ISSAI NAZARBAYEV UNIVERSITY | Institute of Smart Systems and Artificial Intelligence

NAZARBAYEV UNIVERSITY

SCHOOL OF COMPUTING AND ARTIFICIAL INTELLIGENCE

NU IMPACT FOUNDATION

BOOK OF ABSTRACTS

1ST INTERNATIONAL CONFERENCE AND WORKSHOP ON MATHEMATICS AND ARTIFICIAL INTELLIGENCE

Nazarbayev University, Astana

September 9-11, 2026

NAZARBAYEV UNIVERSITY

SCHOOL OF COMPUTING AND ARTIFICIAL INTELLIGENCE

ISSAI NAZARBAYEV UNIVERSITY | Institute of Smart Systems and Artificial Intelligence

НАЗАРБАЕВ УНИВЕРСИТЕТІ
ЕСЕПТЕУ ЖӘНЕ ЖАСАНДЫ ИНТЕЛЛЕКТ МЕКТЕБІ
АҚЫЛДЫ ЖҮЙЕЛЕР ЖӘНЕ ЖАСАНДЫ ИНТЕЛЛЕКТ ИНСТИТУТЫ

НАЗАРБАЕВ УНИВЕРСИТЕТ
ШКОЛА ВЫЧИСЛИТЕЛЬНЫХ НАУК И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
ИНСТИТУТ УМНЫХ СИСТЕМ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

NAZARBAYEV UNIVERSITY
SCHOOL OF COMPUTING AND ARTIFICIAL INTELLIGENCE
INSTITUTE OF SMART SYSTEMS AND ARTIFICIAL INTELLIGENCE

«МАТЕМАТИКА ЖӘНЕ ЖАСАНДЫ ИНТЕЛЛЕКТ БОЙЫНША 1-ХАЛЫҚАРАЛЫҚ КОНФЕРЕНЦИЯ ЖӘНЕ СЕМИНАР»

баяндамалар тезистерінің жинағы

9–11 қыркүйек 2026 ж.

«1-Я МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ И СЕМИНАР ПО МАТЕМАТИКЕ И ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ»

сборник тезисов докладов

9–11 сентября 2026 г.

«1ST INTERNATIONAL CONFERENCE AND WORKSHOP ON MATHEMATICS AND ARTIFICIAL INTELLIGENCE»

Book of Abstracts

September 9–11, 2026

АСТАНА 2026

ӘОЖ / УДК 51+004.8
КБЖ / ББК 22.1+32.813
М29

Редакция алқасы: Д. Сураган, А. Варол, А. Кашкынбаев, Б. Сабитбек, Т. Эдгу, М. Садыбеков, Е. Абсалямов, М. А. Зайгум
Жауапты хатшылар: М. Манарбек, У. Алибек

Редакционная коллегия: Д. Сураган, А. Варол, А. Кашкынбаев, Б. Сабитбек, Т. Эдгу, М. Садыбеков, Е. Абсалямов, М. А. Зайгум
Ответственные секретари: М. Манарбек, У. Алибек

Editorial board: D. Suragan, A. Varol, A. Kashkynbayev, B. Sabitbek, T. Edgu, M. Sadybekov, Y. Absalyamov, M. A. Zaighum
Secretaries: M. Manarbek, U. Alibek

М29 **Математика және жасанды интеллект бойынша 1-халықаралық конференция және семинар** = 1-я Международная конференция и семинар по математике и искусственному интеллекту = 1st International Conference and Workshop on Mathematics and Artificial Intelligence: баяндамалар тезистерінің жинағы = сборник тезисов докладов = Book of Abstracts (9–11 қыркүйек 2026 ж.). – Астана: Назарбаев Университеті, 2026. – 72 б. – Ағылш.

ISBN 978-601-08-6470-2

Жинақта 2026 жылғы 9–11 қыркүйекте Назарбаев Университетінде (Астана, Қазақстан) өткен Математика және жасанды интеллект бойынша 1-халықаралық конференция мен семинарда ұсынылған ғылыми баяндамалардың тезистері берілген. Жинаққа функционалдық талдау, функциялық кеңістіктер және операторлар теориясы, дифференциалдық теңдеулер және кері есептер, комбинаторика, математикалық логика, оңтайландыру, сондай-ақ машиналық оқыту және үлкен тілдік модельдер бойынша жұмыстар енгізілген. Жинақ ғалымдарға, PhD докторанттарға, магистранттарға, сондай-ақ математика мен жасанды интеллект салаларындағы мамандарға арналған.

В сборнике представлены тезисы научных докладов 1-й Международной конференции и семинара по математике и искусственному интеллекту, состоявшихся 9–11 сентября 2026 года в Назарбаев Университете (Астана, Казахстан). В сборник включены работы по функциональному анализу, функциональным пространствам и теории операторов, дифференциальным уравнениям и обратным задачам, комбинаторике, математической логике, оптимизации, а также машинному обучению и большим языковым моделям. Сборник предназначен для учёных, PhD-докторантов, магистрантов и специалистов в области математики и искусственного интеллекта.

The book contains the abstracts of scientific contributions presented at the 1st International Conference and Workshop on Mathematics and Artificial Intelligence, held on September 9–11, 2026, at Nazarbayev University (Astana, Kazakhstan). It includes works on functional analysis, function spaces and operator theory, differential equations and inverse problems, combinatorics, mathematical logic, optimization, as well as machine learning and large language models. The book is intended for researchers, PhD students, master's students and specialists in mathematics and artificial intelligence.

Тезистер авторлардың редакциясында жарияланады. Материалдардың мазмұнына авторлар жауапты.
Тезисы публикуются в авторской редакции. Ответственность за содержание материалов несут авторы.
The abstracts are published as submitted; the authors are responsible for their content.

ӘОЖ / УДК 51+004.8
КБЖ / ББК 22.1+32.813

ISBN 978-601-08-6470-2

© Назарбаев Университеті, 2026
© Авторлар, 2026

Content

ABSTRACTS **6**

Abek A.N., Bokayev N.A., Gogatishvili A, Ünver T Optimal embedding of cone of a generalized fractional maximal function modeled on the Lorentz Spaces into rearrangement invariant spaces . . . 7

Aliyev M. Test-time budgets are a knapsack, not a water-filling problem . . . 9

Assylbekov Y.M. The Geodesic Ray Transform with Finite Measurements: Stability, Reconstruction, and DeepONets . . . 10

Babayeva G. A packing ceiling for colour-vision-deficiency-safe categorical palettes . . . 10

Baktybergen K., Sobolev B. Categorizing factors influencing hip-fracture care standards for causal diagrams . . . 11

Bassarov S.Zh., Tleukhanova N.T. A recovery operator for Haar wavelet spaces . . . 12

Botaibekov A., Askarov A. Uncertainty-Aware Evaluation of Local OCR-VLMs for Russian and Kazakh Financial Document Recognition . . . 13

Bukanay N.U., Assanova A.T., Mirzakulova A.E. Asymptotics of a Doubly Singularly Perturbed Third-Order BVP with Piecewise Constant Argument . . . 13

Esfahani A., Zhardemova N. Local Well-posedness of the Boussinesq Equation with Linear Restoring Force . . . 15

Fazyl Z.G., Markhabatov N.D. Stability Properties of Pseudofinite Abelian Groups . . . 16

Garibli N. AZ-Eval: negative transfer from Kazakh fine-tuning to Azerbaijani is largely orthographic . . . 18

Garibli N. What does a certified self-consistent answer certify? . . . 19

Gogatishvili A, Bokayev N.A., Kuzeubayeva N.K., Ünver T On the characterization of discrete weighted Hardy inequalities of non-increasing sequences . . . 21

Gogatishvili A, Karshygina G, Ünver T Pointwise and Integral Majorants on Cones of Monotone Functions . . . 23

Huseynli F.S. A Fuzzy TOPSIS-Based Intelligent Decision Support Architecture for Human Resource Management in Banking . . . 25

Huseynov E. How many items does a benchmark need when its items come in clusters? . . . 27

Jubatkanov K., Markhabatov N., Tussupov J., Beisenbayeva K. Logical Adversarial Networks Based on Ehrenfeucht–Fraïssé–Taimanov Games . . . 28

Jumadildayev M.A. Enumeration of perfect matchings in line graphs . . . 29

Kairolla M. Mamba state-space model for sequential WiFi indoor localization . . . 30

Kalidolday A.K., Nursultanov E.D. Interpolation theorem in discrete net spaces $n_{p,q}(F)$ . . . 31

Kim A. Fourier-Mode Allocation and Physical Error in a One-Dimensional Heat Problem 33

Kingston P., Fagnoni T. Proof-Carrying Streamed Workflow Graphs . . . 34

Kosmakova M.T., Akhmanova D.M. Explicit solution and well-posedness for a fractional PDE with a loaded term along a moving curve . . . 35

Kuanysh S.K. Numerical Solution of a Nonlocal Problem for a Fredholm-Type Parabolic Integro-Differential Equation . . . 36

Kuanyshov N., Yeginbay I. Computational methods for the simplicial geometric category . 37

Manarbek M., Tleukhanova N.T. Convolution-type Operators in Grand Lorentz Sequence Spaces . . . 39

Molybaikyzy A. Multipoint boundary value problems for impulsive hyperbolic equations with piecewise constant arguments . . . 39

Mosavi A., Igenbayev A. Generative Artificial Intelligence for Strategic University Leadership . . . 41

Mosavi A., Sabdenaliyev B. LMMs for Higher Education Scientific Strategy Planning and Evaluation . . . 42

Mukhambetzhan M.A., Turgumbayev M.Zh., Kenzhebekova G.S. On the weighted integrability of sums of series with respect to the Price multiplicative system with GM and bounded variation coefficients . . . 43
Muratbekov M., Jarbulova K., Abylayeva A. Spectral Estimates for a Mixed Type Differential Operator . . . 44
Murzabekova G.Y. Exhauster-Based Nonsmooth Analysis for Optimization Problems in Machine Learning . . . 45
Omarov M., Ramazanov M., Khalmetov E. Weak Well-Posedness and Perturbation Analysis for a Space-Dependent Variable-Order Fractional Diffusion Equation . . . 49
Orazbayev A., Zaighum M.A Generalized potential operators in weighted grand Herz spaces 50
Orumbayeva N.T., Zhantassova B.B. On the Solution of a Periodic Boundary Value Problem for a Hyperbolic Equation with a Riemann–Liouville Fractional Derivative with Respect to the Spatial Variable . . . 50
Ospanov K.N., Yeskabylova Zh.B. Coercive solvability conditions for a nonlinear third-order singular differential equation in Lebesgue space . . . 52
Sadybekov M., Dildabek G. On the inverse problem for a degenerate parabolic equation with the Hadamard–Caputo derivative under Samarskii–Ionkin conditions . . . 52
Satybaldina A.N. Direct and Inverse Problems for One-Dimensional Transient Heat Conduction in Multilayered Media . . . 54
Tulenbaev K.M., Koshzhanov R.A., Zharykov M.N., Kettebekov Z.K., Kovalenko A.A. Automatic horse recognition in images using convolutional neural networks . . . 55
Tursungaliyev N.N., Fazyl G.G., Markhabatov N.D. Topological properties of products of finite groups . . . 56
Ussipbayev A.A., Mukhamedina A.I. Spectrally Stable Graph Neural Networks for Railway Delay Propagation . . . 57
Verbovskiy V.V. Context-Dependent Structures of Signs and a Mathematical Model of Meaning . . . 59
Yeshkeyev A.R., Issayeva A.K., Popova N.V. The fractal Jonsson theory and online learning 60
Zaighum M.A. On Besov's Problem . . . 62
Zakiyeva N., Otemissov A. Balance constrained sparse group lasso for network time series 63
Zeinulla A. Combinatorial State-Space and Empirical Game Tree Complexity of Togyzkumalak: A Billion-Game Analysis . . . 64
Zeynalov I.B. Capture ceilings for budget-constrained early warning in education . . . 66
Zeynalov I.B. Decomposing prompt-format sensitivity: extraction artifacts, reasoning suppression, and surface form . . . 67
Zeynalov I.B. From score reliability to decision reproducibility in AI-based educational early warning . . . 68
Zibatkaliyev N. Toward Verified Intelligence: A Survey of AI for Mathematics . . . 70
Index of Names . . . 70

ABSTRACTS

Optimal embedding of cone of a generalized fractional maximal function modeled on the Lorentz Spaces into rearrangement invariant spaces

A.N. ABEK[1], N.A. BOKAYEV[2], A GOGATISHVILI[3], T ÜNVER[4]

[1,2] *L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan*
[3]*Institute of Mathematics, Czech Academy of Sciences, Praha, Czech Republic*
[4]*Department of Mathematics, Kirikkale University, Kirikkale, Türkiye*

[1]azhar.abekova@gmail.com, [2]bokayev2011@yandex.ru, [3]gogatish@math.cas.cz, [4]tugceunver@kku.edu.tr

Let $E \equiv E(R^n)$ be a rearrangement-invariant space (RIS; we use here the axioms developed in the book by Bennett and Sharpley [1]). $E' \equiv E'(R^n)$ the associated rearrangement-invariant space and $\widetilde{E} \equiv \widetilde{E}(R_+)$, $\widetilde{E}' \equiv \widetilde{E}'(R_+)$ their Luxembourg representation, i.s. such an RIS is

$$\|f\|_E = \|f^*\|_{\widetilde{E}}, \quad \|g\|_{E'} = \|g^*\|_{\widetilde{E}'} \tag{1}$$

The balls $B(x,r)$ are measurable for all x and $r > 0$.

Definition 1.1. Let $\Phi : R_+ \to R_+$. For the function $f \in E(R^n) \cap L_1^{loc}(R^n)$ the *generalized fractional-maximal function* $M_\Phi f$ is defined as follows:

$$(M_\Phi f)(x) = \sup_{r>0} \Phi(r) \int\limits_{B(x,r)} |f(y)dy|$$

That is, consider the operator $M_\Phi : L_1^{loc}(R^n) \to L_0(R^n)$. Denote by $M_E^\Phi \equiv M_E^\Phi(R^n)$ the cone of functions u, for which there is a function $f \in E(R^n)$ such that

$$u(x) = (M_\Phi f)(x),$$

$$\|u\|_{M_E^\Phi} = \inf\{\|f\|_E : f \in E(R^n), \ \ M_\Phi f = u\}$$

such a set M_E^Φ will be called a cone of decreasing functions of *generalized fractional-maximal function* modeled on E.

The RIS of Lorentz space $E(R^n) = \Lambda^p(w)$ with the weight w is defined as follows:

$$\|f\|_{\Lambda^p(w)} = \left(\int\limits_0^\infty (f^*(t))^p w(t) dt \right)^{\frac{1}{p}}, \ 1 < p < \infty; \tag{2}$$

Definition 1. Let E be a rearrangement-invariant space. An *optimal rearrangement-invariant space* $X_o = X_o(\mathbb{R}^n)$ for the embedding

$$M_E^\Phi(\mathbb{R}^n) \hookrightarrow X_o(\mathbb{R}^n)$$

is defined as the smallest (in the sense of inclusion) RIS into which M_E^Φ embeds. That is, for any other RIS X for which the embedding holds,

$$M_E^\Phi(\mathbb{R}^n) \hookrightarrow X(\mathbb{R}^n),$$

we have $X_o \subseteq X$.

We denote

$$(T_\Phi f)(x) = \sup_{x<t<\infty} \Phi(t^{1/n})\, t\, f^{**}(t), \qquad x > 0,$$

which is a Hardy-type operator acting on the decreasing rearrangement of f.

Using result from [2], our problem reduces to find the optimal rearrangement-invariant space $X_o(\mathbb{R}^n)$ such that

$$M^{\Phi}_{\Lambda_p}(\mathbb{R}^n) \hookrightarrow X_o(\mathbb{R}^n),$$

which is equivalent to determining the smallest rearrangement-invariant space $\widetilde{X}_o(\mathbb{R}_+)$ (Luxemburg representation) such that

$$\|T_\Phi f\|_{\widetilde{X}_o} \le C\|f\|_{\Lambda^p(w)}.$$

For brevity, we denote $\varphi(t) := \Phi(t^{1/n})$.

Theorem 2. *Let $1 < p < \infty$, $q = 1$, and $\varphi(t) = \Phi(t^{1/n})$ be a quasi-concave function on $(0,\infty)$. Let v and w be weights on $(0,\infty)$. Then the embedding*

$$\|T_\Phi f\|_X \le C\|f\|_{\Lambda^p(w)}$$

holds for some rearrangement-invariant space X if and only if there exists a constant $C > 0$ such that for all $g \in \mathfrak{M}^{\downarrow}$,

$$\|g\|_{X'} \approx \left(\int_0^\infty \left(\int_0^\infty \frac{\varphi(s)\, g(s)}{[\varphi(s)+\varphi(t)]^q}\, ds\right)^{p'} d\nu_p(t)\right)^{1/p'},$$

where

$$d\nu_p(t) = \varphi^r(t)\ \operatorname*{ess\,sup}_{s>t} \frac{1}{\varphi^r(s)} \left(\int_0^\infty \frac{U^{p'}(\tau)\, U^{p'}(s)}{U^{p'}(\tau)+U^{p'}(s)}\, v^{1-p'}(\tau)\, d\tau\right)^{r/p'} dt,$$

for some $r \in (0,\infty)$, and $U(t)$ is an appropriate auxiliary function (typically $U(t) = t$).

Using duality, the norm in the optimal rearrangement-invariant space $\tilde{X}_o$ is given by

$$\begin{aligned}\|f\|_{\tilde{X}_o} &= \sup_g \frac{\int_0^\infty f^*(s)g^*(s)\, ds}{\|g^*\|_{X'}} \\ &= \sup_{g\downarrow} \frac{\int_0^\infty f^*(s)g(s)\, ds}{\left(\int_0^\infty \left(\int_0^\infty \frac{\varphi(s)\, g(s)}{[\varphi(s)+\varphi(t)]^q}\, ds\right)^{p'} d\nu_p(t)\right)^{1/p'}}.\end{aligned}$$

This space $\tilde{X}_0$ is the Luxemburg representation of the minimal rearrangement-invariant envelope X_o for the embedding of

$$M^{\Phi}_{\Lambda_{p(w)}}(\mathbb{R}^n) \hookrightarrow X_o(\mathbb{R}^n)$$

Funding: This research by A. N. Abek and A. Gogatishvili was supported and funded by the grant Ministry of Science and Higher Education of the Republic of Kazakhstan (project no: AP22686420). This research is funded by the grant Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26196065). The research of A. Gogatishvili was supported by the Institute of Mathematics of the Czech Academy of Sciences RVO:67985840,

Keywords: Banach functional spaces, rearrangement invariant spaces, non-increasing rearrangements of functions, spaces of generalized Riesz potential, embedding spaces of generalized fractional-maximal functions in rearrangement invariant spaces.

2020 Mathematics Subject Classification: 46E30, 47G10, 42B25

References

[1] Bennett C., Sharpley R. Interpolation of operators. Acad. Press, Boston. 1988.

[2] Bokayev N. A., Gogatishvili A., and Abek A.N, Cones of monotone functions generated by a generalized fractional maximal function, *TWMS J. Pure Appl. Math* **15**:1 (2024), , 127–141.

[3] Ünver T., Gogatishvili A., Bokayev N. A., and Abek A. N. Characterization of weighted inequalities for superpositions of integral and supremal operators, *Bull. Karaganda Univ. Math. Ser.* **2** (2026), 166–180.

Test-time budgets are a knapsack, not a water-filling problem

M. ALIYEV

Baku State University, Baku, Azerbaijan

miraga.aliyev.qasim@bsu.edu.az

A fixed amount of computation must be divided among a benchmark's questions. Recent rules treat per-question accuracy as concave in the budget and equalise marginal gains [2,3], the water-filling answer. For a token budget under greedy decoding that premise fails.

Item i is answered once the budget reaches the length ℓ_i of the chain the model writes for it, and not before, so its accuracy curve is a step and the aggregate $A(b) = n^{-1}\#\{i : \ \ell_i \le b, \text{ answer correct}\}$ is a sum of steps, convex where the mass of ℓ begins and concave afterwards.

Proposition 1. *With step curves, maximising the number of answered items subject to $\sum_i b_i \le B$ is a knapsack with unit values and costs ℓ_i; its optimum funds items in increasing ℓ_i, and the uniform allocation $b_i = B/n$ is optimal only if $B/n \ge \max_i \ell_i$ or all ℓ_i are equal.*

Proposition 2. *Truncation is observable, so ℓ_i can be discovered online: the ladder b_0, rb_0, $r^2b_0, \dots$ stops at the first rung above ℓ_i and spends at most*

$$\frac{r^2}{r-1}\,\ell_i, \tag{1}$$

minimised at $r = 2$ with a factor of 4. No difficulty predictor is needed, because the model reports whether it finished.

We measure ℓ_i on 120 generated algebra and number-theory items with exact keys, answered by a local open-weight 7B model under greedy decoding with a cap of 1400 tokens: accuracy 0.967, every completion inside the cap, ℓ with median 278 and maximum 1032. The aggregate curve is flat up to 64 tokens, reaches 0.308 at 256 and 0.917 at 512; its second differences in $\log b$ change sign five times, so it is not concave and water filling does not apply. The knapsack gap is large: at a total budget of 256 tokens per item the uniform allocation answers 30.8% and the optimum 90.8%.

