniškai nyksta kaip $t^{-d/2}$. Šilumos lygtis veikia kaip etalonas, leidžiantis patikrinti, ar identifikavimo algoritmas geba išskirti *vienintelį* Laplasiano dėmenį.

Fišerio–KPP lygtis (monostabilūs bangos frontai). Logistinio augimo pridėjimas prie difuzijos duoda Fišerio–Kolmogorovo–Petrovskio–Piskunovo modelį

$$u_t \;=\; D\,\Delta u + r\,u(1-u), \qquad r > 0, \tag{2}$$

kuris iš pradžių buvo skirtas aprašyti naudingos mutacijos plitimą erdvėje [7]. Neteigiama pusiausvyra užkariauja nestabilią būseną $u = 0$ judančiais frontais, kurių mažiausias greitis $c^* = 2\sqrt{rD}$. Ši lygtis patikrina, ar retasis atradimas gali atkurti *ir* Laplasiano, *ir* kvadratinį reakcijos dėmenį.

Aleno–Kano lygtis (bistabilios sąsajos). Metalo lydinių fazių atskyrimą aprašo

$$u_t \;=\; D\,\Delta u + \beta\,(u - u^3), \tag{3}$$

kuri yra L^2 srauto gradientas dvigubos duobės potencialui [8]. Sistema bistabili: $u = \pm 1$ – stabilios būsenos, $u = 0$ – nestabili, o susidarančios sąsajos juda kreivumo valdomu greičiu.

Kubinė netiesinė reakcija ir aštrūs erdviniai gradientai leidžia patikrinti algoritmo gebėjimą identifikuoti aukštesnės eilės polinominius dėmenis.

Grėjaus–Skoto modelis (Turingo raštai, savireplikacija). Dviems sąveikaujantiems komponentams U, V galioja

$$U_t = D_U\,\Delta U - UV^2 + F(1-U), \tag{4}$$

$$V_t = D_V\,\Delta V + UV^2 - (F+k)\,V, \tag{5}$$

kur F – tiekimo (feed) greitis, o k – susinaikinimo (kill) greitis [9, 10]. Homogeninės pusiausvyros difuzijos sukeliamas nestabilumas sukuria dėmes, juostas, labirintus ir savireplikuojančias struktūras; tai griežtas raštų formavimosi etalonas.

Šios penkios DLDI kartu apima tiesinę difuziją, monostabilią ir bistabilią kinetiką, daugiakomponentę sąveiką, osciliacijas ir sudėtingą raštų formavimąsi – tai išsamus testų rinkinys retajam sistemų identifikavimui.

2.4 Iššūkiai DLDI identifikavime

Duomenimis grindžiamas dalinių diferencialinių lygčių identifikavimas susiduria su keliais esminiais sunkumais. Pagrindiniai iš jų – triukšmo jautrumas skaičiuojant išvestines, tinkamo kandidatinių funkcijų rinkinio parinkimas bei būtinybė vertinti atrastų modelių galią prognozuoti tyrinėjamą dinamiką, o ne vien suderinimo su naudojamais duomenimis tikslumą. Šie iššūkiai tiesiogiai lemia šio darbo metodologinę kryptį.

Kandidatinio rinkinio sudarymas ir modelio parinkimas. Retosios regresijos prielaida – kad tikroji dinamika priklauso nuo tam tikro kandidatinių funkcijų rinkinio. Jei svarbūs dėmenys praleidžiami arba bibliotekoje gausu stipriai koreliuotų funkcijų, optimizavimas gali praleisti esminę fiziką ar įtraukti klaidingus dėmenis. Nepilnas rinkinys verčia regresiją užpildyti trūkstamą fiziką turimais dėmenimis, o pernelyg plati biblioteka sukelia blogą sąlygiškumą, kai stulpeliai beveik tiesiškai priklausomi [1]. Modelio parinkimą dar labiau komplikuoja poreikis derinti retumą, tikslumą ir gebėjimą apibendrinti – geras priderinimas prie duomenų nebūtinai reiškia, kad modelis gerai veiks su nematytais režimais kaip kitomis pradinėmis ar kraštinėmis sąlygomis.

Triukšmo jautrumas. Išvestinių skaičiavimas iš triukšmingų erdvės–laiko duomenų yra bene problemiškiausias žingsnis retajame DLDI atradime. Net ir žemas triukšmo lygis smarkiai išryškėja atliekant skaitinį diferenciavimą, ypač kai reikia aukštesnių erdvinių išvestinių, tokių kaip u_{xx} ar u_{xxx}. Tai nuolat įvardijama kaip didžiausia kliūtis ankstesniuose darbuose [2, 4], kur parodyta, kad net paprastos baigtinių skirtumų schemos gali sugriauti modelio atkūrimą esant menkiems trikdžiams. Triukšmui slopinti taikomi Gauso filtrai, lokali polinomų aproksimacija, o pastariausiu metu – silpnosios integralinės formos DLDI identifikacija [6], kurioje valdančioji lygtis taikoma integraline prasme, taip mažinamas triukšmo poveikis.

Validacija. Paprastosios SINDy darbo eigos dažnai baigiasi radus retą koeficientų vektorių, tačiau nepatikrinus, ar gauta DLDI simuliacijoje iš tikrųjų atkuria sistemos elgseną. Vien likučio paklaida mokymo duomenyse nebūtinai parodo prognozės galią, ypač triukšmingoje aplinkoje. Simuliacinė validacija – t. y. patikrinimas, ar išmoktas modelis atkuria teisingą ilgalaikę dinamiką ir išlieka teisingas pakitus pradinėms sąlygoms – yra kritiškai svarbi, tačiau literatūroje dažnai praleidžiami [2].

Šio darbo indėlis. Siekiant spręsti minėtus tarpusavyje susijusius iššūkius, šiame darbe kuriama, konfigūruojama ir plečiama DLDI identifikavimo proceso eiga. Ji palaiko sistemingus parametrų rėžius per plačius optimizatorių parametrus, kandidatinių funkcijų rinkinių struktūras ir retumo slenksčius, o kiekvieną identifikuotą modelį vertina plačiu rodiklių rinkiniu: likučio paklaida, R^2 įvertis, santykinė koeficientų paklaida, retumo lygis, tikslumas, atkūrimas ir F1 balas. Šiame darbui pradėta programinė struktūra suteikia galimybę prie skaitinių sprendinių klasės pridėjus norimą tirti dinamiką tyrinėti išsaugant ir kaupiant tyrimų duomenis, palieka galimybę tyrimų rezultatus kaupti galimai kuriant sąsają ir su duomenų baze ar kitomis programinėmis sistemomis.

Sujungus automatines parametrų tinklelių paieškas su griežtomis vertinimo ir validacijos procedūromis, sukuriama tvirta struktūra duomenimis grindžiamam DLDI atradimui tirti.

3 Metodologija

3.1 Metodikos apžvalga

Siūloma darbo eiga, kuri šiame darbe taikoma DLDI identifikavimui remiantis SINDy pagrindu [1, 2].[1] Kiekvienas etapas vykdomas kaip nepriklausomas modulis, todėl skaitinius metodus, rinkinius ar optimizavimo nustatymus galima keisti nemodifikuojant kodo. Modulinė architektūra leidžia vykdyti sistemingus parametrų rėžių (sweep) tyrimus ir gauti lengvai atkartojamus etaloninius rezultatus.

[1] Diagrama pateikiama kitame puslapyje; didelės raiškos versiją žr. priede A.

Etapas 1: **Duomenų generavimas.** Kanoninės DLDI (šilumos, Fišerio–KPP, Aleno–Kano, Grėjaus–Skoto, ir kt.) skaitiškai sprendžiamos pasirinktame erdviniame tinkle ir laiko intervale, gaunant didelės raiškos tikruosius duomenis. Simuliacijos parametrai (laiko žingsnis, erdvinė raiška, kraštinės sąlygos) aprašomi YAML konfigūracijoje, tai leidžia tyrimus vykdyti sistemingai parenkant skaitinių skaičiavimų ir dinaminės sistemos parametrus.

Etapas 2: **Duomenų apdorojimas.** Pirminiai sprendinių laukai apdorojami skaitiniams išvestinių įverčiams gauti. Baigtinių skirtumų schemos arba lokali polinominė aproksimacija pateikia u_t, u_x, u_{xx} ir, jei to reikalautų uždavinys, aukštesnes išvestines. Norint patikrinti tvirtumą, gali būti įtraukiamas Gauso triukšmas, po kurio taikomas pasirenkamas duomenų filtravimas. Rezultatas – tarpusavyje susietas duomenų rinkinys $\{u, \nabla u, \ldots, u_t\}$, su kuriuo vykdoma dinamikos identifikacija.

Etapas 3: **Bibliotekos sudarymas ir retoji regresija.** Sudaroma kandidatinių funkcijų matrica $\Theta(U)$. Taikant retumo prielaidą [1], kiekvienas laiko išvestinių stulpelis (pvz. u_t) regresuojamas į $\Theta(U)$ pasitelkiant retumą skatinantį optimizatorių – STLSQ, SSR ar SR3. Sistemos architektūra leidžia grotelių paieškos būdu (grid search) keisti optimizatoriaus hiperparametrus (slenksčius, ℓ_1 svorius), sugeneruojama daug retų modelių tolesnei lyginamajai analizei.

Etapas 4: **Modelių vertinimas ir atranka.** Kiekviena kandidatė DLDI vertinama pagal kiekybinių metrikų rinkinį: (i) *likučio paklaida* ir R^2 pritaikymo rodiklis, (ii) *retumo lygis.* (iii) *struktūriniai rodikliai* (tikslumas, atkūrimas, F1), kai žinoma tiesa, (iv) *koeficientų santykinė paklaida* parametrų tikslumui, kai žinoma tiesa, Pareto analizės metu atrenkami taupūs modeliai, suderinantys tikslumą ir sudėtingumą.

Ši keturių etapų struktūra sudaro pagrindą visiems eksperimentams. Vėlesniuose Metodologijos skyriaus poskyriuose pateikiami detalūs kiekvieno modulio, nustatymų paruošimo aprašai.

3.2 DLDI sistemų pasirinkimas analizei

Šiame darbe retojo identifikavimo vykdoma eiga tiriama taikant reakcijos-difuzijos lygčių rinkinį, apimantį įvairius dinamikos režimus ir struktūrinį sudėtingumą. Atrinktos sistemos aprėpia tiesinius ir netiesinius difuzijos modelius, bistabilią ir monostabilią kinetiką bei tiek vieno, tiek kelių kintamųjų reakcijos–difuzijos lygtis. Toliau pateikiamos jų matematinės formos ir įtraukimo argumentai.

Šilumos lygtis.

$$\frac{\partial u}{\partial t} = D\,\nabla^2 u \tag{6}$$

Ši tiesinė lygtis aprašo skaliarinio lauko $u(t, x)$ (pvz. temperatūros) grynąją difuziją ir veikia kaip bazinis bandymas. Jos identifikavimas patikrina, ar identifikavimo eiga geba atkurti vienintelį difuzijos dėmenį iš švarių ar triukšmingų duomenų be papildomų reakcijos terminų [2].

Fišerio–KPP lygtis.

$$\frac{\partial u}{\partial t} = D\,\nabla^2 u + r\,u(1-u) \tag{7}$$

Šis monostabilus vienos rūšies modelis turi netiesinį logistinį augimo dėmenį [7]. Jis skirtas patikrinti, ar SINDy gali izoliuoti polinominius netiesiškumus ir atpažinti reakcinės–difuzinės pusiausvyros nulemtus judančius bangos frontus.