The online ladder recovers part of that gap, but its starting rung decides everything. Started at $b_0 = 64$ with $r = 2$, the textbook doubling, it is worse than uniform, because the first rungs buy nothing on a distribution whose mass sits near 300; started near that mass ($b_0 = 384$, $r = 1.25$) it closes 39 to 48% of the uniform-to-optimum gap at tight budgets and all of it once the budget is generous. Per item the ladder costs 2.5 to 3.5 times the oracle, against the bound 4.0 to 4.5 from (1). Re-running 60 items at budgets 64 to 512 reproduces the ladder predicted from the single unconstrained pass on 97.1% of the item-budget pairs; the residual is not sampling noise but scheduling non-determinism, since a request's logits depend on the batch it shares.

Benchmarks that report a compute budget should report the distribution of ℓ as well: it is what decides whether an allocation rule can help at all, and where its ladder must start.

Keywords: test-time compute, budget allocation, knapsack, online algorithms, doubling schedule, language model evaluation.

2020 Mathematics Subject Classification: 90C27, 68W27, 68T05

References

[1] Snell C., Lee J., Xu K., Kumar A. Scaling LLM test-time compute optimally can be more effective than scaling model parameters, arXiv:2408.03314 (2024).

[2] Nguyen M., Gupta S., Le H. Uncertainty-aware budget allocation for adaptive test-time reasoning, arXiv:2605.26849 (2026).

[3] Wan X. et al. The shadow price of reasoning: economic perspective on optimal budget allocation for LLMs, arXiv:2606.03092 (2026).

[4] Damani M. et al. Learning how hard to think: input-adaptive allocation of LM computation, *Proc. ICLR* (2025).

The Geodesic Ray Transform with Finite Measurements: Stability, Reconstruction, and DeepONets

Y.M. ASSYLBEKOV

Super Dispatch, Astana, Kazakhstan

yernat.assylbekov@gmail.com

The geodesic ray transform is a fundamental operator arising in inverse problems. Its inversion has important applications in tomography and seismic imaging. Classical formulations assume complete knowledge of the transform, corresponding to an infinite collection of measurements. In practical settings, however, only finitely many measurements are available, and the unknown is typically approximated within a finite-dimensional model class.

In this work, we study the inversion of the geodesic ray transform on a compact non-trapping Riemannian manifold with strictly convex boundary under this finite-measurement setting. We discretize the measurement space using shape-regular meshes and piecewise constant finite element spaces. This construction provides a finite-dimensional approximation of the geodesic ray transform that models the available measurements.

Using functional analytic techniques, we establish stability estimates of $L^2 \to L^2$ type for the finite-measurement inverse problem. We use these estimates to develop a globally convergent reconstruction procedure. Finally, we show that the inverse map of the geodesic ray transform can be approximated by a DeepONet on the considered finite-dimensional model class.

The proposed framework provides a mathematically justified approach to the inversion of the geodesic ray transform under finite-dimensional measurement constraints and is applicable to computational tomography and related geometric inverse problems.

Keywords: geodesic ray transform, inverse problems, finite measurements, stability estimates, reconstruction, DeepONet, operator learning.

2020 Mathematics Subject Classification: 44A12, 65R32, 68T07, 58J90

A packing ceiling for colour-vision-deficiency-safe categorical palettes

G. BABAYEVA

Baku State University, Baku, Azerbaijan

babayevagunel@bsu.edu.az

A categorical palette asks its colours to stay distinguishable, including for viewers with a colour-vision deficiency. Practice says this fails somewhere between six and ten colours, and palette optimisers quietly relax their separation requirement as the palette grows: the accessible sequences of [1] use minimum perceptual distances of 20, 18 and 16 for six, eight and ten colours. We derive that ceiling, compute it, and measure how far generative models fall below it.

The geometry. Dichromatic simulation [2] is linear in linear RGB, and its singular values, $(1.349, 1.051, 7.5 \cdot 10^{-8})$ for protan and $(1.206, 0.974, 4.1 \cdot 10^{-7})$ for deutan, show the map has rank two: the gamut is projected onto a surface and colours differing only along the lost direction merge; the tritan model keeps a third component of relative size 0.118. Let K be the image of the sRGB cube in the perceptually uniform CAM02-UCS space [3]; by cell counting its area is 4380 for protan, the smallest of the three.

Proposition 1. *If n colours are pairwise at least d apart under a rank-two simulation, the discs of radius $d/2$ about their images are disjoint and lie in $K \oplus B_{d/2}$, so*

$$n \;\le\; \frac{\operatorname{area}\big(K \oplus B_{d/2}\big)}{\pi(d/2)^2}. \tag{1}$$

On the protan region (1) gives $n \le 37.0$ at $d = 16$ and $n \le 25.8$ at $d = 20$. The attainable number is smaller, because a palette must survive normal vision and all three deficiencies at once: maximising the worst-case separation over these four views yields certified palettes of 19, 15, 13 and 9 colours at $d \ge 16, 18, 20, 25$, and inside a usable lightness band $J \in [45, 85]$, which legibility requires and [1] also imposes, only 11, 10, 8 and 6. The designers' folklore is a theorem about a projection, and the relaxations in [1] track this frontier.

What generators produce. Language models already recommend palettes [4], with no accessibility check. We asked three model tiers for 180 palettes of six, eight and ten colours, half with an explicit instruction to be deficiency-safe, and scored each the same way. Without the instruction the median separation is about 3 and 3.3% of palettes match the strongest deployed accessible palette (Okabe–Ito scores 11.0 here); with it, 17.8% do. The instruction helps but does not solve: the rate falls from 33% at six colours to 17% at eight and 3% at ten, and none of the 180 reached $d = 20$, where eight colours provably fit. Independent attempts make a compliant palette cost $\log \delta / \log(1 - p)$ draws: about 15 with the instruction, 88 without.

The constraint is satisfiable and the ceiling computable: generative palette design lacks not room but a check, cheap enough to run inside the generator's loop.

Keywords: colour vision deficiency, categorical palettes, sphere packing, perceptual uniformity, accessibility, generative design.

2020 Mathematics Subject Classification: 52C15, 68U05, 68T05

References

[1] Petroff M.A. Accessible color sequences for data visualization, arXiv:2107.02270 (2021, rev. 2024).

[2] Machado G.M., Oliveira M.M., Fernandes L.A.F. A physiologically-based model for simulation of color vision deficiency, *IEEE Trans. Vis. Comput. Graph.*, **15**:6 (2009), 1291–1298.

[3] Luo M.R., Cui G., Li C. Uniform colour spaces based on CIECAM02 colour appearance model, *Color Res. Appl.*, **31** (2006), 320–330.

[4] Xia D. et al. ColorGPT: leveraging large language models for multimodal color recommendation, *Proc. ICDAR* (2025).

Categorizing factors influencing hip-fracture care standards for causal diagrams

K. BAKTYBERGEN[1], B. SOBOLEV[2]

[1]*Nazarbayev University School of Medicine, Astana, Republic of Kazakhstan*
[2]*School of Population and Public Health, University of British Columbia, Vancouver, Canada*
[1]kazybek.baktybergen@nu.edu.kz, [2]boris.sobolev@ubc.ca

Graphical models are increasingly used in causal inference to guide covariate selection and define causal estimands. However, applying these models to healthcare delivery requires clearly representing the involved factors and their relationships to the assessed care. This proves challenging when candidate factors are derived from heterogeneous literature and span different aspects of patients, healthcare services, resources, clinical actions, and timing.

We are conducting a scoping review to identify and synthesize factors that influence adherence to five standards of hip-fracture care: early surgery, prompt medical optimization, early mobilization, dedicated operating-room availability, and orthogeriatric co-management. Drawing evidence

from quantitative, qualitative, mixed-methods, and interventional studies, we extract and harmonize the reported factors across the literature, alongside their associations with adherence and any proposed mechanisms.

We initially classified these factors using an extended Donabedian framework comprising case-mix, structural, and care-process domains. To support subsequent causal-diagram development, we propose a relational classification that treats an *episode of care* as a primitive, irreducible entity of real-world healthcare delivery. Consequently, properties describing timing, clinical actions, service models, and episode-specific resource arrangements are treated as coordinate properties of the same episode, rather than as separate ontological levels.

For each specified care standard, we classify literature-derived factors into three distinct roles: *adherence-influencing factors*, which affect whether the standard is met; *adherence-enabling conditions*, which make adherence possible; and *clinical grounds for justified deviation*, which provide a clinical rationale for non-adherence during a specific episode. This approach transforms disparate factors from evidence synthesis into a representation that removes the logical tension in causal-diagram development.

Ultimately, hip-fracture care serves as a clinically important case study for the broader methodological challenge of translating heterogeneous evidence about real-world care delivery into representations suitable for causal reasoning.

Keywords: hip fracture, standards of care, causal diagrams, causal inference, episode of care, relational classification.

A recovery operator for Haar wavelet spaces

S.Zh. BASSAROV[1], N.T. TLEUKHANOVA[2]

Nazarbayev University, Astana, Kazakhstan
[1]*Geometry LLP, Astana, Kazakhstan*
[2]*L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*
[1]bassarov.serzhan98@gmail.com, [2]tleukhanova@rambler.ru

We present a constructive approximation of a function using the function's values at certain points. The functions we are approximating have their Fourier—Haar coefficients to decay rapidly.

First we describe a method to compute Fourier—Haar coefficients. We derive an error of the method in terms of Fourier—Haar coefficients. Thus, we analyse the convergence rate of the error. If the function is a Haar polynomial with a degree at most $2^m - 1$ (meaning that $\prod_{i=1}^{d} k_i < 2^m$, where k_i is a degree of ith variable), then the method approximates Fourier—Haar coefficients of the polynomial exactly with at most $O(m^{d-1}2^m)$, d–dimension, points needed for calculation.

The method is based on a sparse grid with weights and nodes given explicitly.

Second we construct a recovery function based on the method described above. The recovery function is a Haar polynomial with coefficients calculated using the function's values. We obtain that the recovery function is exact for the corresponding Haar polynomial. We analyse the recovery function's error in the metric L_∞ and obtain upper bound of the error in terms of m.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23488613).

Keywords: multi-dimension, function approximation, Fourier coefficient, Fourier series, Haar function, Haar system.

2020 Mathematics Subject Classification: 41A10, 41A25

Uncertainty-Aware Evaluation of Local OCR-VLMs for Russian and Kazakh Financial Document Recognition

A. BOTAIBEKOV[1], A. ASKAROV[2]

[1]*Satbayev University, Almaty, Kazakhstan*
[2]*Astana IT University, Astana, Kazakhstan*
[1]asset517@yandex.kz, [2]anuar.askarov@astanait.edu.kz

Financial reports are challenging for OCR-oriented vision-language models because recognition errors differ greatly in practical significance. A minor transcription error in ordinary text may have little impact, whereas an incorrect digit, sign, reporting period, unit, or table association can alter the meaning of a financial indicator. This study evaluates locally deployable OCR-VLMs for Russian- and Kazakh-language financial documents represented as page images.

Two complementary tasks are considered. In the first, rendered PDF pages are provided directly to vision-language models for image-to-text transcription with preservation of document structure. In the second, the same images are used to extract predefined financial information into a structured representation. Extracted values are additionally checked for arithmetic consistency, enabling separate analysis of recognition, structural association, and numerical reasoning errors. Particular attention is given to errors in numerical values, signs, reporting periods, units, table relationships, omissions, and hallucinated values.

The evaluation uses manually verified reference data. Because documents differ in layout, table structure, numerical density, and complexity, model performance is analyzed at the level of individual document elements. A hierarchical Bayesian logistic model is used to compare OCR-VLM configurations while accounting for document-specific variation and differences between element types. Posterior distributions provide uncertainty-aware comparisons rather than relying only on aggregate accuracy estimates.

A central focus is local deployment. Instead of frontier multimodal models accessed through cloud APIs, the study evaluates OCR-VLMs that can run on consumer-grade GPUs and examines the trade-off between recognition reliability and computational requirements.

Keywords: OCR-VLM, LLM, AI, financial document recognition, Russian language, Kazakh language, document AI, Bayesian evaluation, local inference.

Asymptotics of a Doubly Singularly Perturbed Third-Order BVP with Piecewise Constant Argument

N.U. BUKANAY[1], A.T. ASSANOVA[2], A.E. MIRZAKULOVA[3]

[1]*Farabi University, Almaty, Kazakhstan*
[1]*Institute of Mathematics and Mathematical Modelling, Almaty, Kazakhstan*
[2]*Institute of Mathematics and Mathematical Modelling, Almaty, Kazakhstan*
[3]*Farabi University, Almaty, Kazakhstan*

[1]bukanay.nuraiym@kaznu.kz, [2]assanova@math.kz, [3]aziza.mirzakulova@bk.ru

We consider a two-point boundary value problem for a third-order differential equation featuring two small parameters and a piecewise constant argument on $[0, T]$:

$$\varepsilon^2 y'''(t) + \varepsilon A_3(t) y''(t) + A_2(t) y'(t) + A_1(t) y(t) + A_0(t) y(\beta(t)) = F(t), \tag{1}$$

$$B_i y^{(i)}(0, \varepsilon) + C_i y^{(i)}(T, \varepsilon) = d_i, \quad i = 0, 1, 2, \tag{2}$$

here, $\varepsilon > 0$ is a small parameter and $B_i, C_i, d_i,\ i = 0, 1, 2$ are known coefficients. A piecewise constant argument is $\beta(t) = \theta_{r-1}$, if $t \in [\theta_{r-1}, \theta_r),\ r = 1, \ldots, p+1,\ 0 = \theta_0 < \theta_1 < \cdots < \theta_{p+1} = T$.
The following assumptions are made:
I. Let the functions $A_i(t)$, $i = 0, 1, 2, 3$ belongs to $C([0, T], \mathbb{R})$, and $F(t) \in C([0, T], \mathbb{R})$.
II. $A_1(t) > 0,\ \ 0 \le t \le T$.
III. The roots $\mu_1(t), \mu_2(t)$ of the additional characteristic equation

$$\mu^2 + A_3(t)\mu + A_2(t) = 0,$$

satisfy $\mu_1(t) \neq \mu_2(t)$, and $\operatorname{Re} \mu_i(t) \le -\gamma_i < 0,\ i = 1, 2$.
IV. $\Delta(\varepsilon) = B + CP_{1,p+1}(\varepsilon)$.

The solution and its derivatives are bounded at the boundary points $t = 0$ and $t = 1$, but this does not rule out the existence of a boundary layer. In singularly perturbed problems, the solution itself remains bounded, while its derivatives may vary rapidly within a thin boundary layer of width $O(\varepsilon)$. The boundary layer arises due to the reduction in the order of the differential equation as $\varepsilon \to 0$, which leads to an incompatibility between the reduced problem and the imposed boundary conditions. This phenomenon is well understood in the theory of singular perturbations. As $\varepsilon \to 0$, the order of the differential equation decreases, and the reduced problem is unable to satisfy all boundary conditions simultaneously. As a result, boundary layers form at the endpoints of the interval. Since the following estimates hold:

$$y(0,\varepsilon) = O(1), \;\; y'(0,\varepsilon) = O(1), \;\; y''(0,\varepsilon) = O\left(\frac{1}{\varepsilon}\right), \;\; \varepsilon \to 0,$$

$$y(1,\varepsilon) = O(1), \;\; y'(1,\varepsilon) = O(1), \;\; y''(1,\varepsilon) = O\left(\frac{1}{\varepsilon}\right), \;\; \varepsilon \to 0,$$

the solution exhibits a first-order boundary layer at $t = 0$ and $t = 1$. If the roots of the additional characteristic equation $\mu^2 + A_2(t)\mu + A_1(t) = 0$ have opposite signs ($\operatorname{Re}\mu_1(t) \le -\gamma_1 < 0$, $\operatorname{Re}\mu_2(t) \ge \gamma_2 > 0$), then boundary layer phenomena occur at both $t = 0$ and $t = 1$. If both roots are negative $\operatorname{Re}\mu_1(t) \le -\gamma_1 < 0$, $\operatorname{Re}\mu_2(t) \le -\gamma_2 < 0$), the boundary layer forms only at $t = 0$.

To analyze and construct the solution to problem (1)–(2), the parametrization method is applied.

Theorem 1. *Assume that conditions I–IV hold. Then problem (1)–(2) admits a unique solution*

$$\tilde{y}(t,\varepsilon) = \begin{cases} \tilde{y}_1(t,\varepsilon), \; t \in [\theta_0, \theta_1), \\ \tilde{y}_2(t,\varepsilon), \; t \in [\theta_1, \theta_2), \\ \vdots \\ \tilde{y}_{p+1}(t,\varepsilon), \; t \in [\theta_p, \theta_{p+1}], \end{cases} \tag{3}$$

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23485509).

Keywords: piecewise constant, boundary value problem, asymptotical method, Cauchy problem, parametrization method.

2020 Mathematics Subject Classification: 34K10, 34B05, 34E15

References

[1] Mirzakulova A.E., Dauylbayev M.K. Asymptotic expansion of the solution for singularly perturbed boundary value problem with boundary jumps, *Miskolc Mathematical Notes*, **24** (2020), 309–324.

[2] Dzhumabaev D. On one approach to solve the linear boundary value problems for Fredholm integro-differential equations, *Applied Mathematics and Computation*, **294** (2016), 342–357.

[3] Assanova A.T. Solution of the boundary value problem for differential equation with piecewise-constant argument of generalized type, *Kazakh Mathematical Journal*, **22** (2022), 15–27.

Local Well-posedness of the Boussinesq Equation with Linear Restoring Force

A. ESFAHANI[1], N. ZHARDEMOVA[2]

[1,2]*Nazarbayev University, Astana, Kazakhstan*

[1]amin.esfahani@nu.edu.kz, [2]nazerke.zhardemova@nu.edu.kz

We study the initial-value problem for the Boussinesq equation with a linear restoring force, which models the transverse deflections of an elastic rod on an elastic foundation [1].

$$\begin{cases} u_{tt} = -\gamma u + \alpha\Delta u - \Delta^2 u + \Delta(f(u)), \quad \gamma > 0, \alpha \in \mathbb{R}, x \in \mathbb{R}^n,\ t > 0, \\ u(x,0) = g(x), \\ u_t(x,0) = h(x), \end{cases} \tag{1}$$

where $f(u) = \pm u^2, \pm u^3$ and $n < 3$.

The corresponding phase function is more regular than one of classical Boussinesq equation when $\gamma = 0$. The restoring-force term changes the low-frequency behavior, which, in turn, affects the associated energy functional

$$E(u,u_t) = \frac{1}{2}\|(-\Delta)^{-\frac{1}{2}}u_t\|_{L^2}^2 + \frac{\gamma}{2}\|(-\Delta)^{-\frac{1}{2}}u\|_{L^2}^2 + \frac{\alpha}{2}\|u\|_{L^2}^2 + \frac{1}{2}\|\nabla u\|_{L^2}^2 + \int_{\mathbb{R}^n} F(u)\,dx.$$

Unlike the equation without a restoring force, the restoring-force term shows that the natural energy requires control of u at low frequency in a negative homogeneous norm. At the same time, the fourth-order dispersion contributes the high-frequency term $\frac{1}{2}\|\nabla u\|_{L^2}^2$. Therefore the energy suggests a function space that behaves like $\dot{H}^{-1}$ at low frequency and like H^1 at high frequency. Hence, the natural energy space is $Y = H^1 \cap \dot{H}^{-1}$. We prove local well-posedness in the natural energy space generated by the linear restoring force. This observation motivates us to work in a Bourgain-type space X^b adapted to the energy space Y (see [1,2,3]):

$$\|u\|_{X^b} = \left\| \langle\xi\rangle^2 |\xi|^{-1} \langle |\tau| - \varphi(\xi)\rangle^b \widetilde{u}(\xi,\tau) \right\|_{L^2_{\xi,\tau}}, \tag{2}$$

which captures the low-frequency structure created by the restoring force and preserves the required high-frequency Sobolev regularity.

A further difficulty arises from the nonlinear term $\Delta(f(u))$. Although the Duhamel operator contains the smoothing factor

$$\frac{\sin((t-s)\varphi(D))}{\varphi(D)},$$

the Laplacian in the nonlinearity causes a derivative loss. Therefore, to prove well-posedness one needs multilinear estimates that recover the derivatives lost in $\Delta(f(u))$.

We establish the bilinear estimate needed for the quadratic nonlinearity and the corresponding trilinear estimate for the cubic nonlinearity. Combining these estimates with a contraction mapping argument, we prove local well-posedness for the quadratic and cubic equations in Y.

Theorem 1 (Local well-posedness for the quadratic equation)**.** *Let $f(u) = \pm u^2$. Then, the Cauchy problem* (1) *with $\gamma = 1$ and $\alpha = \pm 1$ is locally well-posed in Y.*

Theorem 2 (Local well-posedness for the cubic equation)**.** *Let $f(u) = \pm u^3$. Then, the Cauchy problem* (1) *with $\gamma = 1$ and $\alpha = \pm 1$ is locally well-posed in the energy space Y.*

The proofs are crucially deduced when the following bilinear and trilinear estimates are proved.

Lemma 3. *Let* $n \geq 1$, $\frac{1}{2} < b < 1$, *and* $0 < T < 1$. *Then*

$$\left\| |D|^2 \langle D \rangle^{-2} (u_1 u_2) \right\|_{X^{b-1}(0,T)} \lesssim T^{1-b} \left\| u_1 \right\|_{X^b(0,T)} \left\| u_2 \right\|_{X^b(0,T)}. \tag{3}$$

Lemma 4. *Let* $n \geq 1$, $\frac{1}{2} < b < 1$, *and* $0 < T < 1$. *Then*

$$\left\| |D|^2 \langle D \rangle^{-2} (u_1 u_2 u_3) \right\|_{X^{b-1}(0,T)} \lesssim T^{1-b} \prod_{j=1}^{3} \left\| u_j \right\|_{X^b(0,T)}. \tag{4}$$

Keywords: Boussinesq equation, linear restoring force, local well-posedness, Bourgain spaces, bilinear estimates, nonlinear dispersive equations.

2020 Mathematics Subject Classification: 35Q53, 35A01, 35B30, 42B37

References

[1] Bourgain J. Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations, *Geom. Funct. Anal.*, **3**:2 (1993), 107–156.

[2] Bourgain J. Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations: Part II: The KdV-equation, *Geometric & Functional Analysis*, **3**:3 (1993), 209–262.

[3] Bourgain J. Periodic nonlinear Schrödinger equation and invariant measures, *Communications in Mathematical Physics*, **166**:1 (1994), 1–26.

Stability Properties of Pseudofinite Abelian Groups

Z.G. FAZYL[1], N.D. MARKHABATOV[2]

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

fzhebei@gmail.com

By Szmielew's theorem, two abelian groups are elementarily equivalent if and only if their Szmielew invariants coincide [1]. In the form given by Eklof and Fisher [2], for a prime p and $n \geq 1$ these invariants are

$$\alpha_{p,n}(A) = d_p\big((p^{n-1}A)[p]/(p^nA)[p]\big), \qquad \beta_p(A) = \inf_{k\geq 0} d_p\big((p^kA)[p]\big), \qquad \gamma_p(A) = \inf_{k\geq 0} d_p\big(A/(A[p^k]+pA)\big),$$

where $d_p(V) = \dim_{\mathbb{F}_p} V$ if this dimension is finite and $d_p(V) = \omega$ otherwise, and $\varepsilon(A) \in \{0,1\}$ indicates whether A has unbounded exponent. Every theory of an abelian group is stable by the Baur–Monk theorem [3,4], while ω-stability and superstability are characterized by descending chains of pp-definable subgroups [5,6].

A group is pseudofinite if it is an infinite model of the theory of all finite groups. We use the following criterion.

Theorem 1 [7]. An infinite abelian group A is pseudofinite if and only if $\beta_p(A) = \gamma_p(A)$ for every prime p.

Put $\mathcal{S}(A) = \{(p,n) : \alpha_{p,n}(A) \neq 0\}$ and $\mathcal{I}(A) = \{(p,n) : \alpha_{p,n}(A) = \omega\}$.

Theorem 2. Let A be an infinite pseudofinite abelian group. Then:

(1) $\mathrm{Th}(A)$ is superstable if and only if $\mathcal{I}(A)$ is finite and $\beta_p(A) = \gamma_p(A) < \omega$ for every prime p;

(2) $\mathrm{Th}(A)$ is ω-stable if and only if $\mathcal{S}(A)$ is finite and $\beta_p(A) = \gamma_p(A) = 0$ for every prime p; equivalently, $A \cong \mathbb{Q}^{(\kappa)} \oplus B$ for some cardinal κ and some bounded abelian group B;

(3) $\mathrm{Th}(A)$ is strongly minimal if and only if A is either an infinite elementary abelian p-group or a nonzero divisible torsion-free abelian group.

We also give examples showing that the hierarchy

$$\text{strongly minimal} \subsetneq \omega\text{-stable} \subsetneq \text{superstable} \subsetneq \text{stable}$$

is strict in the class of theories of pseudofinite abelian groups.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26194247).

Keywords: pseudofinite abelian group, Szmielew invariants, strongly minimal group, ω-stability, superstability, stability.

2020 Mathematics Subject Classification: 03C13, 03C20, 03C45.

References

[1] Szmielew W. Elementary properties of Abelian groups, Fundamenta Mathematicae, 41:2 (1955), 203–271.

[2] Eklof P.C., Fisher E.R. The elementary theory of abelian groups, Annals of Mathematical Logic, 4 (1972), 115–171.

[3] Baur W. Elimination of quantifiers for modules, Israel Journal of Mathematics, 25 (1976), 64–70.

[4] Monk L. Elementary-recursive decision procedures, Ph.D. Thesis, University of California, Berkeley (1975).

[5] Macintyre A. On ω_1-categorical theories of abelian groups, Fundamenta Mathematicae, 70:3 (1971), 253–270.

[6] Ziegler M. Model theory of modules, Annals of Pure and Applied Logic, 26 (1984), 149–213.

[7] Fazyl Zh.G. Criteria for pseudofiniteness of abelian groups, submitted to Bulletin of the Karaganda University. Mathematics Series.

AZ-Eval: negative transfer from Kazakh fine-tuning to Azerbaijani is largely orthographic

N. GARIBLI

Baku State University, Baku, Azerbaijan

qaribli.nihat.nemat@bsu.edu.az

Quantitative evaluation of multilingual language models remains concentrated on high-resource languages. Kazakh has been served by a growing suite of adapted benchmarks, developed in large part at ISSAI; Azerbaijani, a Turkic language written in Latin script, has to our knowledge no open parallel evaluation set. The gap invites an appealing shortcut for practitioners working on Turkic languages: adapt a model to Kazakh and expect the gains to carry over. We extend the benchmark methodology to Azerbaijani in order to test that expectation directly.

We introduce *AZ-Eval*, an open benchmark of 356 human-verified short-answer items, each stated in parallel Azerbaijani and English. Of these, 255 were written by hand by a native speaker and 101 were instantiated from eight fact templates, each realised in four syntactic variants assigned round-robin so that phrasing is not confounded with item difficulty. The items cover mathematics, history, geography, science, culture, language and a control stratum of universally known world facts. The set of accepted answers is closed under a rule-based Azerbaijani inflection generator (vowel harmony, case, possessive and copular suffixes), so that a morphologically inflected but factually correct response is not penalised. A majority-class baseline attains only 1.1%, confirming that the reported scores carry signal rather than distributional structure.

Scoring uses exact match and token F_1 under a chain of increasingly permissive normalisations,

$$\textsc{strict} \longrightarrow \textsc{morph} \longrightarrow \textsc{lenient} \longrightarrow \textsc{translit},$$

in which each stage adds one transformation (suffix stripping, diacritic folding and Cyrillic-to-Latin transliteration respectively), applied symmetrically to prediction and reference. The score increment across a stage is therefore an attribution of error mass to that single surface phenomenon. Three open models are evaluated under identical prompts and deterministic greedy decoding. All model comparisons are paired, reported with percentile bootstrap confidence intervals (1000 resamples), assessed by a sign-flipping permutation test, which suits binary exact-match scores that violate normality, and corrected for multiplicity by the Holm procedure.

Every model exhibits a large Azerbaijani–English gap (14.0–26.4 points, all $p < 0.001$ after Holm correction). The central comparison is between the Kazakh-tuned Qolda-AVL-5B and its own base model Qwen3-VL-4B-Instruct, matched in architecture, parameter count and quantisation:

Normalisation	Base model	Kazakh-tuned	Deficit	p (Holm)
STRICT	16.6%	3.1%	13.5 pt	0.0024
TRANSLIT	18.0%	12.1%	5.9 pt	0.0364

The comparison is cleanest on the control stratum, the 96 items whose answers every model demonstrably knows in English. There the two models are statistically indistinguishable in English (59.4% versus 59.4%, $p = 1.0000$), which rules out a general capability difference, while in Azerbaijani the base model reaches 45.8% and the Kazakh-tuned model 10.4%, a deficit of 35.4 points (Holm $p = 0.0030$). Transliteration lifts the Kazakh-tuned model to 33.3% on this stratum, recovering 62% of the deficit. The aggregate figures above are smaller because the Azerbaijan-specific strata, where both models fail, dilute the contrast; they are reported as the primary numbers because only the full sample has the power to establish that the post-transliteration residual is non-zero.

Kazakh fine-tuning thus yields negative, not positive, transfer. The attribution identifies the mechanism as predominantly orthographic: 326 of 356 Azerbaijani answers are produced in Cyrillic, several in Kazakh outright, against 1 of 356 for the base model, while both models answer in Latin script when prompted in English. An explicit instruction to use the Latin alphabet moves only 12 of those 326 answers, so the behaviour is not an artefact of prompting. Transliteration removes 56%

of the deficit but not all of it: the residual 5.9 points remain significant, so the model recovers much of its knowledge under a script-agnostic reading while a smaller non-orthographic loss persists. The decomposition is sharp in the other direction too: the writing-system increment is +0.0 points for six of the eight runs and rises to +9.0 for exactly the Kazakh-tuned model on Azerbaijani, whereas morphology contributes at most +1.1 points and diacritics at most +0.6 anywhere in the table.

The mathematics stratum shows the effect at its most extreme. On these 28 items both larger models answer roughly a quarter of the questions correctly in English (25.0% and 28.6%) and one item in twenty-eight in Azerbaijani (3.6%). The collapse is not ignorance: on the Azerbaijan-specific history stratum, where the models genuinely lack the facts, scores are near zero in *both* languages. Mathematics is therefore the stratum where knowledge is demonstrably present and language alone blocks access to it.

Adaptation to a closely related language written in a different script can therefore degrade rather than improve performance, and the damage is predominantly, though not exclusively, orthographic rather than semantic. This localises the remedy: script-constrained decoding addresses the larger share of the loss and is likely to be cheaper than additional target-language data, while the residual deficit indicates that it will not suffice on its own.

Data and code: the benchmark, the evaluation pipeline and all raw model outputs are openly available at `https://github.com/nihatgaribli/AZ-Eval` (code MIT, data CC BY 4.0). Every reported figure reproduces from a fixed seed.

Keywords: low-resource evaluation, Azerbaijani, cross-lingual transfer, Turkic languages, writing systems, large language models.

2020 Mathematics Subject Classification: 68T50, 68T07, 62F40

References

[1] Efron B. Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7**:1 (1979), 1–26.

[2] Holm S. A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics*, **6**:2 (1979), 65–70.

[3] Rajpurkar P., Zhang J., Lopyrev K., Liang P. SQuAD: 100,000+ questions for machine comprehension of text, in: *Proceedings of EMNLP 2016*, Austin (2016), 2383–2392.

What does a certified self-consistent answer certify?

N. GARIBLI

Baku State University, Baku, Azerbaijan

qaribli.nihat.nemat@bsu.edu.az

Sampling a language model several times and returning the majority answer [1] now carries statistical guarantees: the aggregated answer is the mode of the model's answer distribution with high probability [2], and the limit of the vote is a functional of that distribution [3,4]. Every one of those statements is about the model, not about the task. We measure the difference.

Draw k samples and let $n_{(1)} \geq n_{(2)}$ be the two largest option counts. Conditional on the samples landing on one of the two leaders, the leader's count is $\mathrm{Bin}(n_{(1)} + n_{(2)}, \frac{1}{2})$ under the tie

null, an exact certificate for the modal answer; for a margin Δ between the leading probabilities Hoeffding makes

$$k \;\geq\; \frac{2\ln(1/\alpha)}{\Delta^2} \tag{1}$$

sufficient. Two contamination-free item sets of contrasting difficulty are answered by a local open-weight 7B model at temperature 0.8: 120 easy items with 16 samples and 60 harder ones with 12, 2640 scored responses.

Three ceilings, not one. On the harder set 22.9% of samples contain no extractable answer. Scoring those as wrong, as a naive pipeline does, puts the limit of the vote at 0.617; treating them as the abstentions they are raises it to 0.758; and since five options make a plurality sufficient, the operative limit, the share of items whose modal answer is the key, is 0.800. At $k = 11$ the vote reaches 0.827, marginally above that limit, which is what finite k permits: an item whose modal answer is wrong can still be carried by sampling noise, and only as $k \to \infty$ does the vote settle at 0.800. The distance between the first and the last of these ceilings is the measurement, not the model.

Samples buy coverage, not safety. On the harder set the certificate covers 32.8% of items at $k = 6$ and 58.3% at $k = 12$, while the residual risk moves from 0.014 to 0.029 rather than falling. At $n = 60$ that movement is a single item, so the claim defended here is the weaker and safer one, that the extra samples do not visibly buy safety, not that risk provably rises. Certified items are right 97.1% of the time against 56.0% for uncertified ones, so the rule separates well; what it cannot remove are the items on which the model is stably wrong, which are precisely the ones a large margin protects. On the easy set the same procedure certifies 97.5% with a residual risk of zero: the worth of a certificate belongs to the regime, not to the rule. The median margin on the harder set is 0.58, so by (1) two thirds of its items need more than the twelve samples drawn.

A benchmark that reports self-consistency should therefore report three numbers beside it: the share of samples that abstain, the plurality ceiling, and the rate at which certified answers are wrong.

Keywords: self-consistency, majority voting, certification, sequential testing, answer extraction, language model evaluation.

2020 Mathematics Subject Classification: 62F03, 62L10, 68T05

References

[1] Wang X., Wei J., Schuurmans D. et al. Self-consistency improves chain of thought reasoning in language models, *Proc. ICLR* (2023).

[2] Cordero-Encinar P., Duncan A.B. Certified self-consistency: statistical guarantees and test-time training for reliable reasoning in LLMs, arXiv:2510.17472 (2025).

[3] Bay Y.Y., Yearick K.A. When more sampling hurts: the modal ceiling and correlation ceiling of test-time scaling, arXiv:2606.28661 (2026).

[4] Liu Y. Two calls, two moments, and the vote-accuracy curve of repeated LLM inference, arXiv:2605.03379 (2026).

On the characterization of discrete weighted Hardy inequalities of non-increasing sequences

A GOGATISHVILI[1], N.A. BOKAYEV[2], N.K. KUZEUBAYEVA[3], T ÜNVER[4]

[1]*Institute of Mathematics, Czech Academy of Sciences, Praha, Czech Republic*
[2,3]*L.N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan*
[4]*Department of Mathematics, Kirikkale University, Kirikkale, Türkiye*

[1]gogatish@math.cas.cz, [2]bokayev2011@yandex.ru, [3]nurgul.kuzeubaeva@mail.ru,
[4]tugceunver@kku.edu.tr

We provide a complete characterization of the weighted discrete Hardy inequality restricted to the cone of non-increasing sequences. This is achieved using a reduction technique that transforms the inequality into the corresponding inequality on the cone of non-negative sequences [1–3]. Because two different types of reduction theorems are available, two distinct equivalent characterizations of the discrete weighted Hardy inequality are established on the cone of non-increasing sequences. The first result provides a new proof of the Bennett–Grosse-Erdmann theorem, while the second establishes a new equivalent characterization.

Let $p, q \in (0, \infty)$ and let $\{w(k)\}_{k\in\mathbb{N}}$ and $\{v(k)\}_{k\in\mathbb{N}}$ be non-negative weight sequences. We characterize the conditions under which there exists a positive constant C (independent of $\{x(n)\}_{n\in\mathbb{N}}$) such that the two-weight discrete Hardy inequality

$$\left(\sum_{k=1}^{\infty}\Big(\sum_{n=1}^{k} x(n)\Big)^{q} w(k)\right)^{\frac{1}{q}} \le C\left(\sum_{k=1}^{\infty} x(k)^{p}\, v(k)\right)^{\frac{1}{p}} \tag{1}$$

holds for every non-negative, non-increasing sequence $\{x(n)\}_{n\in\mathbb{N}}$. For a sequence $\{v(k)\}_{k\in\mathbb{N}}$ of non-negative numbers we put

$$V(n) = \sum_{k=1}^{n} v(k), \quad n \in \mathbb{N}, \qquad V(\infty) = \lim_{n\to\infty} \sum_{k=1}^{n} v(k).$$

For brevity, denote

$$\beta(k) := V(k+1)^{-1}\, V(k)^{\frac{1}{1-p}}\, v(k+1), \qquad k \in \mathbb{N}.$$

The main result is stated below.

Theorem 1. *Assume $0 < p, q < \infty$ with $p \neq 1$, and let $\{w(k)\}_{k\in\mathbb{N}}$ and $\{v(k)\}_{k\in\mathbb{N}}$ be sequences of non-negative numbers with $v(1) > 0$. Then inequality* (1) *holds for all non-negative, non-increasing sequences $\{x(n)\}_{n\in\mathbb{N}}$ if and only if*

(i) $1 < p \le q$ *and* $\max\{C_0, C_1, C_2\} < \infty$, *where*

$$C_0 := \Big(\sum_{k=1}^{\infty} k^{q} w(k)\Big)^{\frac{1}{q}} \Big(\sum_{k=1}^{\infty} v(k)\Big)^{-\frac{1}{p}},$$

$$C_1 := \sup_{1\le n<\infty} \Big(\sum_{k=n}^{\infty} w(k)\Big)^{\frac{1}{q}} \Big(\sum_{k=1}^{n} k^{\frac{p}{p-1}} \beta(k)\Big)^{\frac{p-1}{p}}, \qquad C_2 := \sup_{1\le n<\infty} \Big(\sum_{k=1}^{n} k^{q} w(k)\Big)^{\frac{1}{q}} \Big(\sum_{k=n}^{\infty} \beta(k)\Big)^{\frac{p-1}{p}};$$

(ii) $1 < p$, $q < p$ *and* $\max\{C_0, C_3, C_4\} < \infty$, *where*

$$C_3 := \left(\sum_{n=1}^{\infty}\Big(\sum_{k=n}^{\infty} w(k)\Big)^{\frac{q}{p-q}} w(n) \Big(\sum_{k=1}^{n} k^{\frac{p}{p-1}} \beta(k)\Big)^{\frac{(p-1)q}{p-q}}\right)^{\frac{p-q}{pq}},$$

$$C_4 := \left(\sum_{n=1}^{\infty}\Big(\sum_{k=1}^{n} k^q w(k)\Big)^{\frac{q}{p-q}} n^q w(n)\Big(\sum_{k=n}^{\infty}\beta(k)\Big)^{\frac{(p-1)q}{p-q}}\right)^{\frac{p-q}{pq}};$$

(iii) $p<1$, $p\le q$ *and* $\max\{C_5, C_6\}<\infty$, *where*

$$C_5 := \sup_{1\le n<\infty}\Big(\sum_{k=n}^{\infty} w(k)\Big)^{\frac{1}{q}} n\, V(n)^{-\frac{1}{p}}, \qquad C_6 := \sup_{1\le n<\infty}\Big(\sum_{k=1}^{n} k^q w(k)\Big)^{\frac{1}{q}} V(n)^{-\frac{1}{p}};$$

(iv) $q<p<1$ *and* $\max\{C_7, C_8\}<\infty$, *where*

$$C_7 := \left(\sum_{n=1}^{\infty}\Big(\sum_{k=n}^{\infty} w(k)\Big)^{\frac{q}{p-q}} w(n) \sup_{1\le k\le n} k^{\frac{pq}{p-q}} V(k)^{-\frac{q}{p-q}}\right)^{\frac{p-q}{pq}},$$

$$C_8 := \left(\sum_{n=1}^{\infty}\Big(\sum_{k=1}^{n} k^p w(k)\Big)^{\frac{q}{p-q}} n^p w(n)\, V(n)^{-\frac{q}{p-q}}\right)^{\frac{p-q}{pq}}.$$

Furthermore, if C denotes the optimal constant in (1), *then*

$$C \approx \begin{cases} C_0+C_1+C_2 & \text{in case (i)},\\ C_0+C_3+C_4 & \text{in case (ii)},\\ C_5+C_6 & \text{in case (iii)},\\ C_7+C_8 & \text{in case (iv)}, \end{cases}$$

where the constants of equivalence depend only on p and q.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26196065). The research of A. Gogatishvili was supported by the Institute of Mathematics of the Czech Academy of Sciences RVO:67985840.

Keywords: Hardy inequality, monotone sequences, reduction, kernel operator, discrete inequality.

2020 Mathematics Subject Classification: 26D15, 47G10, 46E30

References

[1] Gogatishvili A., Bokayev N. A., Karshygina G., Kuzeubayeva N., Ünver T. Reduction theorems for the discrete Hardy operator on cones of monotone sequences, *J. Math. Sci. (N. Y.)* **291**:2 (2025), 217–223.

[2] Gogatishvili A., Bokayev N. A., Kuzeubayeva N. K. Reduction theorems for discrete Hardy operator on monotone sequence cones ($0<p<1$), *Journal of Mathematics, Mechanics and Computer Science* **128**:4 (2025), 12–24.

[3] Gogatishvili A., Pick L., Ünver T. Weighted inequalities for discrete iterated kernel operators, *Math. Nachr.* **295**:11 (2022), 2171–2196.

Pointwise and Integral Majorants on Cones of Monotone Functions

A GOGATISHVILI[1], G KARSHYGINA[2], T ÜNVER[3]

[1]*Institute of Mathematics, Czech Academy of Sciences, Praha, Czech Republic*
[2]*Astana International University, Astana, Republic of Kazakhstan*
[3]*Department of Mathematics, Kirikkale University, Kirikkale, Türkiye*
[1]gogatish@math.cas.cz, [2]karshygina84@mail.ru, [3]tugceunver@kku.edu.tr

We present a unified example of calculating pointwise and integral majorants on cones of monotone functions. Such cones play an important role in the study of decreasing rearrangements of generalized Bessel and Riesz potentials. Explicit formulas for pointwise and integral majorants are obtained, demonstrating the universality of the method.