Aleno–Kano lygtis.

$$\frac{\partial u}{\partial t} = D\,\nabla^2 u + \gamma\,(u-u^3) \tag{8}$$

Fazės persiskyrimo ir sąsajų judėjimo modelis [8] įveda kubinį bistabilų reakcijos dėmenį. Jis išbando regresiją aukštesnės eilės netiesiškumo sąlygomis..

Grėjaus-Skoto modelis.

$$\frac{\partial u}{\partial t} = D_u\,\nabla^2 u - uv^2 + F(1-u), \tag{9}$$

$$\frac{\partial v}{\partial t} = D_v\,\nabla^2 v + uv^2 - (F+k)v \tag{10}$$

Ši dviejų reagentų reakcijos–difuzijos sistema sukelia Turingo raštus ir turtingą erdvės–laiko dinamiką [9, 10]. Ji įvertina grandinės gebėjimą atkurti kelių lygčių sistemas su trečio laipsnio netiesiniais dėmenimis.

Bendras šių modelių rinkinys sudaro platų ir reprezentatyvų pagrindą, skirtą retųjų DLDI atradimo metodų tikslumui, tvirtumui ir bendrumui patikrinti.

3.3 Duomenų generavimas (skaitinė simuliacija)

Sintetiniai duomenys gaunami skaitiškai integruojant kiekvieną DLDI tolygiame tinklelyje. Erdvinės išvestinės aproksimuojamos antrojo laipsnio centriniais baigtiniais skirtumais, o laiko integracijai taikomos neišreikštinė arba neišreikštinė–išreikštinė (IMEX) schema: netiesiniai reakcijos dėmenys skaičiuojami išreikštine schema, o difuzijos dėmenys – neišreikštiniu būdu, todėl galima naudoti didesnį laiko žingsnį Δt nei reikalauja griežta difuzinė Courant–Friedrichs–Lewy (CFL) riba. Abiejuose srities galuose pritaikomos tolydžios Niumano kraštinės sąlygos. Tinklo žingsniai Δx ir Δt parenkami taip, kad lokali paklaida būtų nereikšminga palyginti su signalo dydžiu, o schemos neišreikštinė dalis išliktų besąlygiškai stabili (pagal klasikinį FTCS stabilumo kriterijų $\Delta t \leq \Delta x^2/(2D)$ šilumos lygčiai) [13].

Šilumos lygtis.

$$u_t = \alpha\, u_{xx}.$$

IMEX su implicite difuzija, $\alpha = 0.01$; tinklelis $N = 128$ intervale $[0, 1]$, $\Delta t = 0.1$. Pradinė sąlyga: centrinis „top-hat“ profilis; kraštinės sąlygos: Niumano.

Fišerio–KPP (monostabili).

$$u_t = D\, u_{xx} + r\, u(1-u), \qquad D = 0.01,\ r = 1.$$

Pradiniai duomenys – Gauso impulso formos; tas pats tinklelis; IMEX schema.

Aleno–Kano lygtis (dvistabilė).

$$u_t = D\, u_{xx} + \kappa\,(u - u^3), \qquad D = 0.01,\ \kappa = 1.$$

Viengubas Gauso kalnelis kaip pradinis profilis; IMEX su implicite difuzija.

Grėjaus–Skoto modelis (raštų formavimasis, dvi rūšys).

$$u_t = D_u\, u_{xx} - uv^2 + F(1-u), \tag{11}$$

$$v_t = D_v\, v_{xx} + uv^2 - (F+k)v, \tag{12}$$

kur $D_u = 0.01$, $D_v = 0.005$, $F = 0.04$, $k = 0.06$. Pradinė būsena $u \approx 1$, $v \approx 0$ plius maža centrinė perturbacija.

Table 1: Skaitinių parametrų suvestinė sintetinėms duomenų generavimo simuliacijoms.

Sistema	N_x	N_t	Δt	Pagrindiniai parametrai
Šilumos	16	16	0.10	$\alpha = 0.01$
Fišerio–KPP	16	16	0.10	$D = 0.01,\ r = 1$
Aleno–Kano	32	16	0.10	$D = 0.01,\ \kappa = 1$
Grėjaus–Skoto	64	64	0.10	$D_u = 0.01,\ D_v = 0.005,\ F = 0.04,\ k = 0.06$

3.4 SINDy metodas DLDI identifikavimui

Netiesinės dinamikos retojo identifikavimo eigos struktūra (Sparse Identification of Nonlinear Dynamics, SINDy) [1, 2] sudaro šio darbo modelių atradimo eigos pagrindą. Turint erdvės–laiko duomenų aibę $\mathbf{U}(x, t)$, sukonstruojama požymių matricą Θ, kurios stulpelius sudaro kandidatinių funkcijų vertės, priklausančios nuo $\mathbf{U}$ ir jos išvestinių, ir taiko retąją regresiją, kad izoliuotų kelis dinamiką geriausiai paaiškinančius dėmenis iš koeficientų stulpelio ξ. DLDI kontekste būtina itin kruopščiai apibrėžti tris technines komponentes:

1. **Skaitinio diferenciavimo schemą**, kurios tikslumas riboja įvesties duomenų kokybę.
2. **Kandidatinių funkcijų rinkinį**, kuri nustato paieškos erdvės raiškumą;
3. **Retosios regresijos optimizatorių** ir jo hiperparametrus, lemiančius kompromisą tarp identifikuoto modelio tikslumo ir retumo;

Toliau pateikiami pasirinkimai ir motyvai kiekvienam iš šių komponenčių.

3.4.1 Kandidatinių funkcijų rinkinys

Retasis identifikavimas remiasi prielaida, kad tikroji DLDI dešinioji pusė yra reta, t.y. yra tik kelių dėmenų suma pakankamai turtingoje funkcijų erdvėje. Praktiškai tai reiškia, jog formuojama *kandidatinių požymių matrica*

$$\Theta(U) \;=\; \big[\,\theta_1(U) \,\big|\, \theta_2(U) \,\big|\, \dots \,\big|\, \theta_M(U)\big] \;\in\; \mathbb{R}^{K\times M},$$

kurios K eilutės atitinka erdvės–laiko ėminius, o M stulpelį sudaro iš anksto apibrėžtos funkcijos θ_j, įvertintos duomenyse [1, 2]. Retoji regresija tuomet ieško koeficientų vektoriaus $\mathbf{\Xi}$, tenkinančio

$$\Theta(\mathbf{U})\,\mathbf{\Xi} \;\approx\; \mathbf{U}_t,$$

, su kiek įmanoma mažiau nenulinių elementų.

kandidatinių funkcijų rinkinys Vienam kintamajam $u(x,t)$ pradedame nuo konstantos ir monomų iš u bei jo erdvinių išvestinių:

$$\{ 1,\ u,\ u^2,\ u^3,\ u_x,\ u_{xx},\ \sin u,\ \cos u,\ \dots \}.$$

Kelių kintamųjų atveju (u,v) rinkinys praplečiamas visomis pasirinkto laipsnio sandaugomis, pvz.

$$\{u,\ v,\ u^2,\ uv,\ v^2,\ u^3,\ u^2v,\ uv^2,\ v^3,\ u_x,\ v_x,\ u_{xx},\ v_{xx},\ \dots\}$$

Toks rinkinio pagrindas pakankamas 3.2 skirsnyje pateiktoms kanoninėms lygtims aprašyti ir išlieka aiškiai interpretuojamas.

Raiškumas ir taupumas Per didelis funkcijų rinkinys kelia kolinearumo ir perteklinio priderinimo riziką, o pernelyg siaura – gali praleisti dinamikai įtakos turinčius dėmenis. Dėmens stabilumą π suprasime kaip empirinę tikimybę, kad funkciją atitinkantis dėmuo bus įtrauktas į sprendinį optimizatorių hiperparametrų gardelės šlavime ar šlavimuose. Taikoma dviejų žingsnių strategija [4]:

1. Pradedame nuo *plačios* bibliotekos, kurioje, tikėtina, yra tikrieji dėmenys;

2. Atliekame kelis identifikavimo bandymus (triukšmo vertės, hiperparametrų optimizatorių gardelės parinkimo, slenkstinių verčių rėžiai) ir vertiname funkcijas pagal jų stabilumą, vidutinį lygties retumą, vidutinę likučio paklaidą kai atitinkama funkcija naudojama identifikuotoje lygtyje.

Įprastai neaktyvūs dėmenys, dėmenys naudojami tik pertekliniuose įverčiuose arba tie, kurių pašalinimas nežymiai padidinalikučio paklaidą, atmetami, paliekant sumažintą rinkinį, kuris išlieka raiškus, bet ir taupus. Tokia sisteminga atranka sumažina klaidingų dėmenų įtraukimo tikimybę didelio matmens rinkiniuose [3] ir sudaro valdymo būdą tirti, kaip kandidatinių funkcijų rinkinio platumas veikia identifikavimo tikslumą tolesniuose eksperimentuose.

3.4.2 Retoji regresija ir parametrų parinkimas

Sudėjus požymių matricą $\Theta(\mathbf{U})$, atradimo uždavinys suvedamas iki

$$\Theta(\mathbf{U})\,\boldsymbol{\Xi} \;\approx\; \mathbf{U}_t,$$

su papildoma prielaida, kad tik keli $\mathbf{\Xi}$ elementai yra nenuliniai. Uždavinys sprendžiamas minimizuojant

$$\frac{1}{2}\left\|\Theta\,\mathbf{\Xi}-\mathbf{U}_t\right\|_2^2 \;=\; \frac{1}{2}\sum_{i=1}^{m}\Big(\sum_{k=1}^{p}\Theta_{i,k}\,\Xi_k \;-\; (U_t)_i\Big)^2$$

Pridėdami retumą skatinančią vertę

$$\min_{\mathbf{\Xi}}\;\tfrac{1}{2}\left\|\Theta\mathbf{\Xi}-\mathbf{U}_t\right\|_2^2 \;+\; \lambda\,\mathcal{R}(\Xi), \tag{13}$$

kur $\mathcal{R}$ koduoja retumo sąvoką, o slenkstinė vertė $\lambda > 0$ valdo jos stiprumą.

Kelios dažniausiai naudojamos $\mathcal{R}$ retumą skatinančios formos:

$$\left\|\mathbf{\Xi}\right\|_0 \;=\; \sum_{k=1}^{p}\mathbf{1}\left[\Xi_k \neq 0\right]$$

$$\left\|\mathbf{\Xi}\right\|_1 \;=\; \sum_{k=1}^{p}\left|\Xi_k\right|$$

$$\left\|\mathbf{\Xi}\right\|_2^2 \;=\; \sum_{k=1}^{p}\left(\Xi_k\right)^2$$

Šiame darbe naudoti optimizavimą įgyvendinantys algoritmai:

1. **STLSQ** [1]. Iteratyviai per l taikomas:

(i) $\Xi^{(l)} \;=\; \arg\min_{\Xi}\Big[\tfrac{1}{2}\|\Theta\Xi-\mathbf{u}_t\|_2^2+\dfrac{\alpha}{2}\|\Xi\|_2^2\Big].$
(ii) Slenkstinei vertei λ, $\quad \xi_j^{(l)}=0, \quad$ jei $\quad |\xi_j^{(l)}|<\lambda.$
(iii) Pritaikomas gautas Ξ. Jei $\Xi^{(l)} \neq \Xi^{(l-1)}$, kartojamas (i).