Let f^* be a decreasing rearrangement of a Lebesgue measurable function $f : \mathbb{R}^n \to \mathbb{R}$, that is, a non-negative, decreasing, right-continuous function on $\mathbb{R}_+ = (0, \infty)$, which is equimeasurable with f:

$$\mu_n\{x \in \mathbb{R}^n : \ |f(x)| > y\} = \mu_1\{t \in \mathbb{R}_+ : \ f^*(t) > y\}, \quad y \in \mathbb{R}_+.$$

The symmetric rearrangement $f^\sharp : \mathbb{R}^n \to [0, \infty]$ is given by

$$f^\sharp(x) = f^*(V_n|x|^n), \quad x \in \mathbb{R}^n,$$

where V_n is the volume of the unit ball in $\mathbb{R}^n$. It is also equimeasurable with f.

By $E \equiv E(\mathbb{R}^n)$ we denote the (quasi) Banach function space (BFS) in the sense of the axiomatics developed by C. Bennett and R. Sharpley, where the notion of norm may be replaced by a quasinorm. The BFS $E(\mathbb{R}^n)$ is called a rearrangement invariant space (RIS) if its (quasi) norm is monotone with respect to decreasing rearrangements:

$$g^* \le f^*, \quad f \in E \Rightarrow g \in E, \ \ \|g\|_E \le \|f\|_E.$$

For the RIS $E(\mathbb{R}^n)$ we denote by $E'(\mathbb{R}^n)$ the associated RIS, with norm

$$\|g\|_{E'} = \sup\left\{\int_{\mathbb{R}_+} g^*(\tau) f^*(\tau)\, d\tau : \ f \in E, \ \|f\|_E \le 1\right\}.$$

By $\tilde{E}(\mathbb{R}_+)$, $\tilde{E}'(\mathbb{R}_+)$ we denote their Luxemburg representations, i.e. RIS on the semiaxis $\mathbb{R}_+$, such that

$$\|f\|_{E(\mathbb{R}^n)} = \|f^*\|_{\tilde{E}(\mathbb{R}_+)}, \quad \|g\|_{E'(\mathbb{R}^n)} = \|g^*\|_{\tilde{E}'(\mathbb{R}_+)}.$$

Let $T \in (0, \infty]$. For $T < \infty$ we denote by $\tilde{E}(0,T)$, $\tilde{E}'(0,T)$ the restrictions of $\tilde{E}(\mathbb{R}_+)$, $\tilde{E}'(\mathbb{R}_+)$ to $(0,T)$. Define

$$\tilde{E}^{\downarrow}(0,T) = \{g \in \tilde{E}(0,T) :\ 0 \le g(t)\downarrow,\ t \in (0,T)\}.$$

We introduce a set of continuous, positive, decreasing functions on $(0,T)$:

$$\Omega(T) = \left\{\varphi \in C(0,T) :\ 0 < \varphi\downarrow;\ \int_0^t \varphi(\tau)\,d\tau < \infty,\ t \in (0,T)\right\}.$$

Now consider the kernel

$$f_\psi(t,\tau) = \min\{\psi(t),\psi(\tau)\} = \psi(\max\{t,\tau\}), \quad t,\tau \in (0,T),$$

with $\psi(t) = \frac{1}{1+t}$.

Construct the cone

$$K_{\psi,E}(T) = \left\{h(t) = \int_0^T f_\psi(t,\tau)g(\tau)\,d\tau : g \in \tilde{E}^{\downarrow}(0,T)\right\}.$$

Theorem 1. *In the above notation, if the conditions*

$$\psi \in \Omega(T), \quad f_\psi(t,\cdot) \in \tilde{E}'(0,T), \quad t \in (0,T),$$

are satisfied, then the pointwise and integral majorants on the cone $K_{\psi,E}(T)$ are given by

$$\lambda_K(t) = \|f_\psi(t,\cdot)\|_{\tilde{E}'(0,T)}, \quad t \in (0,T),$$

$$\tilde{\lambda}_K(t) = \left\|\frac{1}{t}\int_0^t f_\psi(\xi,\cdot)\,d\xi\right\|_{\tilde{E}'(0,T)}, \quad t \in (0,T).$$

Thus, we obtain a unified result: for the kernel $\psi(t) = \frac{1}{1+t}$, the pointwise and integral majorants are expressed in terms of norms in the associated space $\tilde{E}'$. This extends previous results and demonstrates the universality of the method.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26196065). The research of A. Gogatishvili was supported by the Institute of Mathematics of the Czech Academy of Sciences RVO:67985840,

Keywords:Cones of monotone functions · Pointwise and integral coverings · Integral majorants on cones · Generalized Riesz potentials · Cones of decreasing rearrangements of potentials.

2020 Mathematics Subject Classification: 46E30, 47G10, 31C45

References

[1] Bennett C., Sharpley R. Interpolation of operators. Acad. Press, Boston. 1988.

[2] Bokayev N. A., Goldman M. L., Karshygina G. Zh. Cones of functions with monotonicity conditions for generalized Bessel and Riesz potentials. *Math. Notes.* **104**: 3, (2018), 356–373.

[3] Bokayev N. A., Goldman M. L., Karshygina G. Zh. Criteria for embeddings of generalized Bessel and Riesz potential spaces in rearrangement invariant spaces. *Eurasian Math. Journal* **10**:2, (2019), 8–29.

[4] Gogatishvili A., Neves J. S., Opitz B. Optimality of embeddings of Bessel-potential type spaces. Function spaces, differential operators and nonlinear analysis. (2005). pp. 97 – 102.

[5] Goldman M. L. On optimal embedding of generalized Bessel and Riesz potentials. *Proc. Steklov Inst. Mathem.* **269**, (2010). 101–123.

[6] Goldman M. L. On the cones of rearrangements for generalized Bessel and Riesz potentials. *Complex Variables and Elliptic Equations.* **55**: 8-10, (2010), 817–832.

A Fuzzy TOPSIS-Based Intelligent Decision Support Architecture for Human Resource Management in Banking

F.S. HUSEYNLI

Azerbaijan Technical University, Baku, Azerbaijan

fuad201685@gmail.com

Decision-making in banking human resource management is formulated as a multi-criteria optimization problem involving heterogeneous quantitative and qualitative indicators. Classical evaluation approaches usually assume precise numerical information and therefore cannot adequately represent uncertainty originating from expert judgments expressed in linguistic terms. Multi-criteria decision-making (MCDM) methods, particularly the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) and its fuzzy extensions, have proven to be effective tools for solving such problems under uncertainty [2,3]. Building upon these mathematical principles and our previous research on human resource optimization in the banking sector [1], this study proposes an intelligent decision support architecture based on the Fuzzy TOPSIS method integrated into a Human Resource Information System (HRIS). The proposed framework combines mathematical modeling with artificial intelligence technologies to provide transparent, reproducible, and explainable employee evaluation. In the proposed architecture, fuzzy mathematical models are employed to aggregate expert knowledge, manage uncertainty, and improve the reliability of personnel assessment. Furthermore, the developed model provides a systematic mechanism for ranking alternatives, supporting managerial decision-making, and ensuring the consistency and interpretability of evaluation results in complex banking environments. Linguistic assessments are represented by triangular fuzzy numbers

$$\tilde{a} = (l, m, u), \qquad l \le m \le u, \tag{1}$$

where l, m, and u denote the lower, modal, and upper values, respectively. The fuzzy decision matrix is defined as

$$\widetilde{X} = [\widetilde{x}_{ij}], \qquad i = 1, \ldots, m,\ j = 1, \ldots, n. \tag{2}$$

After transforming heterogeneous indicators into fuzzy values, benefit criteria are normalized according to

$$\widetilde{r}_{ij} = \left(\frac{l_{ij}}{u_j^*}, \frac{m_{ij}}{u_j^*}, \frac{u_{ij}}{u_j^*} \right), \tag{3}$$

where $u_j^* = \max_i u_{ij}$. The weighted normalized matrix is then obtained as

$$\widetilde{V} = [\widetilde{v}_{ij}] = w_j \otimes \widetilde{r}_{ij}, \tag{4}$$

where w_j denotes the normalized importance weight of criterion j. The fuzzy positive ideal solution (FPIS) and fuzzy negative ideal solution (FNIS) are subsequently determined, and the distances from these reference solutions are computed as

$$D_i^+ = \sum_{j=1}^{n} d(\widetilde{v}_{ij}, A^+), \qquad D_i^- = \sum_{j=1}^{n} d(\widetilde{v}_{ij}, A^-), \tag{5}$$

where $d(\cdot, \cdot)$ denotes the vertex distance between two triangular fuzzy numbers. The distance between two triangular fuzzy numbers $\tilde{a} = (l_1, m_1, u_1)$ and $\tilde{b} = (l_2, m_2, u_2)$ is evaluated using the vertex method

$$d(\tilde{a}, \tilde{b}) = \sqrt{\frac{(l_1 - l_2)^2 + (m_1 - m_2)^2 + (u_1 - u_2)^2}{3}}. \tag{6}$$

The fuzzy positive and negative ideal solutions are respectively defined as

$$A^+ = (\tilde{v}_1^+, \ldots, \tilde{v}_n^+), \qquad A^- = (\tilde{v}_1^-, \ldots, \tilde{v}_n^-), \tag{7}$$

where $\tilde{v}_j^+ = \max_i \tilde{v}_{ij}$ and $\tilde{v}_j^- = \min_i \tilde{v}_{ij}$.

The final ranking is determined through the closeness coefficient

$$CC_i = \frac{D_i^-}{D_i^+ + D_i^-}, \tag{8}$$

where greater values of CC_i indicate alternatives closer to the fuzzy positive ideal solution.

Unlike conventional Fuzzy TOPSIS applications that are primarily limited to ranking, the proposed architecture models the complete decision lifecycle, including data validation, fuzzy transformation, criteria versioning, explainability, role-based authorization, approval workflow, and audit logging. The integration of mathematical optimization, fuzzy set theory, and intelligent system architecture provides a rigorous framework for reliable decision support under uncertainty and enables scalable implementation in modern banking environments [2,3].

Keywords: Fuzzy TOPSIS, fuzzy sets, intelligent decision support systems, artificial intelligence, banking, multi-criteria decision making.

2020 Mathematics Subject Classification: 03E72, 68T37, 90B50, 91B06.

References

[1] Huseynli F.S. A Multi-Criteria Decision-Making Approach to Human Resource Optimization in Bank. *Universum: Technical Sciences*, No. 8(137), 2025. DOI: 10.32743/UniTech.2025.137.8.20723.

[2] Hwang C.L., Yoon K. *Multiple Attribute Decision Making: Methods and Applications.* Springer-Verlag, Berlin, 1981.

[3] Chen C.-T. Extensions of the TOPSIS for Group Decision-Making under Fuzzy Environment. *Fuzzy Sets and Systems*, **114**(1), 2000, 1–9.

How many items does a benchmark need when its items come in clusters?

E. HUSEYNOV

Baku State University, Baku, Azerbaijan

emil.huseynov.zabil@bsu.edu.az

Error bars on benchmark scores are computed as if items were drawn independently [1]. They rarely are: items arrive in examination sittings, in passages, in template families, and items from one cluster share whatever makes it easy or hard. We measure what this costs, and find the cost very different for one system's score than for a comparison of two.

Let item i lie in cluster $c(i)$ and let the outcome have within-cluster correlation ρ. With unequal cluster sizes m_k the variance of the mean carries the design effect

$$D = 1 + (\tilde{m} - 1)\rho, \qquad \tilde{m} = \frac{\sum_k m_k^2}{\sum_k m_k}, \tag{1}$$

so n items behave like n/D independent ones, and certifying a true gap δ at level α with power $1 - \beta$ needs $n \geq D\,\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2/\delta^2$. Both are known [2,3]; what is not is how large ρ is at the unit where items actually nest, and whether pairing escapes it.

Proposition 1. *Write $X_{s,i} = \mu_s + u_{s,c(i)} + e_{s,i}$ for system s, with cluster effects of variance σ_u^2 and cross-system cluster correlation r. The paired difference $d_i = X_{A,i} - X_{B,i}$ has between-cluster variance $2\sigma_u^2(1 - r)$, so*

$$D_d = 1 + (\tilde{m} - 1)\,\frac{2\sigma_u^2(1 - r)}{\sigma_d^2}, \tag{2}$$

which equals 1 when $r = 1$: a difficulty shared by both systems cancels from the difference, and only the part they disagree on survives.

On a benchmark of 1385 certified-key multiple-choice items from 23 national examination sittings ($\tilde{m} = 68.2$), answered by five systems under two input modalities (13 850 scored responses), a single system's accuracy has within-sitting correlation 0.024 to 0.104: design effects 2.6 to 8.0, median 4.8, and a cluster bootstrap widens the interval by a factor matching $\sqrt{D}$ to within 0.06. For paired differences the design effect is systematically smaller, 1.03 to 4.27 (median 2.45), and falls as the two systems' sitting effects agree, as (2) predicts: over the twenty pairs the rank correlation between D_d and r is -0.54 ($p = .015$), with r from 0.17 to 0.90.

The 1385 items therefore carry the information of 291 independent items for a single accuracy and 566 for a comparison. The smallest certifiable gap is 6.1 points, not the 3.9 an independence calculation promises, and certifying 5 points needs 2092 clustered items where 854 independent ones would do. The nesting unit and its intraclass correlation deserve to be reported as routinely as the number of items.

Keywords: benchmark evaluation, design effect, intraclass correlation, cluster bootstrap, sample size, paired comparison.

2020 Mathematics Subject Classification: 62D05, 62F03, 68T05

References

[1] Miller E. Adding error bars to evals: a statistical approach to language model evaluations, arXiv:2411.00640 (2024).

[2] Kotawala A. Resolution diagnostics for paired LLM evaluation, *ICML Workshop on Hypothesis Testing* (2026), arXiv:2605.30315.

[3] Gönen M. Sample size and power for McNemar's test with clustered data, *Statistics in Medicine*, **23**:14 (2004), 2283–2294.

[4] Bowyer S., Aitchison L., Ivanova D.R. Position: don't use the CLT in LLM evals with fewer than a few hundred datapoints, *Proc. ICML*, **267** (2025), 81143–81184.

Logical Adversarial Networks Based on Ehrenfeucht–Fraïssé–Taimanov Games

K. JUBATKANOV[1], N. MARKHABATOV[2], J. TUSSUPOV[3], K. BEISENBAYEVA[4]

[1,2,3,4]*L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*

[1]kuanysh2002@enu.kz, [2]markhabatovΩ_nd@enu.kz, [3]tussupovΩ_da@enu.kz,
[4]beisenbayevaksh@enu.kz

A generative adversarial network is a minimax game between a generator, which produces artificial objects, and a discriminator, which tries to distinguish them from objects of a target class. Classically, the discriminator is a neural network detecting statistical properties of the data and gives no explicit description of the structure learned by the generator. We replace it with a logical adversary whose distinguishing power is restricted to first-order sentences of bounded quantifier rank.

Let τ be a finite signature, $\mathrm{Fin}(\tau)$ the class of finite τ-structures, and $\mathcal{L}_r$ the set of first-order τ-sentences of quantifier rank at most r. We write $A \equiv_r B$ if A and B satisfy the same sentences of $\mathcal{L}_r$. In the r-round Ehrenfeucht–Fraïssé–Taimanov game $\mathrm{EFT}_r(A, B)$, Spoiler chooses an element of one structure and Duplicator answers in the other; Duplicator wins if the correspondence $a_i \mapsto b_i$, $1 \le i \le r$, is a partial isomorphism. By the classical result [1–4], $A \equiv_r B$ if and only if Duplicator has a winning strategy in $\mathrm{EFT}_r(A, B)$.

Definition 1. A logical generative adversarial network of rank r is a triple $\mathrm{LOGAN}_r = (G_\theta, \mathcal{D}_r, T)$, where $G_\theta : Z \to \mathrm{Fin}(\tau)$ is a generator, T is a target first-order theory, and $\mathcal{D}_r = \{D_\varphi : \varphi \in \mathcal{L}_r\}$ with $D_\varphi(M) = 1$ if $M \models \varphi$ and $D_\varphi(M) = 0$ otherwise.

Theorem 2. *Let $A = G_\theta(z)$ and $B \models T$. The following conditions are equivalent: (i) $A \equiv_r B$; (ii) $D_\varphi(A) = D_\varphi(B)$ for every $D_\varphi \in \mathcal{D}_r$; (iii) Duplicator has a winning strategy in* $\mathrm{EFT}_r(A, B)$.

Thus, the generated structure is indistinguishable from the target one by every logical adversary of quantifier rank at most r.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26194247).

Keywords: generative adversarial networks, model theory, Ehrenfeucht–Fraïssé–Taimanov games, logical discriminator, finite structures.

2020 Mathematics Subject Classification: 03C07, 03C13, 68T07

References

[1] Fraïssé R. Sur une nouvelle classification des systèmes de relations, *C. R. Acad. Sci. Paris* **230** (1950), 1022–1024.

[2] Fraïssé R. Sur quelques classifications des systèmes de relations, *Publ. Sci. Univ. Alger, Sér. A* **1** (1954), 35–182.

[3] Ehrenfeucht A. An application of games to the completeness problem for formalized theories, *Fund. Math.* **49** (1961), 129–141.

[4] Taimanov A. D. The characteristics of axiomatizable classes of models, *Algebra i Logika. Sem.* **1**:4 (1962), 5–31.

Enumeration of perfect matchings in line graphs

M.A. JUMADILDAYEV

Nazarbayev University, Astana, Kazakhstan
[1]*Institute of Mathematics and Mathematical Modelling, Almaty, Kazakhstan*
medetdzhuma@gmail.com

Let $G = (V, E)$ be a simple graph. The line graph $L(G)$ of a simple graph G is obtained by associating a vertex in $L(G)$ with an edge in G and connecting two vertices in $L(G)$ if they are incident in G. A perfect matching is a set of independent edges such that they cover all vertices. Let $pm(G)$ denote the number of perfect matchings in the graph G. We consider the problem of enumerating the number of perfect matchings in line graphs. Dong, Yan, and Zhang [3] found that the lower bound on the number of perfect matchings of line graphs is $2^{|E|-|V|+1}$. Our main result:

Theorem 1. *Let $G = (V, E)$ be a simple graph. If $v \in V$, define $X_v \sim N(0, 1)$ as a random variable of the standard normal distribution. Then, the number of perfect matchings in the line graph $L(G)$:*

$$pm(L(G)) = \mathbb{E}\left[\prod_{(u,v)\in E} (X_u + X_v)\right].$$

In other words, the number of perfect matchings in the line graph can be expressed structurally as the expectation of an n-variate normal distribution. This allows one to use various techniques for integral estimation to enumerate perfect matchings asymptotically. In particular, if G is dense, then it is possible to apply Laplace's method to obtain an asymptotic formula for the integral.

Proposition 2. *Let d_v denote the degree of a vertex $v \in V$. If $d = \Theta(|V|)$, it is possible to use Laplace's method to estimate the number of perfect matchings in $L(G)$.*

A natural example is when $G = K_n$ is the complete graph. The line graph of the complete graph is called the triangular graph [1]. The exact number of maximum matchings in the triangular graph was calculated up to graphs of size 18 in the OEIS sequence A297565 [2]. We obtain the asymptotic formula for the number of perfect matchings in the triangular graph.

Theorem 3. *The number of perfect matchings in the triangular graph $L(K_n)$:*

$$pm(L(K_n)) \sim \frac{e^{\frac{1}{4}(n-1)n(\log(2(n-1))-1)-\frac{8}{81}}}{\sqrt{2^{-n}\left(\frac{3n-4}{n-1}\right)^{n-1}}}.$$

Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR 28713025).

Keywords: perfect matchings, line graphs.

2020 Mathematics Subject Classification: 05A16, 05C70

References

[1] Brualdi R., Ryser H.J. *Combinatorial Matrix Theory*, Cambridge University Press, New York (1991).

[2] Sloane N.J.A. OEIS entry A297565, (2017).

[3] Dong F., Yan W., Zhang F. On the number of perfect matchings of line graphs, *Discrete Applied Mathematics*, **161**:6 (2013), 794–801.

Mamba state-space model for sequential WiFi indoor localization

M. KAIROLLA

Institute of Smart Systems and Artificial Intelligence,

Nazarbayev University, Astana, Republic of Kazakhstan

madiyar.kairolla@nu.edu.kz

In this work, we apply the Mamba structured state-space model (SSM) [1] to sequential WiFi-based indoor localization, formulated as a sequence-learning problem over continuous user trajectories. As a user traverses a building, consecutive received signal strength (RSS) observations from wireless access points (WAPs) exhibit strong temporal dependencies, since each measurement is inherently conditioned on the user's motion history. Recurrent neural networks (RNNs) have been

shown to exploit this temporal structure effectively, achieving the best reported performance on the WiFine dataset [2].

The proposed model processes each complete trajectory $\{r_t\}_{t=1}^{T}$, where $r_t \in \mathbb{R}^d$ denotes the preprocessed RSS vector at step t, through a single-layer Mamba backbone. The backbone maintains a hidden state $h_t \in \mathbb{R}^{d_{\text{inner}} \times d_{\text{state}}}$, updated selectively at each step, and produces position predictions $\hat{p}_t \in \mathbb{R}^3$ and floor logits at every timestep in a sequence-to-sequence protocol.

Additionally, we propose a novel loss function, the temporally weighted Mean Euclidean Distance (wMED),

$$\mathcal{L}_{\text{wMED}} = \sum_{t=1}^{T} w_t \, \|\hat{p}_t - p_t\|_2, \qquad w_t = \alpha + (1-\alpha)\,\frac{t-1}{T-1}, \tag{1}$$

where $p_t \in \mathbb{R}^3$ is the ground-truth position and $\alpha = 0.3$ is the minimum relative weight assigned to the first timestep.