2. SR3 [16, 17]. čia w yra $\boldsymbol{\Xi}$ kopija. Iteratyviai per l taikoma:

$$\boxed{\begin{aligned}&\text{(i) } \boldsymbol{\Xi}^{(l)} \;=\; \arg\min_{\boldsymbol{\Xi}}\Big[\frac{1}{2}\,\|\Theta\,\boldsymbol{\Xi}-\mathbf{U}_t\|_2^2 \;+\; \frac{\nu}{2}\,\|\boldsymbol{\Xi}-w^{(l-1)}\|_2^2\Big],\\ &\quad \text{t. y. } (\Theta^T\Theta+\nu I)\,\boldsymbol{\Xi}^{(l)} = \Theta^T\mathbf{U}_t \;+\; \nu\, w^{(l-1)},\\ &\text{(ii) } w^{(l)} \;=\; \arg\min_{w}\Big[\lambda\,\mathcal{R}(w) \;+\; \frac{\nu}{2}\,\|\boldsymbol{\Xi}^{(l)}-w\|_2^2\Big] \;=\; \operatorname{prox}_{\frac{\lambda}{\nu}\mathcal{R}}\big(\boldsymbol{\Xi}^{(l)}\big),\\ &\quad \text{(negriežtas slenkstis jei } \mathcal{R}=\|\cdot\|_1\text{, arba griežtas slenkstis jei } \mathcal{R}=\|\cdot\|_0\text{)},\\ &\text{(iii) Jei } \boldsymbol{\Xi}^{(l)}\neq\boldsymbol{\Xi}^{(l-1)},\ \text{kartojama (i)–(ii).}\end{aligned}}$$

3. SSR (Stepwise Sparse Regression) [15]. Prasideda nuo pilno modelio su aktyvių indeksų rinkiniu $S^{(0)} = \{1,\dots,p\}$. Iteratyviai per l taikoma:

$$\boxed{\begin{aligned}&\text{(i) Fito regresija su aktyviais nariais } S^{(l)}:\\ &\quad \boldsymbol{\Xi}^{(l)} \;=\; \arg\min_{\boldsymbol{\Xi}}\Big\|\Theta_{S^{(l)}}\,\boldsymbol{\Xi}_{S^{(l)}}-\mathbf{U}_t\Big\|_2^2,\quad \Xi_k^{(l)}=0\quad \text{jei } k\notin S^{(l)},\\ &\text{(ii) Surandamas mažiausiai reikšmingas narys } j^{(l)} \;=\; \arg\min_{k\in S^{(l)}}\big|\Xi_k^{(l)}\big|,\\ &\text{(iii) Atmetamas narys } j^{(l)} \text{ iš } S^{(l)}:\ S^{(l+1)} = S^{(l)}\setminus\{\,j^{(l)}\},\\ &\quad \text{jei } |S^{(l+1)}|\geq 1,\ \text{kartojama (i)–(iii).}\end{aligned}}$$

Naudingas norint vertinti klaidos–retumo kompromisą iš identifikuojamo modelio šalinant dėmenis panariui: kiekviename žingsnyje pašalinamas silpniausias narys, perskaičiuojama regresija ir stebima, kaip kinta modelio paklaida.

4. FROLS (Forward Regression Orthogonal Least Squares) [14]. Prasideda nuo tuščio modelio ir kiekviename žingsnyje pridedamas tas bibliotekos narys, kuris maksimizuoja liekanos klaidos sumažinimą. Metodas yra jautrus triukšmui. Paprastas godus algoritmas, leidžiantis greitai atrinkti reikšmingiausius modelio dėmenis.

Hiperparametrų rėžiai Slenkstinę vertę λ ir kitus optimizavimo tikslo funkciją atitinkančius parametrus varijuojame pasirinktuose intervaluose. Kiekvienam optimizatoriaus parametrų, λ deriniui sistemoje registruojame

- *likučio paklaidą* $\|\Theta\boldsymbol{\Xi}-\mathbf{U}_t\|_2$ ar reliatyvią likučio paklaidą $r = \frac{\|\Theta\boldsymbol{\Xi}-\mathbf{U}_t\|_2}{\|\mathbf{U}_t\|_2}$,
- *retumą* (aktyvių dėmenų skaičių),

- struktūrinius rodiklius (tikslumas, atkūrimas, F1), kai žinoma tiesa,
- koeficientų paklaidą $\|\Xi - \Xi_{\text{true}}\|_2$ (sintetiniams duomenims),
- simuliacijos paklaidą, taikant atrastą DLDI naujoms pradinėms sąlygoms.

Pareto analize pagrįsta modelio atranka Braižant likučio paklaidos ir retumo grafiką gaunama Pareto kreivė. Svarba suteikiama modeliams, kurie yra pakankamai reti bei turi mažiausią likučio paklaidą. Tinkamai atlikus kandidatinio rinkinio bei optimizatoriaus parametrų tinklelio parinkimus, grafike matoma Pareto alkūnė, t.y. optimaliausi pakankamai reti modeliai su mažomis likučio paklaidų vertėmis. Renkami mažiausiai dėmenų turintys modeliai, tenkinantys

$$r \;\leq\; (1+\varepsilon)\,\min r,$$

su tam tikra ϵ verte visuose eksperimentuose. Tokia atranka pirmenybę teikia modelio retumui, bet išlaiko likutį optimaliose ribose; tai atitinka modelių atrankos rekomendacijas (pvz. Maddu *et al.*, 2019).

Automatizavimas. Visi parametrų rėžiai šiame darbe apibrėžiami YAML konfigūracijoje ir vykdomi skaičiavimus apvelkančiu kodu. Visiems identifikuojamiems modeliams metrikos išsaugomos struktūrizuotame aplanke vėlesnei Pareto analizei ir vizualizacijai. Toks automatizavimas garantuoja atkuriamumą ir leidžia nuodugniai ištirti sprendiklių hiperparametrus be rankinio įsikišimo.

Optimizavimo algoritmai STLSQ ir SR3 sudaro pagrindinę identifikavimo eigos ašį, o SSR, FROLS algoritmai galimai naudojami papildomai identifikuojamumo validacijai. Sistemingi hiperparametrų šlavimai ir Pareto paremta atranka užtikrina, kad galutiniai modeliai būtų ne tik reti, bet ir aprašytų realią modelį atitinkančią dinamiką.

3.4.3 Išvestinių skaičiavimas

Tiksliai apskaičiuotos erdvinės ir laiko išvestinės yra būtinos matriciai $\Theta(\mathbf{U})$ sukonstruoti. Darbe naudojami *Baigtiniai skirtumai* tame pačiame tolygiame tinklely, kuris taikomas ir simuliacijai. Esant tinklelio žingsniui Δx ir laiko žingsniui Δt, išvestinėms įvertinti naudojami:

$$u_x(x_i, t_n) \approx \frac{u(x_{i+1}, t_n) - u(x_{i-1}, t_n)}{2\Delta x}, \tag{14}$$

$$u_{xx}(x_i, t_n) \approx \frac{u(x_{i+1}, t_n) - 2u(x_i, t_n) + u(x_{i-1}, t_n)}{\Delta x^2}, \tag{15}$$

$$u_t(x_i, t_n) \approx \frac{u(x_i, t_{n+1}) - u(x_i, t_{n-1})}{2\Delta t}, \tag{16}$$

su $O(\Delta x^2)$ ir $O(\Delta t^2)$ klaidomis pakankamai glodžiai funkcijai.

Triukšmo jautrumas. Baigtiniai skirtumai sustiprina aukšto dažnio triukšmą; net ir nedidelė matavimo paklaida gali ženkliai iškraipyti išvestinių įverčius ir, atitinkamai, pabloginti retosios regresijos rezultatus [2, 4]. Šiame darbe dirbant su triukšmingais duomenimis taikomas išankstinis Savickio-Golajaus filtravimas. Savickio-Golajaus formos duomenų gludinimas prieš išvestinių aproksimaciją leidžia sumažinti triukšmo poveikį ir gauti stabilesnius koeficientų įverčius retosios regresijos etape [18].

Santrauka. Kandidatinių funkcijų rinkinio parinkimas (Skirsnis 3.4.1), retąją optimizaciją (Skirsnis 3.4.2) ir čia aprašyta išvestinių skaičiavimo tvarka (Skirsnis 3.4.3) kartu sudaro darbo eigą taupioms DLDI modelių formoms atkurti iš duomenų. Vėlesniuose skyriuose parodoma, kad derinant sistemingus parametrų rėžius su Pareto optimalia modelio atranka ši konfigūracija patikimai identifikuoja valdančiąsias lygtys įvairiose difuzijos ir reakcijos–difuzijos sistemose.

3.5 Vertinimo metrikos

Išsamus identifikuotos DLDI vertinimas turi patikrinti: *(i)* ar pasirinkti teisingi dėmenys, *(ii)* ar jų koeficientai kiekybiškai tikslūs ir *(iii)* ar gautas modelis atkuria stebimą dinamiką. Tam sekamos papildomos metrikų klasės.

1. Likučio paklaida (regresijos pritaikymas). Remiantis [1] vertinama, kaip gerai modelis atkuria stebimą laiko išvestinę, naudodami *santykinę likučio paklaidą*

$$r = \frac{\| \hat{U}_t - U_t\|_2}{\| U_t\|_2}.$$

Maža likučio paklaida rodo, kiek išmoktos DLDI dešinioji pusė sutampa su duomenimis iki skaitinės bei matavimo paklaidos. Lygčių sistemos atveju likutis galimas tiek kiekvienai lygčiai ir kaip bendras vidurkis.

2. Retumas. Pagrindinė retojo identifikavimo prielaida [1, 2, 3] su

$$s = \|\hat{\boldsymbol{\xi}}\|_0$$

yra tai, kad s yra pakankamai mažas.

Normalizuotas retumas:

$$\tilde{s} = \frac{s}{N_\Theta},$$

kur N_Θ kandidatinių funkcijų rinkinyje skaičius.

3. Struktūrinis tikslumas (dėmenų parinkimo teisingumas). Tegu $\mathcal{S}_{\text{true}}$ ir $\mathcal{S}_{\text{pred}}$ žymi tikrojo ir identifikuoto modelio nenulinių koeficientų indeksų aibes. Apibrėžiame

$$\text{TP} = |\mathcal{S}_{\text{true}} \cap \mathcal{S}_{\text{pred}}|, \qquad \text{FP} = |\mathcal{S}_{\text{pred}} \setminus \mathcal{S}_{\text{true}}|, \qquad \text{FN} = |\mathcal{S}_{\text{true}} \setminus \mathcal{S}_{\text{pred}}|.$$

Iš jų apskaičiuojame

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}}, \qquad \text{recall} = \frac{\text{TP}}{\text{TP+FN}}, \qquad F_1 = \frac{2\,\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}.$$

Didelis *precision* baudžia *nereikšmingus* dėmenis, didelis *recall* – *praleistus* dėmenis; jų harmoninis vidurkis F_1 suteikia vieno skaičiaus įvertį. Visi trys rodikliai skaičiuojami *kiekvienai lygčiai* ir pateikiami kartu su viso modelio vidurkiu.

4. Koeficientų paklaida. Net ir idealiai atkurus dėmenų rinkinį skaitiniai koeficientai gali skirtis nuo tikrųjų. Kiekvienai lygčiai i fiksuojame santykinę ℓ_2 paklaidą

$$E_{\text{coef}}^{(i)} = \frac{\|\hat{\boldsymbol{\xi}}^{(i)} - \boldsymbol{\xi}^{(i)}\|_2}{\|\boldsymbol{\xi}^{(i)}\|_2},$$

kur $\boldsymbol{\xi}^{(i)}$ – tikras koeficientų vektorius, o $\hat{\boldsymbol{\xi}}^{(i)}$ – identifikuotasis. Bendroji koeficientų paklaida gaunama sujungus visų lygčių vektorius ir atspindi modelio *kiekybinį* tikslumą.

5. Simuliacijos paklaida Griežtas testas – plačiai taikomas literatūroje – yra integruoti identifikuotą DLDI į priekį ir palyginti sprendinį su atidėtais tikraisiais duomenimis:

$$r = \frac{\| \hat{U} - U\|_2}{\| U\|_2}.$$

Nors šiame darbe simuliacijos paklaida bendruoju atveju neskaičiuojama, ji įvardijama išsamumo dėlei ir aptariama kaip natūrali ateities darbų kryptis.

Metrikos pateikiamos santykine forma, kad neliktų jautrios mąsteliui ir palyginamos tarp skirtingų kintamųjų. Pateikiant rezultatus kiekvienai lygčiai atskirai išsaugoma įžvalga į daugiakomponentę dinamiką, kad puikūs rezultatai vienoje dalyje neužmaskuotų trūkumų kitoje. Šie kriterijai drauge užtikrina subalansuotą struktūrinį, parametrinį ir aprašomąjį vertinimą, kuris sudaro Pareto analizės pagrindą modelio atrankai 3.4.2 skyriuje.

3.6 Triukšmo valdymas duomenyse

Realūs eksperimentiniai matavimai visuomet yra užteršti triukšmu, todėl atsparumas tokiems trikdžiams yra kritinis bet kurios duomenimis grindžiamos identifikavimo schemos etalonas [2, 4]. Norėdami imituoti šią situaciją nagrinėjant kelis atvejus su triukšmu, sintetiniai sprendiniai prieš bet kokį išvestinių skaičiavimą užteršiami Gauso triukšmu.

Triukšmo modelis. Kiekvienam erdvės–laiko taškui nustatome

$$u_{\mathrm{tr}}(x,t) \;=\; u(x,t) \;+\; \sigma\,\eta(x,t), \qquad \eta \sim \mathcal{N}(0,1),$$

kur $\sigma = \alpha\,\|u\|_\infty$, $\alpha \in [0,1]$. Triukšmas įliejamas iki erdvinio ar laiko diferenciavimo, kad būtų tiksliai atspindėtas jo sustiprinimas baigtiniais skirtumais.

Apdorojimas. Tvarkantis su triukšmu naudojamas Savickio-Golajaus filtravimas prieš diferenciavimą bei Gauso filtravimas pasitelkiant sąsūką su normalizuotu diskrečiuoju Gauso branduoliu erdvinės diskretizacijos kryptimi. Nors DLDI identifikavimas dėl baigtinių skirtumų triukšmo stiprinimo tampa ypač komplikuotas, darbe tyrinėjami keli triukšmą vertinantys identifikavimo atvejai. Vis dėlto pastariausi darbai dinaminių sistemų identifikavimo srityje susitelkia į silpnosios integralinės modelio formuluotės sprendimą, kuris ypač ženkliai sumažina triukšmo įtaką paklaidai [6].

Eksperimentinė procedūra. Tiriama DLDI simuliuojama naudojant skaitinius metodus. Visos identifikavimo metrikos, apibrėžtos 3.5 skyriuje, registruojamos kiekvienai parinkto optimizatoriaus hiperparametrų kombinacijai ir kiekybiškai įvertinti degradavimą *(i)* dėmenų parinkimo teisingumui, *(ii)* koeficientų tikslumui ir *(iii)* likučio paklaidai. Rezultatai, pateikiami **??** skyriuje, parodo triukšmo slenksčius, nuo kurių retasis identifikavimas nebepatikimas.

Kontekstas. Retasis identifikavimas gali būti pritaikytas DLDI atveju [2, 4]. Sistemingai tiriant mes atkartojame ir išplečiame šias išvadas, pateikdami skaidrų metodo tvirtumo įvertinimą. Taip pat nagrinėjame atvejus su triukšmo įtaka tyrinėjamose sistemose. Apibendrindami rezultatus identifikuojame ir nurodome kryptis – tokias kaip apibendrintos simuliacijos ir simuliacijos paklaidų įvertinimo įvedimą, silpnos integralinės formos uždavinių sprendimą potencialiems patobulinimams ateityje.

3.7 Identifikavimo eigos santrauka

Programavimo dalyje kiekvienas aprašomos darbo eigos etapas konstruojamas kaip atskiras modulis ir yra lengvai konfigūruojamas, todėl galima atkartojamai ir masiniu būdu tyrinėti identifikuojamų modelių erdvę. Įprastai SINDy konfigūracijoje [1] skirtingoms optimizatorių parametrų vertėms gali tekti tokius kaip erdvinių išvestinių skaičiavimus kartoti. Siekiant išvengti laiko ir išteklių eikvojimo tokie skaičiavimai optimizuojami ir suformuojami taip, kad būtų atliekami tik vieną kartą, parametrų šlavimo skaičiavimai šiame darbe paralelizuojami išnaudojant visus kompiutacinius išteklius bei rezultatai yra fiksuojami įrašant visus šlavimus skirtingiems optimizatoriams, tai palieka galimybę kaupti ir nagrinėti net ir ypač didelius kiekius rezultatų šioje bei ateities analizėse. Pagrindiniai žingsniai:

1. **Sintetinių duomenų generavimas.** Skaitiškai išsprendžiamos pasirinktos DLDI nustatytais parametrais, gaunant aukštos raiškos erdvės–laiko duomenis.

2. **Triukšmo įvedimas (nebūtinas).** Į simuliuotus laukus pridedamas Gauso triukšmas nustatytu lygiu, taip imituojama eksperimentiškai renkamų duomenų matavimo paklaida.

3. **Išvestinių skaičiavimas.** Erdvės ir laiko išvestinės skaičiuojamos pasitelkiant centrines baigtinių skirtumų schemas pagal pasirinkimą įvedant ir Savickio-Golajaus filtravimą triukšmui mažinti, taip gaunamos Θ matricos įvestys.

4. **Kandidatinių funkcijų rinkinio sudarymas.** Sudaromas galimai perpildyta kandidatinių funkcijų rinkinys, dažnai polinominiai ir išvestinių kombinacijos (žr. 3.4.1).

5. **Tyrimo tipo parametrų šlavimas.** Atsitiktiniu ar bendru parametrų intervalų grotelių ėmimu varijuojami optimizatoriaus hiperparametrai (pvz. slenkstis λ, SR3 relaksacija ν) ir įrašomi kiekvieno bandymo likučio paklaida ir retumas, kitos metrikos.

6. **Dėmenų agregavimas ir genėjimas.** Analizuojami bandymų rezultatai, skaičiuojami vidutiniai likučiai kiekvienai kandidatinei funkcijai ir dėmenų įtraukimo dažniai; nereikalingos funkcijos nuosekliai pašalinamos iš pradinio rinkinio taip jį rafinuojant.

7. **Struktūrizuotas šlavimas sumažintoje bibliotekoje.** Atliekamas sistemingas grotelių paieškos tyrimas susiaurintuose hiperparametrų rėžiuose, sugeneruojant tankų kandidatinių modelių rinkinį.

8. **Retas modelio atradimas.** Sumažintoje bibliotekoje taikomi STLSQ, SR3, galimai ir FROLS bei SSR algoritmai retosioms DLDI formoms identifikuoti (žr. 3.4.2).

9. **Pareto pagrįsta modelio atranka.** Analizuojant retumo ir likučio santykį siekiama, kad būtų rasti optimalūs modeliai. Galutinis modelis (-iai) parenkamas (-ami), derinant tikslumą ir taupumą.

10. **Galutinis vertinimas.** Vertinami tikslumas, atkūrimas, F_1, koeficientų paklaida ir likučio paklaida kiekvienai lygčiai (žr. 3.5), kad būtų patikrintas struktūrinis ir kiekybinis tikslumas.

Visus etapus valdo tyrimo konfigūracijas nusakanti programa, kuri iš YAML failo nuskaito simuliacijos, bibliotekos ir optimizatoriaus nustatymus, vykdo modulius eilės tvarka ir registruoja detalias metrikas tolimesnei analizei. Toks automatizavimas užtikrina, kad eksperimentai būtų atkuriami, plečiami ir lengvai pritaikomi naujoms DLDI sistemoms su minimaliu rankiniu įdirbiu.

4 Eksperimentai ir rezultatai

4.1 Eksperimentų apžvalga

Atskiri tyrimų atvejai parengti taip, kad būtų atsižvelgta į pagrindinius veiksnius, lemiančius DLDI identifikavimą:

1. **DLDI sistemos duomenų paruošimas** (Šilumos, Fišerio–KPP, Aleno–Kano, Grėjaus–Skoto);
2. **Matavimo triukšmo lygis** (Gauso triukšmas);
3. **Kandidatinio rinkinio parinkimas** (pradinis pilnas rinkinys prieš apgenėtą, sumažintą rinkinį, gautą 3.4.1 skyriuje aprašyta tvarka).
4. **Modelio retumo ir likučio optimalumas** (mažiausio likučio paklaidos pakankamai reti identifikuoti modeliai)

Visi bandymai valdomi YAML konfigūracija ir vykdomi pasitelkiant automatizuotą programinę sistemą, todėl leidžiamas pilnas atkuriamumas ir sistemingas 3.5 skirsnyje apibrėžtų metrikų bei pačių identifikuotų modelių registravimas konkretiems optimizatorių parametrų rinkiniams. Toliau pateikti rezultatai analizuojami triukšmo, bibliotekos dydžio ir sistemos sudėtingumo aspektais, parodant siūlomos identifikavimo darbo eigos pranašumus ir apribojimus.

4.1.1 Šilumos lygtis

Skaitinė sąranka. Sintetiniai šilumos lygmens duomenys sugeneruoti labai stambiame tinkle $N_x \times N_t = 16 \times 16$, kai sritis $L = 1$, o žingsniai $\Delta x = L/N_x$, $\Delta t = 0{,}1$. Difuzijos koeficientas $D = 0{,}01$, tad valdančioji DLDI yra $u_t = 0{,}01\, u_{xx}$. Skaitiniam sprendimui taikoma neišreikštinė Eulerio schema (kartu su Thomas algoritmu) su Neumano kraštinėmis sąlygomis, Gauso iškilumu srities x centre kaip pradine sąlyga. Identifikavimui atlikta *10 000* atsitiktinai parinktų STLSQ regresijų su slenkstinėmis vertėmis:

$$\lambda \in [0.01, 10], \qquad \alpha \in [10^{-6}, 10^{-2}],$$

kur λ – retumo slenkstis, o α – svoris, stabilizuojantis mažiausių kvadratų sprendinį, skatinantis retumą.

Table 2: STLSQ modeliai prieš ir po kandidatinio rinkinio genėjimo.

ID	λ	s	r	Identifikuota dešinioji pusė
3727	0.15	3	**0.0776**	$-0.2035\, u^2 + 0.016 \cos u + 0.007\, u_{xx}$
8116	0.17	2	0.1054	$-0.0842\, u^2 + 0.0077\, u_{xx}$
2835	**1.23**	**1**	0.1148	$0.0083\, u_{xx}$

.