This weighting scheme is motivated by the architecture of the proposed model. At $t = 1$, the hidden state $h_0 = 0$ carries no trajectory information, so early predictions are inherently less reliable. As the sequence progresses, the preceding path is progressively captured within the hidden state. By assigning lower weights to the early timesteps, the loss concentrates gradient pressure on the parts of the trajectory where the model has accumulated sufficient context for corrections to be both meaningful and effective. Optimizing (1) directly minimizes the evaluation metric rather than a surrogate objective. An ablation study confirms that wMED converges better than MSE: under identical training conditions, the test MED is 3.11 m with wMED versus 3.72 m with MSE.

On the WiFine test set, the proposed model achieves a Mean Euclidean Distance (MED) of 3.11 m. This is within 0.06 m of the best reported RNN result of 3.05 m [2], while the model requires 52% fewer parameters (31 800 vs. 65 200). Furthermore, the model attains a floor classification accuracy exceeding 98%, a metric not reported in prior work on this dataset.

Keywords: Mamba, WiFi fingerprinting, received signal strength, indoor localization, sequence modelling.

2020 Mathematics Subject Classification: 68T07, 68T10, 94A12

References

[1] Gu A., Dao T. Mamba: Linear-time sequence modeling with selective state spaces, *Conference on Language Modeling (COLM)*, 2024.

[2] Khassanov Y., Nurpeiissov M., et al. Finer-level sequential WiFi-based indoor localization, *IEEE/SICE International Symposium on System Integration (SII)*, 2021, 857–862.

Interpolation theorem in discrete net spaces $n_{p,q}(F)$

A.K. KALIDOLDAY[1], E.D. NURSULTANOV[2]

[1]*Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan*
[2]*Kazakhstan branch of Lomonosov Moscow State University Astana, Kazakhstan*
[1]aitolkynnur@gmail.com, [2]er-nurs@yandex.kz

Definition 1. Let S be the set of all finite sets of indices from $\mathbb{Z}^n$. For a fixed set $M \subset S$ we define the space $n_{p,q}(M)$ $(0 < p, q \le \infty)$ as the set of sequences $a = \{a_m\}_{m\in\mathbb{Z}^n}$ with quasinorm for $0 < p < \infty$, $0 < q < \infty$

$$\|a\|_{n_{p,q}(M)} = \left(\sum_{k=1}^{\infty} k^{\frac{q}{p}-1}(\bar{a}_k(M))^q\right)^{\frac{1}{q}},$$

and for $q = \infty$, $0 < p \le \infty$

$$\|a\|_{n_{p,\infty}(M)} = \sup_{1\le k<\infty} k^{\frac{1}{p}}\bar{a}_k(M),$$

where

$$\bar{a}_k(M) = \sup_{\substack{e\in M\\ |e|\ge k}} \frac{1}{|e|}\left|\sum_{m\in e} a_m\right|,$$

where $|e|$ is the number of indices in e. Here, $\bar{a}_k(M)$ denotes a discrete analogue of the averaging of a function and is defined as the supremum of the average sums of the elements a_m over all $e \in M$.

These spaces, introduced by E.D. Nursultanov, are called *net spaces.*

Let $b > 1$. The parametric family $G_b = \{G_k\}_{k\in\mathbb{N}}$ will be called a local net in $\mathbb{Z}^n$, if

$$G_k \hookrightarrow G_{k+1} \quad \text{and} \quad |G_k| = b^k.$$

Here $|G_k|$ is the number of elements in the set G_k.

The set $F_{G_b} = \{G_k + x\}_{k\in\mathbb{N}, x\in\mathbb{Z}^n}$ will be called the global net generated by the net G.

Example. The set of cubes with edge length 2^k, $k \in \mathbb{N}$ in $\mathbb{Z}^n$ is a global net generated by the net of concentric cubes in $\mathbb{Z}^n$ with edge lengths 2^k, $k \in \mathbb{N}$.

Theorem 2. *Let $G = \{G_t\}_{t>0}$ be a local net, and let $F = \bigcup_{x\in\mathbb{Z}^n}(G + x)$ be the global net generated by the net G. Assume that $0 < p_0 < p_1 < \infty$ and $0 < q_0 \le q_1 \le \infty$, $0 < \theta < 1$, $1 \le \tau \le \infty$,*

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

If a linear operator T satisfies the following inequalities for some constants $M_0, M_1 > 0$

$$\|Ta\|_{n_{q_i,\infty}(G+x)} \le M_i\|a\|_{n_{p_i,1}(G+x)}, \quad x \in \mathbb{Z}^n,\ a \in n_{p_i,1}(G + x),\ i = 0, 1, \tag{1}$$

then for any $a \in n_{p,\tau}(F)$, the following inequality holds:

$$\|Ta\|_{n_{q,\tau}(F)} \le c\max\{M_0, M_1\}\|a\|_{n_{p,\tau}(F)},$$

where the constant $c > 0$ depends only on the parameters $p_0,\ p_1,\ q_0,\ q_1,\ p,\ q,\ \tau,\ \theta$.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP25794959).

Keywords: net spaces, discrete net spaces, Lorentz space, Marcinkiewicz-type interpolation theorem, real interpolation method, linear operators, local nets, global nets.

2020 Mathematics Subject Classification: 46B70

References

[1] Nursultanov, E.D. Net spaces and inequalities of Hardy–Littlewood type, *Sbornik: Mathematics — Collection: Mathematics*, **3**:189, (1998), 399–419.

[2] Kalidolday, A.K., Nursultanov, E.D., Sharipova, A.N. Marcinkiewicz-type interpolation theorem for discrete net spaces, *Bulletin of the Karaganda University. Mathematics Series*, **2**:122, (2026), 80–89.

Fourier-Mode Allocation and Physical Error in a One-Dimensional Heat Problem

A. KIM

Nazarbayev University, Astana, Kazakhstan

atay.kim@nu.edu.kz

Fourier neural operators (FNOs) learn mappings between functions. In the model used here, the input is lifted to c channels and passed through four residual layers,

$$z^{(l+1)} = z^{(l)} + 0.5\ \mathrm{GELU}\Big(\mathcal{K}_m z^{(l)} + W_l z^{(l)}\Big). \tag{1}$$

Here $\mathcal{K}_m$ takes a real Fourier transform, applies learned complex weights to bins $0, \dots, m-1$, and transforms the result back; W_l is a pointwise linear map. Thus m is the number of retained Fourier modes and c is the hidden width. At an approximately fixed parameter count, increasing m requires decreasing c.

The effect of this trade-off was tested for the one-dimensional heterogeneous heat equation

$$-\frac{d}{dx}\left(k(x)\frac{dT}{dx}\right) = 1, \qquad T(0) = T(1) = 0, \tag{2}$$

The network received $[\log k(x), x]$ and predicted $T(x)$. Reference temperatures were computed on 64 uniform nodes by a finite-volume method. Face conductivity was the harmonic mean of neighbouring nodal values, and face flux was approximated by $q_{i+1/2} = -k_{i+1/2}(T_{i+1} - T_i)/\Delta x$. Balancing the two face fluxes produced a tridiagonal system, which was solved by the Thomas algorithm. The tridiagonal system was also assembled as a dense matrix and solved with a general linear solver; both solutions agreed to numerical tolerance, and the reference balance error was effectively zero.

Conductivity fields were generated from Matérn-3/2 Gaussian draws and rescaled realization by realization to the same contrast range. Shorter correlation lengths produced faster spatial changes and more high-frequency input energy. Separate models were trained for each correlation length. Two nearly parameter-matched cases were compared: $(m, c) = (4, 61)$ with 138,416 parameters and $(32, 23)$ with 139,302. Five paired optimization seeds were used.

Temperature, reconstructed heat flux $q(x) = -k(x)T'(x)$, and discrete balance residuals were evaluated. At the shortest correlation length, the mean run-level median temperature errors were 1.170% and 0.709% for $m = 4$ and $m = 32$; the flux errors were 17.062% and 11.204%. From the smooth to the rough family, the $m = 4$ minus $m = 32$ flux-error gap increased by 6.34 ± 1.14

percentage points (mean ± sample standard deviation across the five paired optimization seeds). The increase was positive for all five seeds (exact one-sided sign test, $p = 0.03125$).

The one-cell residual gave the opposite ranking: 312.98% for $m = 4$ and 322.72% for $m = 32$. After the same imbalance was summed over finite-volume blocks, the ranking reversed from two cells onward; the 32-cell errors were 13.79% and 8.50%. The large pointwise values came from localized derivative errors, not a temperature field with more than 300% global error. Since $q'(x) = 1$, the exact flux is affine even for rough $k(x)$, so the benchmark mainly tests whether derivative information is preserved in the predicted temperature. In this setting, the high-mode model recovered flux more accurately for shorter-correlated inputs, although the result depended on the scale of the physical-error measure.

Keywords: Fourier neural operators, operator learning, spectral allocation, heterogeneous diffusion, heat flux, PDE consistency, multiscale conservation.

2020 Mathematics Subject Classification: 68T07, 35J15, 65N08.

Proof-Carrying Streamed Workflow Graphs

P. KINGSTON[1], T. FAGNONI[2]

[1,2]*AppliedAI, Abu Dhabi, United Arab Emirates*

[1]phillip.kingston@opus.com, [2]theo.fagnoni@opus.com

Incremental construction of directed workflow graphs can be formulated as a constrained graph-extension problem: each proposed addition should preserve the existence of a viable completion. In AI systems, an emitted addition can be syntactically correct even when its evidence is inadequate, its scope is invalid, or it makes the remaining policy obligations unsatisfiable. We introduce *proof-carrying streams*, in which every addition updates the committed graph and transforms a certificate witnessing that at least one acceptable completion remains.

We model each step in the graph construction by a state s, which records the current committed graph and all information needed to apply future additions and determine whether the construction remains viable. For an arbitrary state space S, an alphabet Σ of possible additions, a deterministic transition function $\delta : S \times \Sigma \to S$ that assigns a unique successor state to each state and addition, and a set F of acceptable states, we define the canonical *viability residual* $\mathscr{R}_s$. The viability residual $\mathscr{R}_s$ contains exactly the finite addition streams that can be admitted from s while leaving the resulting state viable, i.e., still able to reach an accepting state by some finite continuation. As additions are admitted, the viability residual evolves to describe exactly the futures that remain viable from the resulting state. A certificate is sound at a construction state when every future stream it represents belongs to that state's viability residual. Graph construction that begins with a non-empty sound certificate and proceeds through sound certificate updates preserves completion viability through every admitted addition.

We encode each construction state into a proof state that retains enough information to determine completion viability. Without loss of generality, we restrict every proof state to the image of this encoding, so every proof state represents at least one construction state. A proof-state abstraction consists of the encoding, an update rule that computes the next proof state from the current proof state and the next addition, and a classification of proof states as viable or nonviable. The abstraction is viability-exact when two conditions hold: updating an encoded proof state produces the same result as first applying the addition to the construction state and then encoding

its successor, and the proof state's viability classification agrees exactly with the viability of every construction state it represents. These requirements imply that construction states encoded as the same proof state must have equal viability residuals. Consequently, grouping construction states by equality of their residuals forms the coarsest viability-exact abstraction: every viability-exact abstraction must keep construction states with different residuals separate. A finite viability-exact abstraction exists if and only if the construction states have finitely many distinct residuals.

Keywords: proof-carrying streams, workflow graphs, viability residual, residual languages, Nerode equivalence, incremental verification.

2020 Mathematics Subject Classification: 68Q45, 68Q70, 68Q60, 68T27

Explicit solution and well-posedness for a fractional PDE with a loaded term along a moving curve

M.T. KOSMAKOVA[1], D.M. AKHMANOVA[2]

Buketov Karaganda National Research University, Karaganda, Kazakhstan
[1]*Institute of Applied Mathematics, Karaganda, Kazakhstan*
[2]*Buketov Karaganda National Research University, City, Country*
[1]svetlanamir578@gmail.com, [2]danna.67@mail.ru

We study the initial boundary value problem

$$D_{0t}^{\alpha}u = u_{xx} + \lambda\, D_{0t}^{\beta}u\big|_{x=\gamma(t)} + f(x,t), \qquad (x,t)\in\mathbb{R}\times(0,\infty), \tag{1}$$

subject to

$$\lim_{t\to 0} D_{0t}^{\alpha-1}u = 0, \qquad \lim_{t\to 0} D_{0t}^{\alpha-2}u = 0, \qquad \lim_{|x|\to\infty} u(x,t) = 0, \tag{2}$$

where $1<\alpha<2$, $0<\beta<1$, D_{0t}^{ν} denotes the Riemann–Liouville fractional derivative (or integral if $\nu\le 0$), $\lambda\in\mathbb{C}$ is a given parameter, and $\gamma:[0,\infty)\to[0,\infty)$ is non-decreasing with $\gamma(0)=0$.

The paper provides a complete analysis in the class of regular solutions (smooth in x, continuous in t, vanishing at infinity, with a well-defined trace of the loaded fractional derivative). The main idea is to reduce the nonlocal problem to a Volterra integral equation of the second kind for the unknown function

$$\mu(t) := D_{0t}^{\beta}u(\gamma(t),t).$$

Remarkably, this auxiliary equation admits an explicit closed-form solution $\mu(t)$. Once μ is known, the original solution is reconstructed via

$$u(x,t) = \frac{1}{2}\int_0^t\int_{-\infty}^{\infty}(t-\tau)^{\alpha/2-1}\,\phi\!\left(-\frac{\alpha}{2},\frac{\alpha}{2};-\frac{|x-\xi|}{(t-\tau)^{\alpha/2}}\right)\big[f(\xi,\tau)+\lambda\,\mu(\tau)\delta(\xi-\gamma(\tau))\big]\,d\xi\,d\tau, \tag{3}$$

where ϕ is a Wright-type function.

A key technical contribution is the computation of the exact norm of the integral operator in $C[0,T]$. This yields the sharp contraction condition

$$|\lambda| < \frac{\Gamma(\alpha-\beta+1)}{T^{\alpha-\beta}},$$

which guarantees existence and uniqueness for the Volterra equation via Banach's fixed-point theorem.

Furthermore, we show that the loaded term acts as a weak perturbation. **Funding**: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23488740, 2024–2026.).

Keywords: fractional partial differential equation, Riemann–Liouville derivative, loaded equation, well-posedness, weak perturbation.

2020 Mathematics Subject Classification: 35R11, 45D05, 26A33, 35A01, 35A02, 35B30

Numerical Solution of a Nonlocal Problem for a Fredholm-Type Parabolic Integro-Differential Equation

S.K. KUANYSH

Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan

kuanysh.saule1@gmail.com

This study addresses the numerical solution of a nonlocal problem for a Fredholm-type parabolic integro-differential equation. Such problems arise in mathematical modeling of systems with spatial interactions and memory effects, while the presence of nonlocal conditions and integral terms introduces significant analytical and computational challenges [1].

To construct an approximate solution, the Method of Lines is employed to discretize the spatial variable, reducing the original partial integro-differential problem to a nonlocal boundary value problem for a system of Fredholm integro-differential equations. The resulting system is treated using the Dzhumabaev parameterization method, which enables its reduction to a finite-dimensional system of algebraic equations. Based on this formulation, an efficient numerical algorithm is developed [2,3].

Sufficient conditions for the solvability of the discrete problem and the applicability of the proposed numerical approach are established. The convergence and stability of the numerical scheme are also investigated. Numerical experiments are presented to evaluate the performance of the method, and the computed solutions are compared with corresponding exact solutions. The results demonstrate good agreement and confirm the accuracy and computational effectiveness of the proposed approach.

The developed framework provides a reliable numerical foundation for solving complex nonlocal integro-differential models and may also support future integration with data-driven and artificial intelligence approaches for the analysis and approximation of mathematical models governed by nonlocal dynamics.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23485618).

Keywords: nonlocal problem, parabolic integro-differential equation, Fredholm integral term, Method of Lines, parameterization method, numerical approximation.

2020 Mathematics Subject Classification: 45K05, 58J35, 34K28

References

[1] Borachok I., Chapko R., Palianytsia O., Brown J. On the numerical solution of a parabolic Fredholm integro-differential equation by the RBF methods, *Results Appl. Math.*, **26** (2025), 100559.

[2] Temesheva S.M., Kadirbayeva Zh.M., Kuanysh S.K. Analysis of Semi-periodic Problem for Integro-differential Equations of Parabolic Type, *Lobachevskii Journal of Mathematics*, **46**:12 (2025), 6441–6454.

[3] Kuanysh S.K., Kadirbayeva Zh.M. Analysis of Nonlocal Problem for Loaded Parabolic Equation via Numerical Method, *Lobachevskii Journal of Mathematics*, **46**:10 (2025), 5417–5427.

Computational methods for the simplicial geometric category

N. KUANYSHOV[1], I. YEGINBAY[2]

SDU University, Kaskelen, Republic of Kazakhstan

[1]nursultan.kuanyshov@sdu.edu.kz, [2]islam.yeginbay@gmail.com

The Lusternik–Schnirelmann category is a fundamental invariant in homotopy theory [1–3]. Fernández-Ternero, Macías-Virgós, Minuz and Vilches introduced its combinatorial analogues for finite simplicial complexes, including the simplicial geometric category [4,5], based on the strong homotopy theory of Barmak and Minian [6]. We develop a computational framework for this invariant, combining an exact dominated-vertex algorithm for strong collapsibility with a minimum set-cover formulation and a randomized facet-partition search with parallel CUDA restarts.

Let K be a finite abstract simplicial complex with vertex set $V(K)$ and facets $F_0, \dots, F_{M-1}$; we write $K = \langle F_0, \dots, F_{M-1} \rangle$. A vertex v is *dominated* by a vertex $u \neq v$ if every facet containing v also contains u. Deleting a dominated vertex is an *elementary strong collapse*; a sequence of such deletions is a *strong collapse* $K \searrow\searrow L$. The complex K is *strongly collapsible* if it strongly collapses to a vertex, equivalently, if its strong core (which is unique up to isomorphism) consists of one vertex [6].

Definition 1. The *simplicial geometric category* of a nonempty finite simplicial complex K is

$$\operatorname{gscat}(K) = \min\{m \geq 0 : \ K = C_0 \cup \cdots \cup C_m, \ C_i \subseteq K \text{ strongly collapsible}\}.$$

Exact strong cores. Let $\mathcal{F}$ be the normalized facet family of K (duplicate and non-maximal facets removed). For a vertex v put $\mathcal{F}(v) = \{F \in \mathcal{F} : v \in F\}$ and $D_{\mathcal{F}}(v) = \left(\bigcap_{F \in \mathcal{F}(v)} F\right) \setminus \{v\}$.

Proposition 1. A vertex v is dominated if and only if $D_{\mathcal{F}}(v) \neq \emptyset$, and every $u \in D_{\mathcal{F}}(v)$ dominates v.

Repeatedly deleting a vertex v with $D_{\mathcal{F}}(v) \neq \emptyset$ and renormalizing $\mathcal{F}$ terminates after at most $|V(K)| - 1$ deletions and returns a strong core of K. This gives an exact oracle for strong collapsibility.

Exact computation as set cover. A family of subcomplexes covers K if and only if every facet of K lies in one of them. For small complexes we enumerate all strongly collapsible subcomplexes U, record the facet sets $A(U) = \{i : F_i \subseteq U\}$, discard dominated candidates, and

solve the resulting minimum binary set-cover problem; a minimum cover by q subcomplexes gives $\mathrm{gscat}(K) = q - 1$.

Randomized facet-partition search. For larger complexes we search over facet partitions $\{F_0, \dots, F_{M-1}\} = P_1 \sqcup \cdots \sqcup P_q$ with $C_i = \langle P_i \rangle$. In each restart the facets are processed in a seeded random order; a facet F is placed in the first part P for which $\langle P \cup \{F\} \rangle$ is strongly collapsible, otherwise a new part is opened; compatible parts are then merged. Restarts are independent and are executed in parallel on the GPU, while the CPU verifies coverage and the strong collapsibility of every part of the best partition.

Proposition 2. If the search returns q verified parts, then $\mathrm{gscat}(K) \le q - 1$.

Thus randomization affects only the search, not the validity of the certificate; agreement with an independent lower bound determines $\mathrm{gscat}(K)$ exactly. The current bit-mask implementation supports complexes with at most 64 vertices and 512 facets.

Experiments. We used the 6-vertex $\mathbb{R}\mathrm{P}^2$ and a 10-vertex genus-2 surface M_2 from Lutz's census [7,8], Kühnel's 9-vertex $\mathbb{C}\mathrm{P}^2$ [9], and the 12-vertex $S^3 \times S^2$ of Datta [10] (AMD Ryzen 5 7535HS, NVIDIA RTX 4050 Mobile, CUDA 13.1). Table 1 compares exact values with the bounds found by 1000 restarts.

Table 1. Exact values, randomized bounds (1000 restarts) and CPU/CUDA times (s).

Complex	$\lvert V\rvert$	Facets	Exact	Bound	CPU	CUDA	Speedup
$\mathbb{R}\mathrm{P}^2$	6	10	2	2	0.045	0.003	15.0
$\mathbb{C}\mathrm{P}^2$	9	36	2	2	3.254	0.110	29.6
M_2	10	24	2	2	1.959	0.073	26.8
$S^3 \times S^2$	12	112	2	3	30.457	2.306	13.2

For $S^3 \times S^2$ the bound 3 persists up to 10^4 restarts, and the exact value 2 is reached after 10^5 restarts (57.6 s on CUDA). The CUDA implementation is 13–30 times faster than the CPU version.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR31714735).