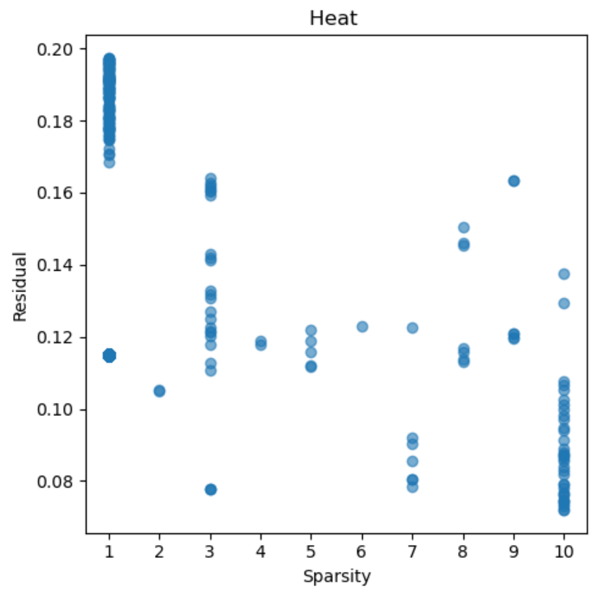

Figure 1: **(a)** Likučio paklaida ir retumo grafikas (10 000 STLSQ bandymų).

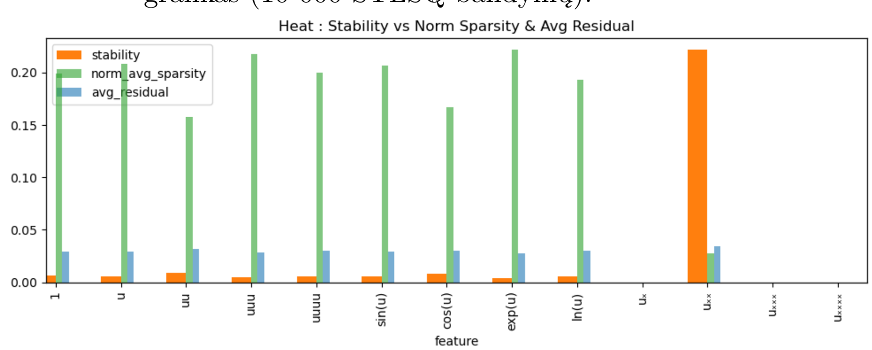

Figure 2: **(b)** Dėmenų stabilumas, normalizuotas vidutinis retumas bei vidutinis likutis.

Interpretacija.

- Mažiausių likučių taškai (ID 3727, 8116) yra Pareto optimaliame priekyje pav. 4.1.1(a), bet abu turi klaidingų dėmenų. Taškų dėmenų žema stabilumo vertė 1 pav. 4.1.1(b) ($\pi_j < 0.02$) ir didelis vidutinis retumas rodo perteklinį regresijos prisitaikymą prie duomenų.
- Bibliotekos genėjimas, remiantis vidutiniu retumu, pašalina retai pasirenkamus dėmenis. Pakartotinai paleidus STLSQ su slenkstine $\lambda \in [0.7, 2]$ ir keliomis α reikšmėmis, nuosekliai gaunamas vienas dėmuo – šilumos lygties Laplasianas.
- Gautas rezultatas atkuria vadovėlinį šilumos lygmens pavyzdį [2], patvirtindamas, kad pašalinus klaidingus stulpelius SINDy patikimai aptinka difuzijos dėsnį net labai stam-

biame tinkle.

- Analogiškas bandymas su triukšmingais duomenimis leidžia patikimai atkurti šilumos lygtį iki 3

4.1.2 Fišerio–KPP lygtis: identifikavimas 16×16 tinkle

Skaitinė sąranka. Vieno komponento Fišerio–KPP sistema

$$u_t = 0.01\, u_{xx} + u - u^2, \qquad x \in [0, 1],$$

sprendžiama tame pačiame 16×16 erdvės–laiko tinkle kaip ir šilumos lygmens teste. Difuzija integruojama implicitiškai (Thomas sprendikliu), logistinis reakcijos dėmuo $u - u^2$ – eksplicitiškai, o pradinė sąlyga – Gauso kalnelis. Atlikta *10 000* atsitiktinių STLSQ regresijų, šluojant $\lambda \in [0.01, 10]$ ir $\alpha \in [10^{-6}, 10^{-2}]$.

Table 3: Reprezentatyvūs STLSQ modeliai ($s = 3$) Pareto fronte.

ID	λ	s	r	Identifikuota dešinioji pusė
1719	1.25	3	0.1122	$+0.944\, u \;-\; 0.944\, u^2 \;+\; 0.0084\, u_{xx}$
2988	2.27	3	0.1177	$-0.773\, u^2 \;+\; 0.900\, \sin(u) \;+\; 0.0085\, u_{xx}$

Interpretacija.

- *Teisinga grupė.* Tankus taškų klasteris ties retumu $s = 3$ (ID 1719, 1800, . . .) atkuria tikėtinus Fišerio dėmenis u, $-u^2$ ir u_{xx} su likučio paklaida $r \approx 0.112$.
- *Sinusinė pakaitinė išraiška.* Antras klasteris su beveik tapačia paklaida linearų augimo dėmenį u pakeičia $\sin(u)$. Kadangi mažoms amplitudėms $\sin(u) \simeq u$, abu stulpeliai požymių matricoje labai koreliuoti, todėl aukštesnių slenksčių atveju gali likti bet kuris jų (5).
- *Bibliotekos genėjimo sprendimas.* Dėmenų statistika (1 pav. 4.1.2(b)) rodo, kad u šiek tiek stabilesnis už $\sin(u)$, o fizika eliminuoja periodinį jėgą, todėl po pirmojo genėjimo ciklo paliekame u ir atmetame $\sin(u)$.
- *Rezultatas.* Naudojant genėtą biblioteką ir patikslintą grotelių paiešką algoritmas nuosekliai grąžina kanoninę Fišerio–KPP formą; koeficientų paklaidos neviršija 5 net ir tokiame stambiame tinkle.
- Triukšmingo Fišerio–KPP atvejo rezultatai pateikiami vėlesniame skyriuje.

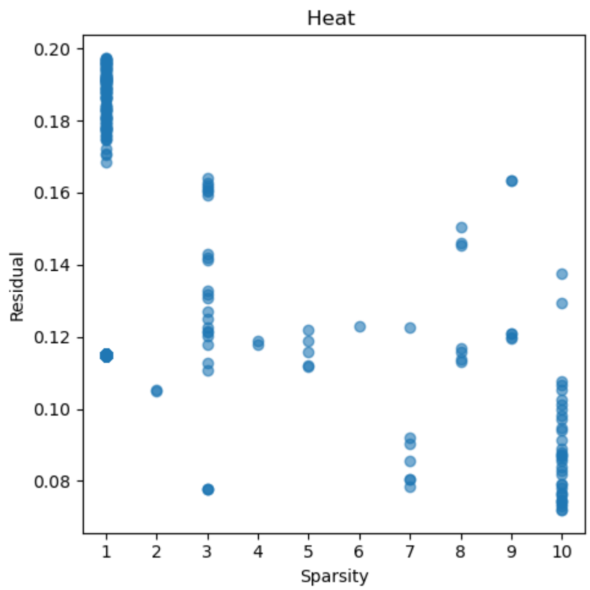

Figure 3: **(a)** Likučio paklaida prieš retumą (10 000 STLSQ bandymų).

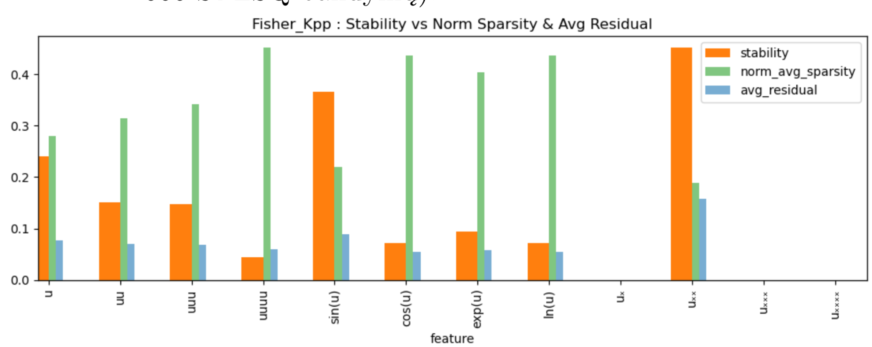

Figure 4: **(b)** Dėmenų stabilumas priklausomai nuo normalizuoto retumo ir vidutinės likučio paklaidos.

Šie pastebėjimai atliepia ankstesnių darbų įspėjimus, jog be fizinių prielaidų nereikėtų be saiko plėsti požymių bibliotekos [3, 6]. Stipriai koreliuoti nepolinomiai dėmenys gali imituoti tikrąją dinamiką ir išlikti po retumo sankcijų, jei slenkstis parenkamas neatsargiai.

4.1.3 Aleno–Kano lygtis

Stambus bazinis atvejis ($N_x = N_t = 16$). 1 lentelėje 4 apibendrintos šešios vyraujančios modelių šeimos, gautos iš 10^4 STLSQ šlavimų 16×16 duomenų aibėje. Pareto „alkūnė" pasirodo ties retumu $s = 4$, kur likučio paklaida mažiausia, tačiau papildomas konstantos dėmuo (c_0) neturi fizinio pagrindo kanoninėje Aleno–Kano lygties formoje $u_t = D\,u_{xx} + \gamma\,(u - u^3)$. Kadangi konstanta pasitaiko retai (stabilumas $\pi_{c_0} < 0.15$) ir tik esant mažiems

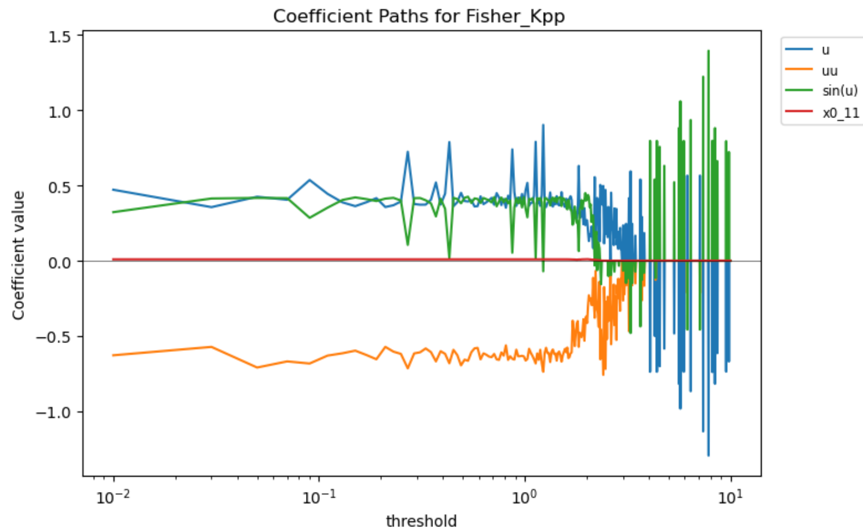

Figure 5: Koeficientų vidutinės vertės per visas optimizatorių realizacijas kintant slenkstinei vertei

slenksčiams ($\lambda < 0.5$), vienas stabilumo–genėjimo ciklas ją pašalina ir atskleidžia teisingą trijų dėmenų struktūrą.

Table 4: Dominuojančios modelių klasės, atgautos 16×16 tinklelyje.