Keywords: simplicial Lusternik–Schnirelmann category, simplicial geometric category, strong collapse, strong core, set cover, randomized algorithm, CUDA.

2020 Mathematics Subject Classification: 55M30, 55U10, 68W20

References

[1] Lusternik L., Schnirelmann L. Sur le problème de trois géodésiques fermées sur les surfaces de genre 0, *C. R. Acad. Sci. Paris* **189** (1929), 269–271.

[2] Fox R. H. On the Lusternik–Schnirelmann category, *Ann. of Math. (2)* **42** (1941), 333–370.

[3] Cornea O., Lupton G., Oprea J., Tanré D. Lusternik–Schnirelmann category. AMS, Providence. 2003.

[4] Fernández-Ternero D., Macías-Virgós E., Vilches J. A. Lusternik–Schnirelmann category of simplicial complexes and finite spaces, *Topology Appl.* **194** (2015), 37–50.

[5] Fernández-Ternero D., Macías-Virgós E., Minuz E., Vilches J. A. Simplicial Lusternik–Schnirelmann category, *Publ. Mat.* **63** (2019), 265–293.

[6] Barmak J. A., Minian E. G. Strong homotopy types, nerves and collapses, *Discrete Comput. Geom.* **47** (2012), 301–328.

[7] Lutz F. H. The Manifold Page: Surfaces. Technische Universität Berlin.

[8] Lutz F. H. Enumeration and random realization of triangulated surfaces, in: *Discrete Differential Geometry*, Oberwolfach Seminars **38**, Birkhäuser, Basel, 2008, 235–253.

[9] Milin L. A combinatorial computation of the first Pontryagin class of the complex projective plane, *Geom. Dedicata* **49** (1994), 253–291.

[10] Datta B. Minimal triangulations of manifolds, *J. Indian Inst. Sci.* **87** (2007), 429–449.

,

Convolution-type Operators in Grand Lorentz Sequence Spaces

M. MANARBEK[1], N.T. TLEUKHANOVA[2]

L. N. Gumilyov Eurasian National University, Astana, Kazakhstan

[1]manarbek@math.kz, [2]tleukhanova@rambler.kz

We introduce and study the discrete analogue of the grand Lorentz space recently defined by Nursultanov, Rafeiro and Suragan, namely the *grand Lorentz sequence space* $Gl^{\theta}_{p,q}$, where the aggrandisation is applied to the principal index p. We establish elementary properties and embeddings, a generalised Hölder inequality, and both Young's and O'Neil's inequalities for the discrete convolution in this new scale of spaces. As an application, we obtain boundedness results for a discrete power-logarithmic Riesz-type operator between grand Lorentz sequence spaces. **Funding**: This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23488549).

Keywords: non-linear differential equation, Lebesgue space, intermediate term, solvability, fixed point.

2020 Mathematics Subject Classification:34A34, 34B40, 34C11

Multipoint boundary value problems for impulsive hyperbolic equations with piecewise constant arguments

A. MOLYBAIKYZY

Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan
Mukhamedzhan Tynyshbayev ALT University, Almaty, Kazakhstan

altynaimolybai@gmail.com

Let $\Omega = [0,T] \times [0,\omega]$. In the domain Ω we consider the following boundary value problem for an impulsive hyperbolic equation with a generalized piecewise constant argument:

$$\frac{\partial^2 u}{\partial t\, \partial x} = A(t,x)\frac{\partial u(t,x)}{\partial x} + B(t,x)\frac{\partial u(t,x)}{\partial t} + C(t,x)u(t,x) + A_0(t,x)\frac{\partial u(\gamma(t),x)}{\partial x} + B_0(t,x)\frac{\partial u(\gamma(t),x)}{\partial t} + C_0(t,x)u(\gamma(t),x) + f(t,x), \quad t \neq \theta_j, \quad j = \overline{1,N-1}, \tag{1}$$

$$P_0(x)\frac{\partial u(0,x)}{\partial x} + \sum_{s=1}^{N-1} P_s(x)\frac{\partial u(\zeta_s,x)}{\partial x} + P_N(x)\frac{\partial u(T,x)}{\partial x} = \varphi_0(x), \quad x \in [0,\omega], \tag{2}$$

$$\lim_{t\to\theta_p+0} \frac{\partial u(t,x)}{\partial x} - \lim_{t\to\theta_p-0} \frac{\partial u(t,x)}{\partial x} = \varphi_p(x), \quad x \in [0,\omega], \quad p = \overline{1,N-1}, \tag{3}$$

$$u(t,0) = \psi(t), \quad t \in [0,T]. \tag{4}$$

Here $\gamma(t) = \zeta_s$ for $t \in [\theta_{s-1}, \theta_s)$, $s = \overline{1,N}$, where

$$0 = \theta_0 < \theta_1 < \ldots < \theta_{N-1} < \theta_N = T, \qquad \theta_{s-1} < \zeta_s < \theta_s, \quad s = 1,2,\ldots,N.$$

Here $u(t,x)$ is the unknown function; the functions $A(t,x)$, $B(t,x)$, $C(t,x)$, $A_0(t,x)$, $B_0(t,x)$, $C_0(t,x)$ and $f(t,x)$ are continuous on Ω; the functions $P_0(x)$, $P_s(x)$, $s = \overline{1,N-1}$, and $P_N(x)$ are continuous on $[0,\omega]$; the functions $\varphi_0(x)$ and $\varphi_p(x)$, $p = \overline{1,N-1}$, are continuously differentiable on $[0,\omega]$; and $\psi(t)$ is continuously differentiable on $[0,T]$.

The problem under consideration combines impulsive effects, a generalized piecewise constant argument and multipoint boundary conditions. Problems of this type arise in the mathematical modeling of processes in which continuous dynamics, discrete updating of information and instantaneous control actions occur simultaneously.

Our approach is based on D. S. Dzhumabaev's parametrization method. The original boundary value problem is reduced to an equivalent problem consisting of a family of Cauchy problems with parameters and a system of functional equations with respect to the introduced functional parameters. Sufficient conditions for the existence and uniqueness of a solution are established in terms of the invertibility of the resulting matrix of parameters. In addition, we propose a constructive iterative algorithm that allows the solution to be constructed step by step.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26194447).

Keywords: impulsive hyperbolic equation, piecewise constant argument, multipoint boundary value problem, parametrization method, algorithm.

2020 Mathematics Subject Classification: 35L20, 34K45, 34B10.

References

[1] Imanchiyev A.E., Assanova A.T., Molybaikyzy A. Properties of a nonlocal problem for hyperbolic equations with impulse discrete memory, Lobachevskii Journal of Mathematics, 44(10) (2023), 4299–4309.

[2] Molybaikyzy A., Yuldashev T.K. Multipoint boundary value problems for impulsive hyperbolic equations with piecewise constant arguments, Journal of Mathematics, Mechanics and Computer Science, 3(131) (2026), 86–100.

[3] Molybaikyzy A., Kabdrakhova S.S., Iskakova N.B., Minglibayeva B.B. On the solution of a boundary value problem for an impulsive hyperbolic equation with a piecewise-constant argument, Herald of the Kazakh-British Technical University, 23(2) (2026), 61–74.

Generative Artificial Intelligence for Strategic University Leadership

A. MOSAVI[1], A. IGENBAYEV[2]

[1,2]*Obuda University, Budapest, Hungary*

amir.mosavi@nik.uni-obuda.hu

University leaders face increasing challenges in strategic planning, resource management, research development, and academic quality assurance. Current leadership practices often depend on fragmented reports, manual analysis, and delayed information, which limits timely and evidence based decision making. This paper proposes a domain specific Generative Artificial Intelligence (GAI) based Decision Support System (DSS) for strategic university leadership. The proposed system uses a specialized Large Language Model (LLM) trained on institutional data, academic publications, policy documents, and global higher education benchmarks. The system supports decision making in academic operations, research management, financial planning, human resources, and institutional development. Recent studies demonstrate the potential of AI and big data integration for improving higher education management and strategic evaluation [1]. Generative AI also provides new opportunities for sustainability analysis, operational optimization, and institutional planning [2]. Furthermore, GAI supports innovation in educational systems and administrative processes by enhancing data analysis and knowledge generation [3,4,5]. The main contribution of this research is the development of an executive AI approach designed specifically for university governance, enabling real time recommendations, predictive insights, and adaptive strategic planning. The proposed system also considers responsible AI requirements, including data privacy, transparency, ethical governance, and institutional readiness. This research provides a foundation for intelligent university leadership in the digital era.

References

1. Kadhem, H. (2025). Comprehensive Framework for Artificial Intelligence and Big Data Integration in Higher Education. Communications in Computer and Information Science.

2. Hundekari, S., Praveen, R.V.S., Shrivastava, A., et al., Lakhanpal, S. (2026). Sustainability in Campus Operations: Data Driven Evaluation of College Initiatives. In Generative AI and Sustainability in Higher Education: Applications, Opportunities, and Challenges.

3. Sharma, P.B., Kathuria, P., Aggarwal, B., Pathak, R. (2026). AI for Education: Augmenting Pedagogy, Creativity, and Institutional Learning Systems. In AI in Public Policy and Business: Strategies, Ethics, and Impact.

4. Sharma, S., Aggarwal, M., Kaur, H., et al., Al-Said, K. (2026). Generative AI in Higher Education: Advancing Pedagogical Innovation and Addressing Emerging Challenges. In Generative AI and Sustainability in Higher Education: Applications, Opportunities, and Challenges.

5. Gering, Z., Feher, K., Harmat, V., Tamassy, R. (2025). Strategic Organisational Responses to Generative AI Driven Digital Transformation in Leading Higher Education Institutions. International Journal of Organizational Analysis.

LMMs for Higher Education Scientific Strategy Planning and Evaluation

A. MOSAVI[1], B. SABDENALIYEV [2]

[1,2]*Obuda University, Budapest, Hungary*

amir.mosavi@nik.uni-obuda.hu

Strategic planning and evaluation are essential for improving scientific performance and institutional development in higher education. However, current evaluation practices often rely on manual assessment, retrospective analysis, and static indicators, which limit real time monitoring and future prediction. This paper proposes an intelligent approach based on Large Language Models (LLMs) and machine learning (ML) for scientific strategy planning and evaluation in higher education. The proposed approach integrates academic data analysis, research performance indicators, faculty evaluation, resource utilization, and institutional benchmarking through supervised, unsupervised, and hybrid learning methods. LLMs are applied to support knowledge extraction, academic evaluation, strategic analysis, and scenario generation. Recent studies show that LLMs can improve educational analytics, personalized learning, and academic decision support [1][2]. They also provide new opportunities for automated research evaluation and scalable assessment processes [3]. Furthermore, LLM based assessment methods demonstrate potential for improving feedback quality and reducing evaluation workload [4]. The main contribution of this paper is the development of an integrated LLM and ML approach for dynamic scientific strategy evaluation in higher education. The proposed methodology supports evidence-based decision making, identifies institutional improvement opportunities, and enables adaptive planning. Responsible adoption of LLMs is also considered through transparency, privacy protection, bias reduction, and governance requirements.

References

1. Jothikumar, K., Velusamy, G., Meenakshi, A.J., et al., Anisha, S. (2026). Design and Implementation of an AI Integrated Educational Assistant for Learning. SN Computer Science.

2. Galdames, I.C.S., Chaple-Gil, A.M. (2026). From Algorithms to Adaptation: How Generative AI Is Reshaping Personalized Learning in Higher Education. In Artificial Intelligence in Higher Education: Generative AI, Personalized Learning, Digital Transformation.

3. Cheng, X., Li, J., Yang, J., Yan, Z. (2024). Exploring the Application of Large Language Models in Academic Paper Evaluation. Library and Information Service.

4. Yeung, C., Yu, J., Cheung, K.C., et al., Fujii, K. (2025). A Zero-Shot LLM Framework for Automated Assignment Grading in Introductory Higher Education Courses. Communications in Computer and Information Science.

On the weighted integrability of sums of series with respect to the Price multiplicative system with GM and bounded variation coefficients

M.A. MUKHAMBETZHAN[1], M.Zh. TURGUMBAYEV[2], G.S. KENZHEBEKOVA[3]

Astana IT University, Astana, Kazakhstan

[1]*Astana IT University, Astana, Kazakhstan*

[2]*Karaganda Buketov University, Karaganda, Kazakhstan*

[3]*L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*

[1]m.mukhambetzhan@astanait.edu.kz, [2]mentur60@mail.ru, [3]gulmira0703@yandex.kz

In this abstract we consider the problem of weighted inequalities for Fourier series with respect to the multiplicative systems with generalized monotonic coefficients. Let $\mathbb{Z}$ be the set of integers. Let $\{p_k\}_{k=1}^{\infty}$ be a sequence of natural numbers $p_k \geq 2,\ k \in \mathbb{N}$, with $\sup\limits_k p_k = N < \infty$. The system $\{\psi_n\}_{n=0}^{\infty}$, known as the Price system, is a complete orthonormal system in $L^1(0,1)$ [1]. If $p_k = 2$ for all k, then $\{\psi_n\}_{n=0}^{\infty}$ coincides with the Walsh system in the Paley ordering. Let $1 \leq p < \infty$, $\frac{1}{p} + \frac{1}{p'} = 1$, and $\omega(x) \geq 0$ be some locally integrable function on $[0,1]$.

Definition 1. A sequence of complex numbers $a = \{a_n\}_{n=1}^{\infty}$ is called generalized monotone (denoted by $a \in GM$) [2] if

$$\sum_{v=n}^{2n-1} |a_v - a_{v+1}| \leq C|a_n|$$

for every integer n, where the constant C is independent of n.

Theorem 2. *Let $a = \{a_n\}_{n=1}^{\infty} \in GM$, $\sup_n p_n = K < \infty$, $\sum_{n=1}^{\infty} \frac{|a_n|}{n} < \infty$. Then the series $\sum_{k=0}^{\infty} a_k \psi_k(x)$ converges for every x, except possibly at the points $x = 0$ and $k \in \mathbb{Z}$, and converges uniformly on every interval $[\varepsilon, 1-\varepsilon]$, where $0 < \varepsilon < 1$.*

Moreover, for every $x \in (0,1)$ and $n \geq 1/x$, the following estimate holds:

$$|f(x) - S_n(f,x)| \leq \frac{C}{x}\left(|a_n| + \sum_{v=n+1}^{\infty} \frac{|a_v|}{v}\right).$$

Theorem 3. *Let $1 \leq p < \infty$, $\sup p_n = K < \infty$, $\lambda_k \in GM$, and $f(x) = \sum_{k=0}^{\infty} \lambda_k \psi_k(x)$.*

Assume that the weight function ω satisfies $\omega \in RB_p \cap \overline{M}_p$.

Then the following estimate holds:

$$\|f\|_{L^p_\omega(0,1)} \leq C\left(\sum_{n=1}^{\infty} B_n^p \int_{\frac{1}{n+1}}^{\frac{1}{n}} \frac{\omega(x)}{x^p}\, dx\right)^{1/p},$$

where $B_n = \sum_{k=\lfloor n/\gamma \rfloor}^{\infty} \frac{|\lambda_k|}{k}$.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26196065).

Keywords: multiplicative systems, weighted Lebesgue spaces, generalized monotone sequences, Hardy operator, integrability conditions.

2020 Mathematics Subject Classification: 35Q79, 35K05, 35K20

References

[1] Golubov B.I., Efimov A.V., Skvortsov V.A. *Walsh Series and Transforms: Theory and Applications*,Moscow, Nauka, 1987

[2] Tikhonov S.Yu. *Trigonometric Series with General Monotone Coefficients*, *J. Math. Anal. Appl.*, **326**:1 (2007), 721–735.

Spectral Estimates for a Mixed Type Differential Operator

M. MURATBEKOV[1], K. JARBULOVA[2], A. ABYLAYEVA[3]

[1]*M.H. Dulaty Taraz University, Taraz, Kazakhstan*
[2,3]*L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*
abylayevab@mail.ru

In the case of a bounded domain, depending on the boundary conditions and the geometry of the domain, the spectral properties of operators of mixed type were studied in papers [1–2] and the papers cited there. In applications related to parabolic-hyperbolic heat conduction and equations with a spectral parameter, such operators naturally arise on unbounded domains, where the coefficient determining the type of the operator may be an arbitrary, possibly rapidly oscillating, function, while the lower-order coefficients may grow unboundedly at infinity. Under these conditions, the classical requirement of uniform ellipticity/hyperbolicity fails, which makes it relevant to build a solvability theory and sharp spectral estimates for this class of operators.

The aim of the work is to construct a solvability theory (invertibility, separability, resolvent compactness) and to obtain sharp two-sided estimates for the s-numbers and eigenvalues of the operator

$$Lu = k(y)u_{xx} - u_{yy} + b(y)u_x + q(y)u \tag{1}$$

on the cylinder $\Omega = \{|x| \le \pi,\ y \in \mathbb{R}\}$ with periodic conditions in x, where $k(y)$ is an arbitrary piecewise continuous (including oscillating) function and $b(y)$, $q(y)$ are unbounded as $|y| \to \infty$.

The continuous invertibility of $L + \mu I$ is proved, and its resolvent is represented as a series over the resolvents of l_n. Under slow-variation conditions on b, q and the relation $q \le C_0 b^2$, separability of the operator is established. A compactness criterion for the resolvent is obtained. A two-sided estimate for the counting function $N(\lambda)$ of singular numbers is derived in terms of the level-set measures of $Q_n(y)$, $K_n(y)$. For a model example ($k(y) = \sin(e^{10|y|})$, $b(y) = q(y) = |y| + 1$), explicit power-type two-sided s-number estimates are obtained, and, via Weyl's inequality, eigenvalue estimates.

The theoretical component of the work is based on the research of the work [3]. Throughout this note we mainly use techniques from the paper [4].

References

[1] Bitsadze A.V. *Equations of Mixed Type*, Moscow (1959).

[2] Kalmenov T.S. On spectrum of the Tricomi problem for the Lavrent'ev–Bitsadze equation, *Differ. Equ.*, **13**:8 (1977), 14–18.

[3] Otelbaev M. Coercive estimates and separability theorems for elliptic equations in $\mathbb{R}^n$, *Trudy Mat. Inst. Steklov.*, **161** (1983), 195–217.

[4] Muratbekov M., Abylayeva A., Muratbekov M. Estimates of singular numbers (s-numbers) and eigenvalues of a mixed elliptic-hyperbolic type operator with parabolic degeneration, *Math. Meth. Appl. Sci.* (2022), 1–13. DOI: 10.1002/mma.8908.

Exhauster-Based Nonsmooth Analysis for Optimization Problems in Machine Learning

G.Y. MURZABEKOVA

QAIRU, Astana, Kazakhstan

guldenmur07@gmail.com

Modern machine learning is fundamentally based on mathematical optimization. A typical learning problem consists of minimizing a loss function with respect to the model parameters $\theta \in \Theta \subseteq \mathbb{R}^m$, while controlling model complexity and seeking good generalization and robustness. Although machine learning generally favors simpler, more parsimonious models, regularization, constraints, and piecewise-defined loss functions may lead to nonsmooth and nonconvex optimization problems.

This work investigates the potential of upper and lower exhausters and convexificators from nonsmooth analysis as a mathematical framework for studying such optimization problems. The approach is based on representing the directional behavior of nonsmooth functions in terms of families of convex compact sets and on reducing certain nonsmooth problems to structured convex subproblems. Particular attention is paid to the sensitivity and stability of solutions defined implicitly by nonsmooth systems. As an application, the framework is considered for wheat yield prediction using machine learning and climatic, soil, and agronomic data from Northern Kazakhstan.

Training a machine learning model can be formulated as an optimization problem

$$\min_{\theta \in \Theta} F(\theta), \tag{1}$$

where θ denotes the vector of model parameters, $\Theta \subseteq \mathbb{R}^m$ denotes the parameter space, and F is an objective function combining predictive error with a regularization term:

$$F(\theta) = \frac{1}{N} \sum_{i=1}^{N} L\big(y_i, f(x_i;\theta)\big) + \lambda R(\theta), \tag{2}$$

where $L(\cdot,\cdot)$ is a per-sample loss function (e.g., squared or absolute error), $R(\theta)$ is a regularization penalty (e.g., an ℓ_1- or ℓ_2-norm penalty), and $\lambda \geq 0$ is a regularization parameter.

An important principle of machine learning is to control model complexity in order to achieve good generalization. However, simpler models do not necessarily lead to simpler optimization problems. For example, ℓ_1-regularization introduces the nonsmooth term

$$\|\theta\|_1 = \sum_{j=1}^{m} |\theta_j|,$$

so that

$$F(\theta) = \mathcal{L}(\theta) + \lambda \|\theta\|_1$$

is not differentiable at points where some components of θ vanish.

This distinction between model complexity and optimization complexity motivates the use of mathematical tools for nonsmooth functions. Upper and lower exhausters and convexificators provide such a framework. These objects provide geometric representations of nonsmooth functions and can be used in nonsmooth optimization, where certain problems can be reduced to structured convex subproblems [1].

Convexificators and Exhausters. Let $h : \mathbb{R}^n \to \mathbb{R}$ be a continuous positively homogeneous function (in particular, h may represent the directional derivative $F'(\theta;\cdot)$ of a nonsmooth objective function at a fixed θ).

Definition 1. A convex compact set $C \subset \mathbb{R}^n$ is called an *upper convexificator* of h if

$$h(g) \le \max_{v\in C}\langle v, g\rangle, \qquad g \in \mathbb{R}^n, \tag{3}$$

and a *lower convexificator* if

$$h(g) \ge \min_{w\in C}\langle w, g\rangle, \qquad g \in \mathbb{R}^n. \tag{4}$$

A set satisfying both (3) and (4) is called a *convexificator* of h [1, 2].

The defining inequalities (3)–(4) provide lower concave and upper convex approximations of a nonsmooth function by means of a convex compact set. In the original construction, such sets arise as limiting objects (e.g., via upper and lower Dini directional derivatives at nearby points) [1].

Definition 2. For a continuous positively homogeneous function h, an *upper exhauster* $E^+(h)$ provides the representation

$$h(g) = \min_{C\in E^+(h)} \max_{v\in C}\langle v, g\rangle, \tag{5}$$

while a *lower exhauster* $E^-(h)$ provides

$$h(g) = \max_{C\in E^-(h)} \min_{w\in C}\langle w, g\rangle. \tag{6}$$

Thus, the local behavior of a nonsmooth function can be described by families of convex compact sets rather than by a single classical gradient. Such minimax representations provide a geometric description of directional behavior and can be used to derive optimality conditions and identify descent directions [1, 3].

Convexificators and Implicit Functions. A second important direction of the theory concerns nonsmooth systems

$$f(x, y) = 0, \tag{7}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^k$, and the problem is to determine whether y can be represented locally as a function of x.

In the smooth case, the classical implicit function theorem is based on the nonsingularity of the Jacobian with respect to y. In the nonsmooth case, the Jacobian is replaced by an appropriate set-valued construction based on convexificator mappings.

Sufficient conditions for the existence of implicit functions in nonsmooth systems can be formulated in terms of convexificator mappings [2]. These conditions apply to both fully determined and underdetermined systems and are expressed in terms of $k \times k$ matrices constructed from elements of the convexificators.

A principal condition has the form

$$|\det A| \ge \alpha_0 > 0 \quad \text{for all } A \in B(x_0, y_0), \tag{8}$$

where $B(x_0, y_0)$ is the compact set of $k \times k$ matrices constructed from the corresponding convexificator mappings at (x_0, y_0). Under appropriate upper semicontinuity assumptions, condition (8) is sufficient for the existence of an implicit function and, under additional assumptions, for its uniqueness and Lipschitz continuity [4].

These results provide a mathematical bridge to optimization problems in machine learning, where optimal model parameters can themselves be viewed as solutions of parameter-dependent optimization systems.

Exhausters and Sensitivity of Machine Learning Models. To study sensitivity, it is convenient to make the dependence of F on the training sample explicit. Let $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$ denote the training data, and write $F(\theta; \mathcal{X})$ for the objective in (2) viewed as a function of both θ and $\mathcal{X}$. Suppose, for simplicity, that the optimal parameters of a machine learning model are defined by

$$\theta^*(\mathcal{X}) = \arg\min_{\theta} F(\theta; \mathcal{X}). \tag{9}$$

Changes in the data $\mathcal{X}$, regularization parameters, or other model conditions may lead to changes in $\theta^*(\mathcal{X})$. Hence, an important question is whether the optimal parameters depend continuously on perturbations of the training data and remain stable under such perturbations.

For smooth problems, this type of sensitivity analysis is closely related to the implicit function theorem. For nonsmooth problems, however, classical differentiability may fail. The use of upper and lower exhausters for nonsmooth implicit functions allows the problem of finding an implicit function to be reduced to a quasilinear system involving exhauster representations [3].

For the directional derivative $F'(\theta; d)$ of $F(\cdot; \mathcal{X})$ at θ, an upper-exhauster representation may be written as

$$F'(\theta; d) = \min_{C \in E^+\left(F'(\theta;\cdot)\right)} \max_{v \in C} \langle v, d \rangle. \tag{10}$$

Consequently, a direction d satisfying

$$F'(\theta; d) < 0 \tag{11}$$

is a direction of local decrease.

This suggests a research perspective in which exhausters are used not simply to replace gradients, but to characterize the local geometry, sensitivity, and stability of optimization problems arising in machine learning.

Application to Wheat Yield Prediction. As an application, we consider wheat yield prediction in Northern Kazakhstan. Let

$$x = (T, P, S, M, A, \dots), \tag{12}$$

where T, P, S, M, and A denote temperature-related variables, precipitation, soil characteristics, moisture-related variables, and agronomic parameters, respectively. The predicted yield is

$$\widehat{Y} = f(x; \theta). \tag{13}$$

Among the models considered, linear regression, ℓ_1- and ℓ_2-regularized regression, and support vector regression are naturally formulated in terms of a finite-dimensional continuous parameter vector θ and can be trained by minimizing an objective of the form

$$\theta^* = \arg\min_{\theta} \left[\frac{1}{N} \sum_{i=1}^{N} \big(y_i - f(x_i; \theta)\big)^2 + \lambda R(\theta) \right]. \tag{14}$$

For these models, the exhauster/convexificator framework developed in (3)–(11) can be applied to the corresponding objective functions and their optimal solutions.

Random Forest and LightGBM, which in previous experiments provided the best predictive performance among the considered models (MAPE ≈ 0.155 and 0.156, respectively), are tree-based ensemble methods that use recursive partitioning and boosting, respectively, rather than continuous minimization over θ as in (14). We therefore treat them as predictive benchmarks; extending the nonsmooth-analysis framework to such models would require a separate formalization of the relevant nonsmooth structure (e.g., the split-selection objective) and is left as a direction for future work, distinct from the θ-based sensitivity analysis developed above for the regression-type models.

These empirical results provide an application context for further mathematical analysis. In particular, the following questions arise: How stable is the optimal solution θ^* with respect to perturbations of the climatic and soil variables entering $\mathcal{X}$? Under what conditions can the optimal parameters be characterized by an implicit nonsmooth system? Can convexificators or exhausters provide useful sufficient conditions for the stability or local sensitivity of $\theta^*(\mathcal{X})$ for the regularized regression models considered above?

These questions establish a connection between the mathematical theory developed in [1, 3, 4] and contemporary machine learning applications.

The increasing use of machine learning does not eliminate the need for mathematical analysis of optimization problems. On the contrary, regularization, sparsity, constraints, and piecewise-defined model components can produce nonsmooth optimization structures even when the resulting model is deliberately kept simple.

Upper and lower exhausters and convexificators provide a geometric framework for describing such structures. Convexificators establish upper and lower convex approximations of nonsmooth functions, while exhausters provide minimax representations of directional behavior [1]. Their application to nonsmooth implicit systems provides conditions for the existence, uniqueness, continuity, and Lipschitz properties of implicit solutions [3, 4].

Keywords: nonsmooth analysis, exhausters, convexificators, nonsmooth optimization, machine learning, artificial intelligence, implicit functions, wheat yield prediction.

2020 Mathematics Subject Classification: 90C26, 90C31, 90C56, 68T05

References

[1] Demyanov V.F. Exhausters and Convexificators – New Tools in Nonsmooth Analysis, in: *Quasidifferentiability and Related Topics*, Springer, Boston, MA (2000), 85–137.

[2] Demyanov V.F., Jeyakumar V. Hunting for a smaller convex subdifferential, *Journal of Global Optimization*, **10**:3 (1997), 305–326.

[3] Murzabekova G.Y. Exhausters and implicit functions in nonsmooth systems, *Optimization*, **59**:1 (2010), 105–113.

[4] Demyanov V.F., Murzabekova G.E. Convexificators and implicit functions in nonsmooth systems, *Computational Mathematics and Mathematical Physics*, **39**:2 (1999), 212–223.

Weak Well-Posedness and Perturbation Analysis for a Space-Dependent Variable-Order Fractional Diffusion Equation

M. OMAROV[1], M. RAMAZANOV[2], E. KHALMETOV[3]

[1]*Buketov Karaganda National Research, 28 Universitetskaya Street, Karaganda, Kazakhstan*
[2]*Buketov Karaganda National Research, 28 Universitetskaya Street, Karaganda, Kazakhstan*
[3]*Kyoto University of Advanced Science, 18 Yamanouchi Gotanda-cho, Ukyo-ku, Kyoto, Japan*

[1]madiomarovt@gmail.com, [2]muratr124@gmail.com, [3]halmetovelnur@gmail.com

Let $Q_T = (0,1) \times (0,T)$, $H = L^2(0,1)$, $V = H_0^1(0,1)$, and $V^* = H^{-1}(0,1)$. We consider the Dirichlet problem

$$\partial_t^{\alpha(x)} u - u_{xx} = f \quad \text{in } Q_T, \qquad u(0,t) = u(1,t) = 0, \qquad u(x,0) = u_0(x), \tag{1}$$

where $\alpha \in C([0,1])$ and $0 < \alpha_0 \le \alpha(x) \le \alpha_1 < 1$. The variable-order Gerasimov–Caputo operator is represented by the Volterra kernel $k(x,t) = t^{-\alpha(x)}/\Gamma(1-\alpha(x))$. A weak solution satisfies $u \in L^2(0,T;V)$, $k * (u-u_0) \in H^1(0,T;V^*)$, $(k*(u-u_0))(0) = 0$, and

$$\big\langle \partial_t\big(k*(u-u_0)\big)(t), \varphi\big\rangle + (u_x(t), \varphi_x) = (f(t), \varphi), \qquad \varphi \in V,$$

for almost every $t \in (0,T)$.

The principal estimate is the coercivity inequality

$$\int_0^T \langle \partial_t (k*w)(t), w(t)\rangle \, dt \ge \frac{\mu_T}{2} \|w\|^2_{L^2(Q_T)}, \qquad \mu_T := \min_{a\in[\alpha_0,\alpha_1]} \frac{T^{-a}}{\Gamma(1-a)}. \tag{2}$$

Its weighted form contains the coefficient $\omega_T(x) = T^{-\alpha(x)}/\Gamma(1-\alpha(x))$. These inequalities, together with a variational approximation argument, imply existence, uniqueness, and continuous dependence of the weak solution for $f \in L^2(Q_T)$ and $u_0 \in H_0^1(0,1)$:

$$\|u_1 - u_2\|_{\mathcal{E}(Q_T)} \le C\left(\|f_1 - f_2\|_{L^2(Q_T)} + \|u_{0,1} - u_{0,2}\|_{H_0^1(0,1)}\right), \quad \|v\|_{\mathcal{E}} := \|v\|_{L^2} + \|v_x\|_{L^2}.$$

For the affine order $\alpha(x) = \bar\alpha + \delta(x - \frac{1}{2})$, additional temporal regularity gives

$$u_\delta = u^{(0)} + \delta U + R_\delta, \qquad \|R_\delta\|_{\mathcal{E}(Q_T)} \le C\delta^2,$$

where $u^{(0)}$ solves the constant-order problem of order $\bar\alpha$. If $f = 0$ and $u_0(x) = \sin(n\pi x)$, the Fourier sine coefficients of U vanish whenever $m+n$ is even; in particular, the coefficient with $m = n$ is zero. Hence the first-order correction contains only eigenfunctions of the Dirichlet Laplacian whose indices have parity opposite to that of n. A nonconstant spatial order also precludes the classical separation-of-variables reduction.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23488729).

Keywords: space-dependent variable order, Gerasimov–Caputo derivative, weak solution, coercivity estimate, continuous dependence, Fourier sine expansion.

2020 Mathematics Subject Classification: 35R11, 35K05, 45K05.

References

[1] Y. Kian, E. Soccorsi, M. Yamamoto, On time-fractional diffusion equations with space-dependent variable order, *Ann. Henri Poincaré*, **19** (2018), 3855–3881.

[2] K. Van Bockstal, Existence of a unique weak solution to a non-autonomous time-fractional diffusion equation with space-dependent variable order, *Adv. Difference Equ.*, **2021** (2021), 314.

Generalized potential operators in weighted grand Herz spaces

A. ORAZBAYEV[1], M.A ZAIGHUM[2]

Nazarbayev University, Astana, Kzakhstan
[1]*Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan*
[2]*Zhubanov University (Herriot-Watt campus), Aktobe, Kazakhstan*
[1]almat.orazbayev@nu.edu.kz, [2]zaighum.asad@nu.edu.kz

In this talk, we extend two-weight norm inequalities for potential-type operators from weighted Herz spaces to weighted Grand Herz spaces. We define the homogeneous and inhomogeneous versions of the weighted Grand Herz spaces and establish the boundedness of potential-type operators under appropriate Muckenhoupt-type and reverse Hölder-type conditions on the weights. As a particular case, we show that Fefferman-Phong type condition ensures the boundedness of fractional integrals on weighted grand Herz spaces.

Funding: This research was funded by Nazarbayev University under Collaborative Research Program Grant o20122022CRP1601.

Keywords: Potential Operators, Grand Lebesgue spaces, Herz spaces.

2010 Mathematics Subject Classification: 46E30; 47B38

On the Solution of a Periodic Boundary Value Problem for a Hyperbolic Equation with a Riemann–Liouville Fractional Derivative with Respect to the Spatial Variable

N.T. ORUMBAYEVA[1], B.B. ZHANTASSOVA[2]

[1]*Buketov Karaganda National Research University, 28 Universitetskaya Street, Karaganda, Kazakhstan*
[2]*Buketov Karaganda National Research University, 28 Universitetskaya Street, Karaganda, Kazakhstan*
[1]orumbayevanurgul@gmail.com, [2]zh.botagoz.b@gmail.com

Let $\Omega = [0,T] \times [0,X]$. We consider the periodic boundary value problem

$$w_{tx} = A(t,x)w_t + B(t,x){}_0D_x^{\mu}w + f(t,x), \quad w(0,x) = \varphi(x), \quad w_t(t,0) = w_t(t,X), \tag{1}$$

where $0 < \mu < 1$, A, B are continuous matrix functions, and

$${}_0D_x^{\mu}w(t,x) = \frac{1}{\Gamma(1-\mu)}\frac{d}{dx}\int_0^x \frac{w(t,\xi)}{(x-\xi)^{\mu}}\,d\xi$$

is the Riemann–Liouville fractional derivative [1]. Setting $u = w_t$ gives $w(t,x) = \varphi(x) + \int_0^t u(\tau,x)\,d\tau$ and reduces (1) to an integro-differential equation for u with the periodic condition $u(t,0) = u(t,X)$ and $F(t,x) = B(t,x)\,{}_0D_x^{\mu}\varphi(x) + f(t,x)$.

Following the parameterization approach [2], the interval $[0,X]$ is partitioned into N subintervals of length $h = X/N$. For $u_r = u|_{[(r-1)h,rh)}$ we introduce $\lambda_r(t) = u_r(t,(r-1)h)$ and $v_r(t,x) = u_r(t,x) - \lambda_r(t)$. The original problem is then equivalent to a closed system consisting of a family of Cauchy-type integral equations for v_r and the parameter equation

$$Q(h)\lambda(t) = -W(t,h)\int_0^t \lambda(\tau)\,d\tau - V_1(t,h,v) - V_2(t,h,v) - \widehat{F}(t,h). \tag{2}$$

This representation yields an iterative algorithm: starting with $v^{(0)} = 0$, solve (2) for $\lambda^{(0)}$, determine $v^{(0)}$ by successive approximations, and repeat the two steps.

Assume that $Q(h)$ is invertible and $\|Q(h)^{-1}\| < \gamma(h)$. If

$$\begin{aligned} q(h) = h^2 \max\{1,h\}\left(\alpha + \frac{\beta T h^{1-\mu}}{\Gamma(2-\mu)}\right)^2 \times \\ \exp\left[h\gamma(h)\max\{1,h\}\frac{\beta T h^{1-\mu}}{\Gamma(2-\mu)} + h\left(\alpha + \frac{\beta T h^{1-\mu}}{\Gamma(2-\mu)}\right)\right] < 1, \end{aligned} \tag{3}$$

then the boundary value problem has a unique solution w^* and the constructed approximations converge geometrically. In particular,

$$\|w^* - w^{(k)}\| \le C(h,\mu,T)\frac{q(h)^k}{1-q(h)}\,M(h)\|f\|_1.$$

Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23488729, 2024–2026).

Keywords: hyperbolic equation, Riemann–Liouville fractional derivative, parameterization method, integro-differential equations, Gronwall–Bellman inequality, boundary value problem.

2020 Mathematics Subject Classification: 34B05, 35A05, 35A20, 35A22.

References

[1] S.G. Samko, A.A. Kilbas, O.I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications*, Gordon and Breach Science, 1993.

[2] D.S. Dzhumabayev, Criteria for the unique solvability of a linear boundary-value problem for an ordinary differential equation, *USSR Computational Mathematics and Mathematical Physics*, **29**(1) (1989), 34–46.

,

Coercive solvability conditions for a nonlinear third-order singular differential equation in Lebesgue space

K.N. OSPANOV[1], Zh.B. YESKABYLOVA[2]

L. N. Gumilyov Eurasian National University, Astana, Kazakhstan

[1]kordan.ospanov@gmail.com, [2]julie92@mail.ru

The work is devoted to the study of the following nonlinear differential equation.

$$Ly = -y''' + q(x,y)y' + s(x,y)y = f(x), \tag{1}$$

where q is continuously differentiable, s is continuous, $x \in \mathbb{R} = (-\infty, +\infty)$, and $f \in L_p = L_p(\mathbb{R})$ $(1 < p < \infty)$. We give sufficient conditions for the existence of a solution to equation (1) and prove the following relation

$$\|y'''\|_p + \|q(x,y)y'\|_p + \|(|s(x,y)| + 1)y\|_p < \infty.$$

Note that third-order differential equations arise in many practical problems. Equation (1) with an unbounded intermediate coefficient describes interactions of elastic bodies, processes in media with memory, incoming-flow effects in hydrodynamics, and propagation of electrical impulses in living organisms (Hodgkin–Huxley and Nagumo models). We assume that the intermediate coefficient q is a real-valued function independent of the lower-order coefficient s. This case is typical, for example, for a stationary Korteweg–de Vries type equation. Previously, the nonlinear equation (1) had been investigated only for the very special case $p = 2$. Although the solution to a singular differential equation of odd order is smooth, it may not belong to a Sobolev space. These and other facts indicate the importance of studying equation (1).

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23488049).

Keywords: non-linear differential equation, Lebesgue space, intermediate term, solvability, fixed point.

2020 Mathematics Subject Classification:34A34, 34B40, 34C11

On the inverse problem for a degenerate parabolic equation with the Hadamard–Caputo derivative under Samarskii–Ionkin conditions

M. SADYBEKOV[1], G. DILDABEK[2]

[1]*Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan*
[2]*Al-Farabi Kazakh National University, Almaty, Kazakhstan*
[1]sadybekov@math.kz, [2]dildabek@math.kz

Let $\Omega = \{(x,t) : 0 < x < 1,\ 0 < t < T\}$. In the domain Ω, we consider the inverse problem of determining the pair of functions $\{u(x,t), f(x)\}$. The main differential equation contains a degenerate fractional differential operator of the Hadamard–Caputo type with respect to the time variable t:

$$t^{-\beta} \cdot {}^{HC}D^{\alpha}_{0+,t}u(x,t) - u_{xx}(x,t) = f(x), \quad (x,t) \in \Omega, \tag{1}$$

where $\beta > 0$ is a degeneracy parameter, and ${}^{HC}D^{\alpha}_{0+,t}$ is the Hadamard–Caputo fractional derivative of order $\alpha \in (0,1]$, defined as follows:

$${}^{HC}D^{\alpha}_{0+,t}g(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \left(\ln \frac{t}{\tau}\right)^{-\alpha} \left(\tau \frac{d}{d\tau}\right) g(\tau) \frac{d\tau}{\tau}.$$

In the case $\alpha = 1$ we assume that ${}^{HC}D^{1}_{0+,t}g(t) = tg'(t)$.

The following nonlocal boundary conditions are considered with respect to the spatial variable x:

$$\begin{cases} U_1(u) \equiv u_x(0,t) + u_x(1,t) + au(1,t) = 0, \\ U_2(u) \equiv u(0,t) + bu(1,t) = 0, \end{cases} \quad , \quad 0 \le t \le T, \tag{2}$$

where a and b are given real constants. We assume that $b \neq -1$ in order to ensure the compatibility of the boundary conditions, and we focus on the case $b \neq 1$, $a \neq 0$, which leads to a substantially more complicated spectral problem.

To determine the unknown source $f(x)$ and the solution $u(x,t)$, we impose the initial condition and the final overdetermination condition:

$$u(x,0) = \varphi(x), \quad 0 \le x \le 1, \tag{3}$$

$$u(x,T) = \psi(x), \quad 0 \le x \le 1. \tag{4}$$

Here $\varphi(x)$ and $\psi(x)$ are given sufficiently smooth functions consistent with the boundary conditions (2).

Definition 1. A pair of functions $\{u(x,t), f(x)\}$ is called a *strong generalized solution* of the inverse problem (1)–(4) if the following conditions are satisfied: $f(x) \in L_2(0,1)$, $u(x,t) \in L_2(\Omega)$, $u_{xx}(x,t) \in L_2(\Omega)$, and $t^{-\beta HC} D^{\alpha}_{0+,t} u(x,t) \in L_2(\Omega)$. Moreover, equation (1) holds almost everywhere in Ω, and conditions (2)–(4) are understood in the sense of traces of Sobolev functions.

The paper most closely related to the present study is [1], where an inverse source problem for a degenerate equation with a time-fractional Hadamard–Caputo derivative was investigated. In the cited work, Bitsadze–Samarskii type conditions were considered, where the value of the unknown function at the boundary is linked to its value at an interior point. The corresponding spectral problem is significantly non-self-adjoint; however, the authors succeeded in proving that its system of eigenfunctions possesses the basis property. This allows for the application of the Fourier method to solve the problem. In the present work, we consider nonlocal conditions of the Samarskii–Ionkin type, which lead to a system of eigenfunctions that does not form an unconditional basis. This necessitated the development of a specialized technique for constructing an auxiliary basis.