Tipinė RHS	s	r intervalas	λ/α intervalas
$c_1 u - c_3 u^3 + 0.0083\, u_{xx}$	3	0.081–0.103	$0.5 \lesssim \lambda \lesssim 1.8,\ \alpha \lesssim 4 \times 10^{-4}$
$c_0 + c_1 u - c_3 u^3 + 0.0083\, u_{xx}$	4	0.070–0.083	$0.2 \lesssim \lambda \lesssim 0.5,\ 4 \times 10^{-3} \lesssim \alpha \lesssim 8 \times 10^{-3}$
$c_1 u - c_2 u^2 + 0.0082\, u_{xx}$	3	0.081–0.104	didelė λ
$c_0 + c_1 u + c_2 u^2 - c_3 u^3 + 0.009\, u_{xx}$	5	0.082–0.100	$\lambda < 0.2,\ 6 \times 10^{-3} \lesssim \alpha \lesssim 9 \times 10^{-3}$
$c_1 u - c_3 u^3 + 0.009\, u_{xx}$	3	0.100–0.115	labai didelė λ
$0.0083\, u_{xx}\ (+c_0)$	1–2	0.11–0.13	ekstremali λ

Tankus skaičiavimas (64×64, $\Delta t = 0.025$). Keturgubai padidinta erdvinė raiška ir sumažintas laiko žingsnis stabilizuoja koeficientų įverčius ir paverčia konstantos dėmenį statistiškai nereikšmingu. Atliekame 1 000 atsitiktinių šlavimų tiek STLSQ, tiek SSR metodais, naudodami polinominę biblioteką iki 3 laipsnio:

$$\lambda \in [0.01, 10], \quad \alpha \in [10^{-6}, 10^{-2}], \quad (\text{SSR } \kappa \in [10^{-4}, 0.1]).$$

- Pareto frontas vis dar *siūlo* optimalų retumą $s = 4$, tačiau papildomas dėmuo yra c_0 arba u^2; abiejų stabilumas $\pi < 0.05$, vidutinis retumas > 0.7, todėl genėjimo taisyklė juos atmeta ($\pi < \pi_{\min} = 0.4$ arba $\bar{s} > s_{\max}$).

- Po genėjimo visi sprendikliai konverguoja į tris dėmenis turinčią Aleno–Kano lygtį su likučio paklaida $r \approx 0.08$ ir koeficientų paklaida <5 %.
- Patikrinus atsparumą: pridedant iki 3 % Gauso triukšmo prieš diferenciavimą teisingas modelis vis dar atgaunamas, jei naudojamas tankus tinklelis ir Savitskio–Golajaus filtras šiuo atveju netaikomas (filtravimo prireikia triukšmui >3 %).

Pagrindinė pamoka. Stambioje raiškoje vien likučio–retumo „alkūnė“ gali klaidinti; derinant dėmenų stabilumo statistiką su vidutiniu tinklelio patankinimu pakanka nuslopinti klaidingus konstantos ir kvadratinius dėmenis ir, be rankinio derinimo, gauti kanoninę Aleno–Kano lygtį.

4.1.4 Grėjaus–Skoto sistema

Skaitiniai parametrai. Difuzijos koeficientai $D_u = 0.01,\ D_v = 0.005$, tiekimo/šalinimo greičiai $F = 0.04,\ k = 0.06$. Laukai inicijuojami dviem Gauso kalneliais (fragmentas žemiau). Stambiame 64×64 tinkle su $\Delta t = 0.1$ sistema integruojama neišreikštine–išreikštine (IMEX) schema; tiek STLSQ, tiek SR3 atliekama po 10^4 atsitiktinių šlavimų hiperparametrų aibėje $\lambda \in [0.01, 10],\ \alpha \in [10^{-6}, 10^{-2}],\ \nu \in \{0.1, 0.98\}$.

$$\begin{cases} u_t = D_u\, u_{xx} \ - \ u\, v^2 + F\,(1-u), \\ v_t = D_v\, v_{xx} \ + \ u\, v^2 - \big(F + k\big)\, v, \end{cases} \qquad \big(D_u = 0.01,\ D_v = 0.005,\ F = 0.04,\ k = 0.06\big).$$

$$\sigma = 0.075\, L, \qquad c_1 = 0.3\, L, \qquad c_2 = 0.7\, L,$$

$$G(x) = e^{-\frac{(x-c_1)^2}{2\sigma^2}} \ + \ 0.7\, e^{-\frac{(x-c_2)^2}{2\,(1.5\,\sigma)^2}},$$

$$u(x,0) = 1 - G(x), \qquad v(x,0) = G(x).$$

Stambiojo 64×64 tinklelio identifikacija. Po *penkių* dėmenų stabilumo genėjimo ratų kandidatų bibliotekoje lieka $\{1, u, v, uv, v^2, uv^2, uv^3, u_{xx}, v_{xx}\}$, tačiau vien bendrąja logika jos nepavyksta toliau sumažinti. Pareto frontas nepakankamai informatyvus ir leidžia dvi vienodai tikėtinas keturių dėmenų formas:

$$\textbf{A:}\begin{cases} u_t &= c_0 + c_1 v - c_3 uv^2 + D_u u_{xx}, \\ v_t &= c_4 uv^2 + c_5 uv^3 - c_6 v^2 + D_v v_{xx}; \end{cases} \qquad \textbf{B:}\begin{cases} u_t &= -c_3 uv^2 + D_u u_{xx}, \\ v_t &= -c_7 v + c_4 uv^2 + D_v v_{xx}, \end{cases}$$

abiems $r \approx 0.07$. Modelis A turi klaidingą sąryšį uv^3, o modelis B praleidžia tiekimo dėmenį $F(1-u)$.

Tankus 256×256 **tinklelis (**$\Delta t = 0.0125$**).** Atliekant po 1 000 šlavimų kiekvienam sprendikliui ta pačia biblioteka dvi alternatyvos iškart atsiskiria – dėmuo uv^3 neišgyvena pirmojo genėjimo – ir Pareto alkūnė išryškėja (1 pav. 6). STLSQ ir SR3 pateikia *kanoninį* retumo raštą $(s_u, s_v) = (4, 3)$:

$$\begin{cases} u_t = +0.030\,(1-u) - 0.918\,uv^2 \; + \; 0.0096\,u_{xx}, \\ v_t = -0.088\,v + 0.945\,uv^2 \; + \; 0.0049\,v_{xx}. \end{cases}$$

Likučio paklaida $r = 1.44 \times 10^{-2}$, koeficientų paklaida $E_{\mathrm{coef}} \approx 3.8\%$, tikslumas/atkūrimas $= 1$.

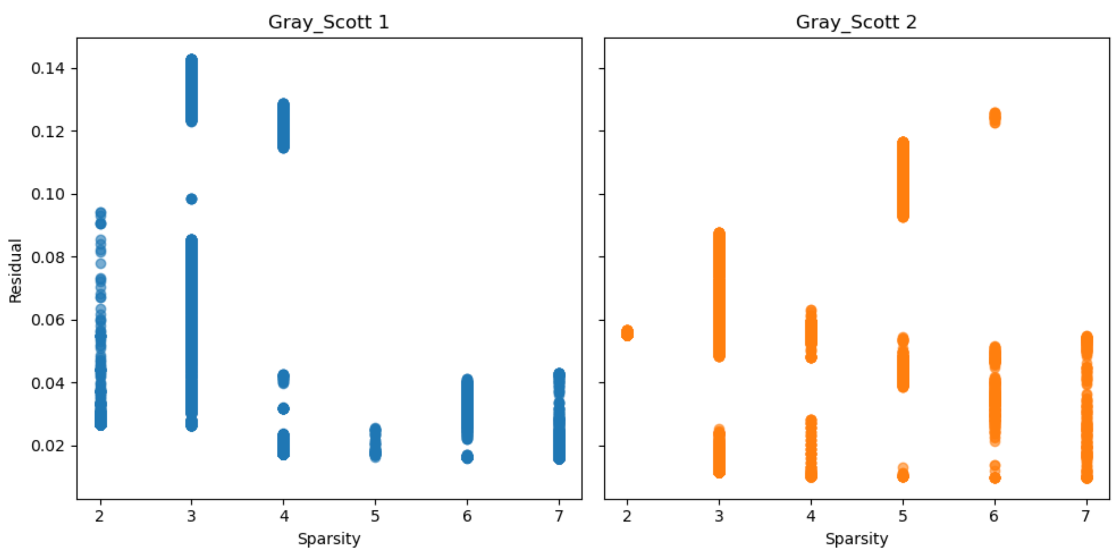

Figure 6: Tankus tinklelis: aiški Pareto minima ties $(s_u, s_v) = (4, 3)$, nė vienas kubinis uv^3 dėmuo neišlieka.

Svarbiausios išvados.

- Stambiame tinklelyje koreliuoti aukštos eilės dėmenys, pvz. uv^3, gali imituoti Grėjaus–Skoto kinetiką ir praeiti pro retumo filtrus, nebent panaudojama papildoma statistika (stabilumas, vidutinis retumas).

- Vienas bibliotekos genėjimo etapas ir vidutinė tinklelio patankinimo priemonė pašalina dviprasmiškumą nekeisdama optimizavimo parametrų.

4.2 Triukšmas

4.2.1 Identifikavimo jautrumas triukšmui

Retoji regresija naudoja taškines baigtinių skirtumų išvestines (3.4.3); todėl net ir nedidelio Gauso triukšmo įnešta paklaida skaitiniame diferenciavime ženkliai sustiprinama. Šį poveikį kiekybiškai įvertiname vienos (Fišerio–KPP) ir dviejų lygčių (Grėjaus–Skoto) sistemoms.

4.2.2 Fišerio–KPP: 0–10 % triukšmo optimizatorių šlavimai

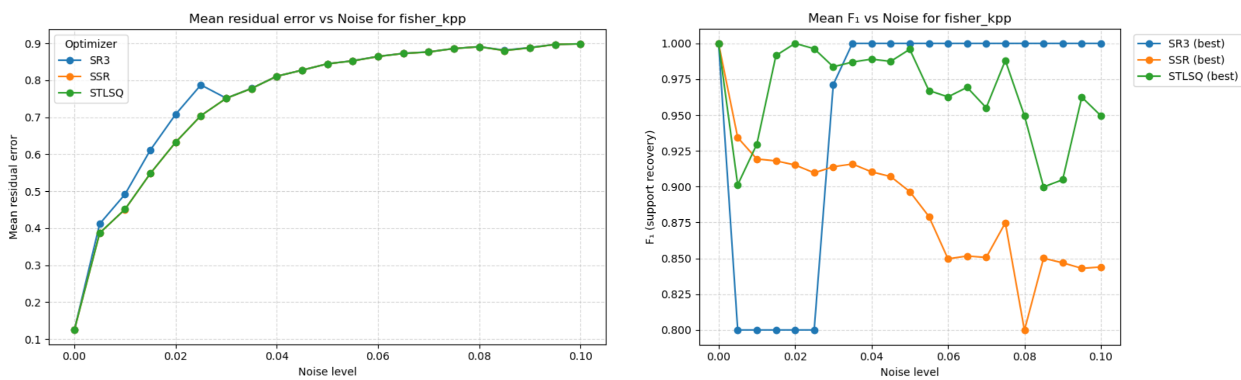

Figure 7: Fišerio–KPP identifikavimas esant adityviniam Gauso triukšmui (1000 įverčių kiekvienam optimizatoriaus taškui grafikuose). Kairėje – vidutinė likučio paklaida; dešinėje – vidutinis dėmenų parinkimo F_1 rodiklis.

Pastebėjimai. 1 pav. 7 vaizduoja likučio paklaidą ir F_1 rodiklį didėjant triukšmui nuo 0 iki 10 %.

- Jau ties 0.5% triukšmu stebimas **staigus likučio šuolis**, rodantis dideles išvestinių skaičiavimo paklaidas.
- Su tiksliu kandidatinių funkcijų rinkiniu (tik $\{1, u, u^2, u_{xx}\}$) teisingas dėmenų rinkinys atstatomas iki maždaug 1 % triukšmo; F1 įvertis prastėja, bet yra veikiamas ir perteklinio prisitaikymo prie duomenų.

- Didinant retumo slenkstį mažėja klaidingai įtrauktų dėmenų, bet kartu atmetami ir tikrieji, tad tikslumo–atkūrimo balansas sparčiai blogėja, kai triukšmas viršija 1 %.

4.2.3 Grėjaus–Skoto modelis: 0.1–0.3 % triukšmas

0.1 % triukšmo. Naudojant 256×256 tinklelį ir Savickio–Golajaus filtravimą (langas 31, polinomų laipsnis 3) beveik pavyksta atkurti kanoninę keturių–/trijų–dėmenų struktūrą; vis dėlto įsiskverbia menkos difuzijos liekanos (u_{xx} v_t lygtyje ir atvirkščiai) su $|\hat{c}| \approx 10^{-3}$.

0.2 % triukšmo. Kryžminės difuzijos dėmenys išlieka, atsiranda papildomų klaidingų reakcijų, minimali likučio paklaida didėja $r > 0.6$, tikrasis signalas vos atskiriamas.

0.3 % triukšmo. Identifikavimas žlunga. Visi optimizatoriai grąžina penkių–šešių dėmenų modelius, kuriuose dominuoja konstantos, tiesiniai ir kvadratiniai dėmenys (likutis $r \gtrsim 0.9$). Vienintelis būdas mažinti likučio paklaidą tampa parinkti skaitinei simuliacijai tankesnį tinklelį, tačiau net ir vienos dimensijos atveju skaičiavimai pasidaro brangūs. Net ir tikslus kandidatinis rinkinys nebesugeba atkurti tikrųjų koeficientų, galima tvirtinti, kad baigtinių skirtumų išvestinės per stipriai paveiktos triukšmo.

Pasekmės ir silpnųjų integralinių formų poreikis Šie eksperimentai patvirtina, kad išvestinėmis grindžiami SINDy metodai greitai tampa nepatikimi prie nedidelių triukšmo lygių $\gtrsim 1.5\%$ difuzijos sistemoms ir $\gtrsim 0.1\%$ sudėtingesnėms, šiuo atveju reakcijos-difuzijos sistemoms. Pagrindinė priežastis – blogai apibrėžta skaitinio diferencijavimo užduotis, kuri sustiprina aukšto dažnio triukšmą. Silpnoji integralinė forma atsisako taškinių išvestinių, uždavinio sprendimas įgyvendinamas integraline prasme ir taip suteikia keliais laipsniais didesnį triukšmo toleravimą. Reinbold *ir kt.* parodė, kad reakcijos–difuzijos lygtį galima atkurti esant net $\sim 30\%$ triukšmui, tuo tarpu įprastos formos identifikavimas žlunga jau virš 1% [6]. Silpnosios integralinės formos įtraukimas yra aiškiausias kelias stiprinti mūsų identifikavimo struktūros tvirtumą ir atsparumą triukšmui.

4.2.4 Triukšmingų duomenų švelninimo strategijos

Baigtinių skirtumų „siaurasis kaklelis“. Standartinės SINDy atmainos erdvines ir laiko išvestines skaičiuoja baigtinių skirtumų būdu; net ir nedidelis Gauso triukšmas ženkliai sustiprinamas, kas iššaukia likučio šuolius ir klaidingų dėmenų patekimą į modelį [2, 4]. Mūsų bandymai tai patvirtina: vos prie 0.1% triukšmo likučio paklaida visose penkiose bazinėse sistemose padidėja maždaug dešimteriopai, o F_1 krinta žemiau 0.8.

Paprastos priemonės. detect-weight=true

- **Savickio–Golajaus gludinimas** prieš diferenciaciją (S-G filtrui parenkami–lango dydis 9–17, polinomo laipsnis 3 Grėjaus–Skoto atveju) sumažina aukšto dažnio svyravimus ir atkuria F1 $\approx$ 0.9 iki maždaug 3 % triukšmo—tai atitinka filtruoto diferenciavimo praktiką, rekomenduotą Champion *et al.* [18].
- **Slenksčio pertvarkymas.** Padidinus retumo slenkstį λ sumažėja klaidingai įtrauktų dėmenų, tačiau neišvengiamai nukenčia *recall*;.

Integralioji / silpnoji SINDy forma. Daug atsparesnė alternatyva – *perkelti diferenciavimą ant žinomų, glodžių testinių funkcijų* dauginant kandidatines lygčių formas iš branduolio ir integruojant dalimis. Tokia silpnoji forma triukšmą vidurkina, o ne stiprina [6]. Reinbold *ir kt.* parodė, kad silpnoji forma sėkmingai atkuria reakcijos–difuzijos dėsnį, kai duomenys teršiami 5–10 % triukšmu—tokioje srityje klasikinė stiprioji SINDy jau žlunga. Metodas iš esmės pakeičia taškinį likučio minimizavimą integruotais (kontrolinio tūrio) likučiais, drastiškai pagerindamas atsparumą triukšmui.

Tolimesnė kryptis mūsų identifikavimo eigai. Aukščiau pateikti pastebėjimai rodo, kad vien filtravimo ir slenksčių derinimo *nepakanka*, kai triukšmas viršija kelis procentus. Ateityje į identifikavimo sistemą vertėtų įtraukti silpnos formos metodiką, kuri kartu su ansambliniu vidurkinimu ir tikslesniu skaitiniu vertinimu turėtų išplėsti patikimo identifikavimo ribas gerokai virš šiame darbe pastebėtų triukšmo lubų (Pagal panašias išvadas [3, 6, 18]).

4.3 Kandidatinio rinkinio kompleksiškumas

Dinamikos atradimo sėkmė tiesiogiai priklauso nuo *kandidatinių funkcijų rinkinio* $\Theta(\mathbf{u})$. Jei Θ per maža, nebeįmanoma aprašyti duomenyse vyraujančios dinamikos; jei per didelė, atsiranda koreliuotų pakaitinių dėmenų, kurie gali tiek pat kiek tikrieji dėmenys patekti į randamą modelį. Abu kraštutinius atvejus stebėjome Grėjaus–Skoto sistemoje (64×64 tinklelis, 10^4 STLSQ šlavimų); kokybiškai tie patys dėsningumai pasitvirtino ir vieno komponento testuose.

4.3.1 Minimalus ir išplėstas rinkiniai

Galima išbandyti du polinominius rinkinius:

$$L = \{1,\, u,\, v,\, u^2,\, v^2,\, uv,\, u_{xx},\, v_{xx}\},$$

$$L \cup \{u^3,\, v^3,\, u^2v,\, uv^2,\, u^4,\, u^3v,\, u^2v^2,\, uv^3,\, v^4\}.$$

Grėjaus–Skoto sąveikos dėmuo uv^2 *nėra* įtrauktas į pirmąjį, bet yra Θ_{ext}. Abiem rinkiniais atlikome po 10^4 STLSQ šlavimų tame pačiame $64{\times}64$.

Rezultatai.

- *Minimalus rinkinys (Θ_{min}).* Kadangi uv^2 nėra, SINDy kompensuoja didindamas konstantinį įtekėjimo (c_0) ir linearų v dėmenis. Geriausios paklaidos „užstringa“ ties $\bar{r} \approx 4.6 \times 10^{-2}$, apie tris kartus blogiau nei su pilnu modeliu, o tikrosios fizikos recall krenta iki ≈ 0.75.
- *Išplėstas rinkinys* Įtraukus uv^2, recall pakyla iki 1.0, mažiausia likučio paklaida sumažėja iki 1.4×10^{-2} (plg. 4.1.4). Precision kiek smunka: žemoms slenkstinėms vertėms pro filtrą praeina klaidingas uv^3 dėmuo. Jo žymuo akivaizdus— $\pi_{uv^3} < 0.05$, $\bar{s}_{uv^3} > 7$ ir $\bar{r}_{uv^3}$ beveik negerėja—todėl vienas stabilumo pagrįstas genėjimo ciklas jį pašalina ir sugrąžina kanoninę Grėjaus–Skoto struktūrą.
- Panašiai iš kandidatinio rinkinio pašalinus pačius difuzijos dėmenis, nė vienas optimizatorius neranda reto ir pritaikyto modelio.

Svarbiausia išvada. Per maža biblioteka praleidžia esminę fiziką; per didelė įsileidžia pakaitinius dėmenis, kurie imituoja tikrąją dinamiką. Stebint π_j, $\bar{s}_j$ ir $\bar{r}_j$ cikle „fit $\to$ rank $\to$ prune“ gaunamas objektyvus ir lengvas standartas, kuris padeda išlaikyti visus tikruosius dėmenis ir pašalinti klaidingus mažiau nei per du iteracinius rinkinio mažinimo žingsnius— tai atitinka bendras rekomendacijas [3, 6].

4.4 Rezultatų santrauka

Bendras identifikavimo tikslumas. Triukšmo nepaveiktuose baziniuose eksperimentuose sukonstruota identifikavimo eiga atgavo visus tikruosius dėmenis. Tai patvirtina, kad esant paprasčiausiais atvejais net ir pakankamai grubiai erdvės–laiko raiškai ir tinkamai atlikus kandidatinių funkcijų rinkinio, optimizavimo hiperparametrų intervalų bei Pareto analizės dalis, naudojantis pasiūlyta darbo eiga galima sėkmingai atrasti valdančiąsias DLDI įvairioms reakcijos–difuzijos sistemoms [2].

Matavimo triukšmo poveikis. Gauso triukšmo pridėjimas atskleidė išvestinėmis grįstų metodų „Achilo kulną“. Vienos komponentės uždaviniuose sistema paprastai tampa sunkiai identifikuojama esant vos 1%-3% triukšmui, o daugiakomponentė Grėjaus–Skoto sistema tapo neatkuriama viršijus vos 0.3 % triukšmą. Tai atitinka ankstesnes išvadas, kad baigtinių skirtumų gradientai sustiprina aukšto dažnio paklaidas ir destabilizuoja retumo filtrą [6]. Savickio–Golajaus filtravimas šiek tiek praplėtė naudotiną ribą, bet esminiam patobulinimui būtina pereiti prie silpnųjų integralinių formų (pvz. weak-SINDy ar RK4-SINDy), kurios vengia taškinių išvestinių ir pasižymi daug didesniu atsparumu triukšmui [18].

Kandidatinio rinkinio sudėtingumo poveikis. Per maža biblioteka praleidžia svarbius dėmenis (blogėja recall), o per didelė – padidina klaidingai teigiamus atvejus (blogėja precision) ir skatina perteklinį priderinimą. Empirinė dėmenų stabilumo statistika pasirodė veiksmingiausia išeitis: dėmenys su maža pasirodymo tikimybe ($\pi_j < \pi_{\min}$) optimizavimo šlavimų įgyvendinimuose arba dažniausiai atsirandantys didelio retumo / didelio likučio modeliuose buvo pirmiausia pašalinti; vidutinis retumas ir likutis nulėmė antrą genėjimo ciklą. Šis dviejų etapų genėjimas paprastais difuzijos atvejais uždavinį leido išspręsti vos atlikus šį rinkinio mažinimą, o reakcijos-difuzijos atveju buvo kertinis žingsnis ieškant Pareto optimalaus įverčio ieškomoms sistemoms.

Reikšmė identifikavimo eigos struktūrai. Rezultatai atsako į 1.2 skirsnyje keliamus klausimus:

- **(bazinis identifikavimas).** Visos etaloninės DLDI atkurtos tam tikromis sąlygomis.
- **(atsparumas triukšmui).** Identifikavimas nuspėjamai blogėja didėjant triukšmui; silpnosios formos yra logiška tolesnė kryptis.
- **(hiperparametrų ir kandidatinių rinkinių vertinimas).** Dėmenų stabilumu grįstas genėjimas būtinas valdant retumo–likučio bei recall–precision kompromisus dideliuose rinkiniuose.

Bendrai, patikimiausias automatinio DLDI atradimo triukšmingiems duomenims kelias apima *(i)* vidutinio tankio tinklelio (t. y. pakankamos raiškos) naudojimas, *(ii)* dėmenų stabilumu grįstas rinkinio genėjimas ir *(iii)* Pareto analize, triukšmo įtakos mažinimo sprendimais pagrįstą modelio atrinkimą.

5 Apibendrinimas

Šiame skyriuje atsitraukiame nuo 4 skyriaus techninių detalių ir pažvelgiame į „didelį paveikslą“: *kokią* prasmę turi rezultatai, *kodėl* jie svarbūs duomenimis grindžiamam DLDI atradimui ir *kur* dabartinė grandinė vis dar neatitinka lūkesčių.

5.1 Rezultatų interpretacija

Nuo difuzijos iki raštų formavimosi. SINDy be vargo atkūrė vadovėlinę šilumos lygtį net ir itin stambiame 16×16 tinklelyje—tai nestebina, žinant ankstesnius darbus, kur valdančioji lygtis atrandama naudojantis atsitiktinio klaidžiojimo eksperimento duomenimis [2]. Netiesinėse sistemose situacija tampa sudėtingesnė: teisingai identifikuoti logistinio augimo ($u - u^2$) ir Grėjaus–Skoto sąveikos (uv^2) dėmenys rodo, kad turint raiškią biblioteką reta regresija pajėgi atpažinti susietus reakcijos–difuzijos dėmenis sudėtinguose režimuose. Mūsų Grėjaus–Skoto identifikavimui užteko mažiau tinklelio taškų nei ankstesniame reakcijos difuzijos sistemos PDE-FIND pavyzdyje (64×64 vs. 128×128 [1]), pabrėžiant čia pristatyto dėmenų stabilumu grįsto genėjimo vertę.

5.2 Pasekmės DLDI atpažinimui

- **Duomenų kokybė svarbiausia.** Vos 1 % Gauso triukšmo sužlugdė Grėjaus–Skoto identifikavimą, jei nebuvo taikoma silpnoji forma ar stiprus filtravimas— tai atitinka literatūros įspėjimus [6]. Sprendinio lauko eksperimentai privalo naudoti mažatriukšmius jutiklius arba kaupti perteklinius matavimus vidurkinimui.
- **Kandidatiniai rinkiniai gali būti subtilūs.** Per siauras rinkinys atveda prie nepakankamai retų arba netikslių įverčių. Dviejų etapų genėjimas (stabilumas → retumas) pasirodė praktiškas kompromisas ir galėtų būti pilnai automatizuotas ateities įrankiuose.
- **Aiškios formos atsiperka.** Skirtingai nei neuroniniai PDE pakaitalai, SINDy pateikia uždaros formos lygčių sistemas, kurios iškart tinkamos analizei ar valdymo sintezei—tai lemiamas privalumas inžineriniuose taikymuose.

5.3 Apribojimai

- **Tik sintetiniai duomenys.** Visi etalonai buvo simuliuoti; realūs matavimai dažnai turi koreliuotą, ne Gauso triukšmą.
- **Nedidelis sudėtingumas.** Eksperimentai vienmačiai, daugiausia su dviem kintamaisiais. Aukštesnių matmenų ar chaotiniai srautai (pvz. Navjė–Stoukso turbulencija) lieka neapžvelgti.
- **Baigtiniai skirtumai ir slenksčių derinimas.** Taškinės išvestinės ypač stiprina triukšmą, o slenkstis λ tebepasirenkamas šlavimo būdu; automatinės ar Bajeso analize informuota atranka galėtų būti geras indėlis į gero identifikavimo įrankio sukūrimą.

- **Rinkinio prielaidų poreikis.** Šiame darbe rinkiniai buvo sudaryti remiantis srities žiniomis; visiškai nežinomos fizikos atvejais gali tekti daug resursų paskirti būtent funkcijų, kurios potencialiai galėtų aprašyti dinamiką, įvedimui.

5.4 Palyginimas su alternatyviais metodais

Silpnos formos atmainos, tokios kaip integralinė SINDy, STRIDE ar RK4-SINDy, taškinių išvestinių nenaudoja ir demonstruoja atsparumo triukšmui didėjimą bene dešimt kartų dydžio rėžiais (iki 20 % triukšmo 2-D Grėjaus–Skoto sistemoje [18]). Physics-informed neural nets (PINN) mokosi rezidualų neišreikštiniu būdu, tačiau reikalauja daug skaičiavimo ir pateikia sunkiai interpretuojamas parametrizacijas. Retoji regresija išlieka lengva ir skaidri; pritaikius silpnos formos idėjas būtų galima sujungti šiuos privalumus su modernių PINN/STRIDE triukšmo atsparumu.

5.5 Svarbiausi

- *pasiūlyta dinamikos atpažinimo eiga veikia—bet tik tada, kai duomenys pakankamai švarūs, o rinkinys kruopščiai kuruojamas.*
- Dėmenų stabilumu grįstas genėjimas – paprastas, automatizuojamas būdas suvaldyti dideles bibliotekas neprarandant teisingų dėmenų.
- Silpnos formos ir tikimybinės plėtiniai – logiški žingsniai kovojant su eksperimentiniu triukšmu ir modelio neapibrėžtumu.

6 Išvados ir būsimi darbai

6.1 Išvados

Šiame darbe pritaikėme retosios identifikacijos (SINDy) karkasą, kad atkurtume valdančiąsias DLDI įvairiems difuzijos ir reakcijos–difuzijos etalonams. Triukšmo neturinčiuose duomenyse visi kanoniniai dėmenys buvo identifikuoti su F_1 rodikliais > 0.9. Sistemingi bandymai su adityviniu Gauso triukšmu ir kandidatinių funkcijų nagrinėjimu kiekybiškai nustatė tam tikras praktines modelių atpažinimo ribas bei parodė dviejų etapų *stabilumas* $\to$ *retumas* genėjimo metodo privalumus. Rezultatai įrodo, kad—turint kuruotą biblioteką ir vidutiniškai švarius duomenis—SINDy suteikia interpretacijai patogų ir skaičiavimams lengvą kelią lygčių atradimui, taip įvykdydamas **??** skirsnyje iškeltus tikslus.

6.2 Būsimi darbai

- **Eksperimentiniai ir aukštesnių matmenų duomenys.** Grandinės išplėtimas realiems laboratoriniams matavimams (arba 2-D/3-D simuliacijoms) patikrins jos mastelį ir triukšmo toleranciją tikroviškomis sąlygomis.
- **Pažangus triukšmo valdymas.** Integruoti silpnos formos / integralines SINDy ar STRIDE atmainas, kurios taškines išvestines pakeičia erdvės-laiko vidurkinimu ir Gray–Scott sistemose toleruoja $\sim$ 20% triukšmą [18]; Bajeso analize patobulintas optimizatorių hiperparametrų rinkimas bei dėmenų vertinimas galėtų papildomai patobulinti rezultatus.
- **Automatinis rinkinio konstravimas.** Automatizuoti būdai tikslinti kandidatines funkcijas leistų dinamiškai generuoti ir genėti dėmenis, sumažinant rankinį bandymų skaičių ir apsaugant nuo perteklinio priderinimo prie duomenų, kai stokojama tyrinėjamos srities žinių.
- **Dalinis stebimumas.** Daugelyje eksperimentų stebima tik dalis kintamųjų ar erdvinių taškų. SINDy sujungimas su būsenos rekonstrukcijos metodais, galintis kompensuoti trūkstamus duomenis, lieka atviru ir praktiškai svarbiu uždaviniu.

Baigiamoji pastaba. Duomenimis grindžiamas lygčių atradimas žada iš esmės pakeisti modeliavimo procesą moksle ir inžinerijoje. Patvirtinęs SINDy veiksmingumą, struktūrizuotą būdą atlikti identifikavimo darbus įvairioms DLDI ir nubrėžęs jo veikimo ribas, šis darbas sukuria tvirtą pagrindą ir aiškų veiksmų planą ateities, triukšmui atsparių ir visiškai automatizuotų PDE identifikavimo grandinių plėtrai.

Darbui parašytas programas rasti galima rasti: https://github.com/Io-1/PDE-Find

Literatūra

[1] S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proc. Natl. Acad. Sci. U.S.A. 15, 3932–3937 (2016).

[2] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Data-driven discovery of partial differential equations*, Sci. Adv. 3(4), e1602614 (2017).

[3] N. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, *Implicit sparse identification of nonlinear dynamics with rational functions*, arXiv:1605.08368 (2017).

[4] H. Schaeffer, *Learning partial differential equations via data discovery and sparse optimization*, Proc. R. Soc. A 473(2197), 20170388 (2017).

[5] M. Schmidt and H. Lipson, *Distilling free-form natural laws from experimental data*, Science **324**, 81–85 (2009).

[6] P. A. K. Reinbold, D. R. Gurevich, and R. O. Grigoriev, *Using noisy or incomplete data to discover models of spatiotemporal dynamics*, Phys. Rev. E **101**(1), 010203 (2020).

[7] R. A. Fisher, *The wave of advance of advantageous genes*, Ann. Eugen. **7**, 355–369 (1937).

[8] S. M. Allen and J. W. Cahn, *A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening*, Acta Metall. **27**, 1085–1095 (1979).

[9] P. Gray and S. K. Scott, *Autocatalytic reactions in the isothermal, continuous stirred tank reactor: isolating and characterising the pattern–forming system*, Chem. Eng. Sci. **39**, 1087–1097 (1984).

[10] J. E. Pearson, *Complex patterns in a simple system*, Science **261**, 189–192 (1993).

[11] I. Prigogine and R. Lefever, *Symmetry breaking instabilities in dissipative systems*, J. Chem. Phys. **48**, 1695–1700 (1968).

[12] J. J. Tyson, *Oscillations and waves in an autocatalytic reaction system*, J. Chem. Phys. **58**, 3919–3925 (1973).

[13] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*, Society for Industrial and Applied Mathematics (2007).

[14] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio–Temporal Domains*, John Wiley & Sons, Chichester (2013).

[15] L. Boninsegna, F. Nüske, and C. Clementi, *Sparse learning of stochastic dynamical equations*, J. Chem. Phys. **148**(24), 241723 (2018).

[16] P. Zheng, K. Champion, Z. Qin, W. Pan, A. Y. Aravkin, S. L. Brunton, and J. N. Kutz, *A unified framework for sparse relaxed regularized regression: SR3*, IEEE Access **7**, 1404–1423 (2018).

[17] K. Champion, P. Zheng, A. Y. Aravkin, S. L. Brunton, and J. N. Kutz, *A unified sparse optimization framework to learn parsimonious physics–informed models from data*, IEEE Access **8**, 169259–169271 (2020).

[18] K. Champion, P. Zheng, S. L. Brunton, and J. N. Kutz, *Discovery of nonlinear dynamical systems using a Runge–Kutta sparsification*, Proc. R. Soc. A **478**(2260), 20210883 (2022).