This report will present new scientific results obtained using artificial intelligence, which helped accelerate the research process. The authors are the sole authors of the main idea behind the work and the choice of research methods. Artificial intelligence was also used to refine the English language of this publication and for final editing.

The main mathematical result of this work is the following theorem.

Theorem 2. *Let the given functions $\varphi(x), \psi(x) \in W_2^2(0,1)$ and satisfy the boundary conditions (2). Then there exists a unique strong generalized solution $\{u(x,t), f(x)\}$ to the inverse problem (1)–(4).*

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR24993094).

Keywords: inverse problem; ultraslow diffusion equation; Hadamard–Caputo derivative; nonlocal conditions; Samarskii–Ionkin conditions; Riesz basis; spectral expansion.

2020 Mathematics Subject Classification: 35R30; 35R11, 35K20, 35P10, 35P15.

References

[1] Ashurov R.R., Kadirkulov B.J., Turmetov B.Kh. On the inverse problem of the Bitsadze–Samarskii type for a fractional parabolic equation, *Probl. Anal. Issues Anal.*, **12**:30 (2023), 20–40.

Direct and Inverse Problems for One-Dimensional Transient Heat Conduction in Multilayered Media

A.N. SATYBALDINA

L.N.Gumilyov Eurasian National University, Astana, Kazakhstan
[1]*Kazakh-British Technical University, Almaty, Kazakhstan*
a.satybaldina@kbtu.kz

Transient heat conduction in multilayered media is an important problem in mathematical modeling and thermal engineering, particularly when the thermophysical properties of individual layers must be identified from temperature measurements. Analytical and inverse approaches for determining thermal properties of multilayered media have been considered in previous studies, while inverse methods have also been applied to nonlinear heat-transfer problems.

In this study, a one-dimensional transient heat conduction problem in a three-layered radial medium is formulated and solved analytically. The considered domain is $r_0 \le r \le r_3$, where the thermal conductivity, density, and specific heat capacity are piecewise constant:

$$C(r) = \begin{cases} C_1, & r_0 \le r \le r_1, \\ C_2, & r_1 \le r \le r_2, \\ C_3, & r_2 \le r \le r_3, \end{cases} \qquad k(r) = \begin{cases} k_1, & r_0 \le r \le r_1, \\ k_2, & r_1 \le r \le r_2, \\ k_3, & r_2 \le r \le r_3. \end{cases} \tag{1}$$

In each layer, the temperature field $T^{(i)}(r,t)$ satisfies the radial transient heat conduction equation

$$C_i\rho_i \frac{\partial T^{(i)}}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(rk_i \frac{\partial T^{(i)}}{\partial r}\right), \qquad i = 1,2,3. \tag{2}$$

The inner boundary is prescribed by the Dirichlet condition

$$T(r_0,t) = T_0, \tag{3}$$

whereas heat exchange with the surrounding environment at the outer boundary is described by the convective condition

$$k_3 \frac{\partial T^{(3)}}{\partial r}\bigg|_{r=r_3} = -h\left[T^{(3)}(r_3,t) - T_a(t)\right]. \tag{4}$$

At the interfaces $r = r_1$ and $r = r_2$, continuity of temperature and heat flux is imposed:

$$T^{(i)}(r_j,t) = T^{(i+1)}(r_j,t), \qquad k_i \left.\frac{\partial T^{(i)}}{\partial r}\right|_{r=r_j} = k_{i+1} \left.\frac{\partial T^{(i+1)}}{\partial r}\right|_{r=r_j}, \qquad j = 1,2. \tag{5}$$

The analytical solution of the direct problem is constructed using the separation of variables and an eigenfunction expansion. The transient component is represented in terms of radial eigenfunctions expressed through Bessel functions of the first and second kinds. The corresponding eigenvalues are determined from characteristic equations obtained from the homogeneous boundary and interface conditions. The orthogonality properties of the eigenfunctions are then used to determine the expansion coefficients. This analytical construction follows the methodology developed in the corresponding sections of the study.

The obtained analytical solution is subsequently used to formulate the inverse problem. Given experimentally measured temperature values T_0, T_1, T_2, T_3 and the ambient temperature T_a, the unknown thermal conductivities

$$k_1, \quad k_2, \quad k_3$$

are determined from the interface conditions and the outer convective boundary condition.

The proposed approach provides an analytical framework for identifying the thermal properties of individual layers from transient temperature observations and extends the inverse analysis of multilayered media to a time-dependent radial heat-conduction problem.

Keywords: transient heat conduction, multilayered medium, radial coordinates, analytical solution, inverse problem, thermal conductivity, Bessel functions, parameter identification.

2020 Mathematics Subject Classification: 35K05, 35R30, 35Q79

References

[1] Umbricht G.F., Rubio D., Tarzia D.A. Determination of Thermal Conductivities in Multilayer Materials, *WSEAS Transactions on Heat and Mass Transfer*, **17**:188–195 (2022).

[2] Rysbaiuly B., Sinitsa A., Capsoni A. Analytical Inverse Analysis Methodological Approach for Thermo-Physical Parameters Estimation of Multilayered Medium Terrain with Homogenized Sampled Measurements, *Symmetry*, **14**:2248 (2022).

[3] Alpar S., Rysbaiuly B. Determination of Thermophysical Characteristics in a Nonlinear Inverse Heat Transfer Problem, *Applied Mathematics and Computation*, **440**:127656 (2023).

Automatic horse recognition in images using convolutional neural networks

K.M. TULENBAEV[1], R.A. KOSHZHANOV[2], M.N. ZHARYKOV[3], Z.K. KETTEBEKOV[4], A.A. KOVALENKO[5]

[1]*Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan*
[2]*Civil Aviation Academy, Almaty, Kazakhstan*
[3]*Satbayev University, Almaty, Kazakhstan*
[4]*Satbayev University, Almaty, Kazakhstan*
[5]*Satbayev University, Almaty, Kazakhstan*

[1]kasart4@mail.ru, [2]rasul87@list.ru, [3]cmalikz.h@gmail.com, [4]zhanaidar.kettebekov@gmail.com, [5]kovalenko.a@stud.satbayev.university

This article is devoted to the problem of automatic horse recognition in images using convolutional neural networks (CNN). During the research, image processing and classification methods were studied, and modern approaches in this field were analyzed. In the course of the work, a dataset was created, images were preprocessed and prepared for training. A neural network model was developed, and its training and testing stages were carried out. The results of the study showed that the proposed approach demonstrates good performance in the task of horse recognition, confirming its effectiveness and potential for practical application.

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR31714735).

Keywords: neural networks, convolution, backpropagation algorithm.

2020 Mathematics Subject Classification: 68T07, 68T05, 62M45

References

[1] Goodfellow I., Bengio Y., Courville A. *Deep Learning*, MIT Press, (2016).

[2] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton ImageNet Classification with Deep Convolutional Neural Networks, *Communications of the ACM*, **1**:2 (2017), 84–90.

Topological properties of products of finite groups

N.N. TURSUNGALIYEV[1], G.G. FAZYL[2], N.D. MARKHABATOV[3]

[1,2]*Department of Fundamental Mathematics,*

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

[3]*Department of Cryptology, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan*

[1]tursungaliyevnn@enu.kz, [2]fzhebei@gmail.com, [3]markhabatovnd@enu.kz

We define a profinite group to be a topological group which is isomorphic (as a topological group) to a projective limit of finite groups. Using the new formalism, Zoltan Kocsis [1] proved that every profinite group admits a finite approximation in the sense of Zilber [2]. The question naturally arises about the pseudofiniteness of profinite groups.

Definition 1 ([3])**.** A group G is called pseudofinite if it is an infinite model for the first-order theory of finite groups or, equivalently, if G is elementarily equivalent to an ultraproduct of finite groups.

The well-known Feferman–Vaught theorem implies that the direct product of finite and pseudofinite groups is also pseudofinite. In this paper, we show that the direct product of finite groups is profinite.

Theorem 2. *Let $\{G_i : i \in I\}$ be an arbitrary family of finite groups, each equipped with the discrete topology. Then the Cartesian direct product $G = \prod_{i\in I} G_i$, equipped with the product topology and coordinatewise group operations, is a profinite group.*

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP26194247).

Keywords: pseudofinite group, profinite group, approximations of theories, topological group.

2020 Mathematics Subject Classification: 35Q79, 35K05, 35K20

References

[1] Kocsis Z. Development of group theory in the language of internal set theory, *PhD Thesis*, 2019.

[2] Zilber B. Perfect infinities and finite approximation, *Infinity and Truth* (2014), 199–223. DOI: 10.1142/9789814571043

[3] Felgner U. Pseudo-endliche Gruppen, *Proceedings of the 8th Easter Conference on Model Theory (Humboldt-Universität, Berlin, 1990)*, 82–96.

Spectrally Stable Graph Neural Networks for Railway Delay Propagation

A.A. USSIPBAYEV[1], A.I. MUKHAMEDINA[2]

[1,2]*Astana IT University, Astana, Kazakhstan*

[1]Amirbek.Ussipbayev@astanait.edu.kz, [2]a.mukhamedina@astanait.edu.kz

Railway delays propagate as cascades through connected station networks. A disruption at one station transmits forward through platform occupancy, timetable conflicts, and capacity bottlenecks [1,2,3]. Ignoring cascade structure causes systematic underestimation of corridor reliability. This work proposes DDP-GCN (Directed Delay-Propagation Graph Convolutional Network), a physics-informed neural architecture for cascade delay forecasting. The main contribution is threefold: (1) DDP-GCN learns the directed transmission matrix $P(t)$ from edge features via row-normalized softmax, unlike standard GCNs using fixed symmetric adjacency [4,5]; (2) Theorem 1 and Corollary 2 provide an explicit forecast horizon, replacing empirical depth selection; (3) a spectral penalty enforces stability during training.

Let $G = (V, E)$ be a directed graph where V is the set of stations and E is the set of track segments. The delay state is

$$\mathbf{d}(t) = \big(d_1(t), \ldots, d_n(t)\big)^\top \in \mathbb{R}^n_+, \tag{1}$$

where $d_i(t) \geq 0$ is the accumulated delay at station i at time t. The cascade dynamics follow:

$$\mathbf{d}(t+1) = P\,\mathbf{d}(t) + \boldsymbol{\eta}(t+1), \tag{2}$$

where $P \in \mathbb{R}^{n\times n}$ is the non-negative delay transmission matrix, $\boldsymbol{\eta}(t)$ is the primary disturbance vector with $\mathbb{E}[\boldsymbol{\eta}(t)] = \boldsymbol{\mu}_\eta$. Here, $P\mathbf{d}(t)$ represents propagated delays, while $\boldsymbol{\eta}(t+1)$ represents new primary disruptions. The entry P_{ij} is the delay transmission coefficient from station j to station i per time step; the condition $\sum_j P_{ij} \leq 1$ reflects partial absorption of incoming delay by timetable reserves and recovery actions.

Theorem 1 (Spectral Stability)**.** *Let $\rho(P) < 1$. Then there exists a unique steady state $\mathbf{d}^* = (I-P)^{-1}\boldsymbol{\mu}_\eta$. Moreover, for every number r such that $\rho(P) < r < 1$, there exists a constant $C_r > 0$ such that*

$$\big\|\mathbb{E}[\mathbf{d}(t)] - \mathbf{d}^*\big\|_2 \leq C_r r^t. \tag{3}$$

This estimate shows how fast the influence of the initial delay decreases over time.

Since $\rho(P) < 1$, the matrix $(I - P)$ is invertible. Recursion on (2) gives $\mathbb{E}[\mathbf{d}(t)] = P^t\mathbf{d}(0) + (I - P)^{-1}(I - P^t)\boldsymbol{\mu}_\eta$. Subtracting $\mathbf{d}^*$ gives $\mathbb{E}[\mathbf{d}(t)] - \mathbf{d}^* = P^t(\mathbf{d}(0) - \mathbf{d}^*)$. By the spectral radius formula, for any r with $\rho(P) < r < 1$, $\|P^t\|_2 \le C_r r^t$, which yields (3).

Corollary 2. *For a tolerance $\varepsilon > 0$, the forecast horizon can be estimated as*

$$T_\varepsilon = \left\lceil \frac{\log(\varepsilon/C_r)}{\log r} \right\rceil, \quad \rho(P) < r < 1. \tag{4}$$

Network depth K should satisfy $K \ge T_\varepsilon$.

For a directed edge (v_j, v_i), let $\mathbf{r}_{ij}(t) = (l_{ij}, v_{ij}^{\max}, b_{ij}(t), q_{ij}(t))^\top$ collect segment length, speed limit, occupancy, and traffic-intensity or delay-correlation indicators. The time-dependent coefficient can be learned as $P_{ij}(t) = \sigma_P(\mathbf{w}_P^\top \mathbf{r}_{ij}(t) + b_P)$, where σ_P is row-normalized and includes a virtual "no-transmission" option. The DDP-GCN graph layer is:

$$\mathbf{H}^{(k+1)} = \sigma\Big(P(t)\,\mathbf{H}^{(k)}\mathbf{W}_{\text{prop}}^{(k)} + \mathbf{H}^{(k)}\mathbf{W}_{\text{self}}^{(k)}\Big), \tag{5}$$

where $\mathbf{H}^{(k)}$ is the matrix of node representations at layer k, $\mathbf{W}_{\text{prop}}^{(k)}$ and $\mathbf{W}_{\text{self}}^{(k)}$ are trainable weight matrices, the first term encodes cascade propagation, and the second term encodes local station dynamics. The spectral penalty

$$\mathcal{L}_{\text{spec}} = \max\big\{0,\ \rho(P) - (1 - \delta)\big\}^2, \quad \delta \in (0, 1), \tag{6}$$

discourages learned matrices whose spectral radius approaches the unstable region and enforces the condition of Theorem 1 during training [6]. A dual-head decoder produces joint point forecasts $\hat{\mathbf{d}}(t{+}1)$ and uncertainty estimates $\hat{\boldsymbol{\sigma}}^2(t{+}1)$ without post-hoc calibration.

On a synthetic five-node corridor with known ground-truth P^* ($\rho(P^*) = 0.553$, $T = 5000$ steps, split 70/15/15%), DDP-GCN recovers the spectral radius with 0.7% error versus 10.7% for a standard GCN baseline [4,5], and achieves 1.7× lower RMSE at the five-step horizon. The experiment uses synthetic data only; no confidential or operational data of any railway operator are used in this study. The analytical steady state $\mathbf{d}^* = (I - P^*)^{-1}\boldsymbol{\mu}_\eta$ matches the simulated mean with error 0.023 min, which is consistent with Theorem 1. These results indicate that learning directed transmission matrices with spectral guarantees improves both interpretability and stability of railway delay forecasts. Although the present study focuses on railway corridors, the same stability-aware graph learning principle can be adapted to other transport and logistics networks with cascade-like delay propagation.

Funding: The authors received no external funding for this research.

Keywords: railway delay propagation, graph neural networks, spectral stability, physics-informed learning, cascade dynamics, directed transmission matrix.

2020 Mathematics Subject Classification: 90B06, 90B15, 39A30, 68T07

References

[1] Goverde R.M.P. Railway timetable stability analysis using max-plus system theory, *Transportation Research Part B: Methodological*, **41**:2 (2007), 179–201.

[2] Büchel B., Spanninger T., Corman F. Empirical dynamics of railway delay propagation identified during the large-scale Rastatt disruption, *Scientific Reports*, **10** (2020), Article 18584.

[3] Monechi B., Gravino P., Di Clemente R., Servedio V.D.P. Complex delay dynamics on railway networks: from universal laws to realistic modelling, *EPJ Data Science*, **7** (2018), Article 35.

[4] Scarselli F., Gori M., Tsoi A.C., Hagenbuchner M., Monfardini G. The graph neural network model, *IEEE Transactions on Neural Networks*, **20**:1 (2009), 61–80.

[5] Kipf T.N., Welling M. Semi-supervised classification with graph convolutional networks, in: *International Conference on Learning Representations*, ICLR (2017).

[6] Raissi M., Perdikaris P., Karniadakis G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, **378** (2019), 686–707.

Context-Dependent Structures of Signs and a Mathematical Model of Meaning

V.V. VERBOVSKIY

Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan

verbovskiy@math.kz

We propose a mathematical framework in which meaning is not introduced as a primitive notion. Instead, the primary object is a dynamic structure of signs. Let V be a finite set of signs, where a sign may represent a word, an image, an algorithm, a mathematical construction, a rule, or even an entire theory. For every context $Y \subseteq V$, we introduce a context-dependent weight function $w_Y : V \times V \longrightarrow \mathbb{R}$, where $w_Y(a, b)$ represents the strength of the semantic relation from a to b in the context Y. Thus, a structure of meaning can be viewed as a family of context-dependent weighted directed graphs $S = (V, w_Y)_{Y \subseteq V}$, or equivalently as a family of generally non-symmetric matrices W_Y.

The central hypothesis is that the meaning of a sign is determined not by the sign itself but by its position in this structure. Consequently, understanding is modeled not as a logical inference or as an assignment of a fixed interpretation, but as a transition $S_1 \rightsquigarrow S_2$. Such a transition may change existing weights, create new relations, and introduce new signs. Learning is then represented as a sequence of acts of understanding.

Two different limitations are distinguished. The *depth* of a knowledge domain $A \subseteq V$ characterizes how large a context is intrinsically required to determine its semantic relations. In a simplified form, A has depth k if for every $B \subseteq A$ there exists $I \subseteq B$, $|I| \leq k$, such that $w_B(x_i, x_j) = w_I(x_i, x_j)$ for the relevant signs. In contrast, the *capacity* k_O of an observer is the maximal number of signs that the observer can simultaneously retain as context. Hence the complexity of a domain and the limitations of the system processing it are separated explicitly.

Communication between two agents is described locally. If the agents have structures S_A and S_B, communication activates substructures $H_A \subseteq S_A$ and $H_B \subseteq S_B$ and transforms them into H'_A and H'_B. Successful communication corresponds to increasing structural similarity between these local substructures. The resulting changes may subsequently propagate through the full structures.

The framework also suggests a possible extension of large language model training. Standard training updates model parameters from external data, whereas human reasoning also includes internal reorganization of already acquired knowledge. We therefore propose an additional *structural closure* phase between training stages:

$$\text{data} \longrightarrow \text{structure update} \longrightarrow \text{structural closure} \longrightarrow \text{next training stage}.$$

During structural closure no new external data are supplied. Instead, the system derives implicit relations, modifies existing ones, and may form new higher-level signs from stable substructures. Fuzzy, probabilistic, or continuous logics provide natural candidates for constructing such closure operators.

The proposed model is intended as a framework rather than a completed theory. Its main purpose is to shift the mathematical question from "What is meaning?" to "How does a context-dependent structure of signs change during understanding, learning, and communication?"

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR28713025).

Keywords: meaning, signs, semantic structure, context-dependent graph, understanding, knowledge representation, large language models, structural closure.

2020 Mathematics Subject Classification: 68T30, 68T27

The fractal Jonsson theory and online learning

A.R. YESHKEYEV[1], A.K. ISSAYEVA[2], N.V. POPOVA[3]

[1,2,3]*NLS Buketov Karaganda National Research University, Karaganda, Kazakhstan*

[1]aibat.kz@gmail.com, [2]isaevaaiga@gmail.com, [3]popovanv092024@gmail.com

In this work, we present a novel model-theoretic framework for formalizing fractality within first-order logic. This approach is built upon elementary endomorphisms on models paired with elementary syntactic endomorphisms on theories, anchored by the central concept of the fractal corridor. Under the assumption of coordinated syntactic and semantic fractality, we rigorously establish fractal soundness, fractal completeness, and a refined formulation of fractal compactness.

Let L be a first-order language and T a theory in L. Let Fm_L denote the set of all L-formulas.

Definition 1 (A.R. Yeshkeyev)**.** Let $M \models T$. An endomorphism $f : M \to M$ is called *elementary* if

$$\forall \varphi(\bar{x}) \in \mathrm{Fm}_L \; \forall \bar{a} \in M \Big(M \models \varphi(\bar{a}) \Rightarrow M \models \varphi(f(\bar{a})) \Big).$$

Definition 2 (A.R. Yeshkeyev)**.** Let L be a first-order language. A mapping $\sigma : L \to L$ is called a syntactic endomorphism if it induces a map on terms and on quantifier-free formulas such that:

1. For every function symbol f,
$$\sigma(f(t_1, \ldots, t_n)) = f(\sigma(t_1), \ldots, \sigma(t_n)).$$
2. For every predicate symbol P,
$$\sigma(P(t_1, \ldots, t_n)) = P(\sigma(t_1), \ldots, \sigma(t_n)).$$
3. Boolean structure is preserved:
$$\sigma(\neg\varphi) = \neg\sigma(\varphi), \qquad \sigma(\varphi \wedge \psi) = \sigma(\varphi) \wedge \sigma(\psi),$$
and similarly for other connectives.

Definition 3 (A.R. Yeshkeyev)**.** A syntactic endomorphism σ is called elementary if it extends to all formulas of L by the rules

$$\sigma(\forall x\,\varphi) = \forall x\,\sigma(\varphi), \qquad \sigma(\exists x\,\varphi) = \exists x\,\sigma(\varphi),$$

and therefore acts on the whole set of sentences of L.

Definition 4 (A.R. Yeshkeyev)**.** Let T be a theory in L. An elementary syntactic endomorphism σ is called T-elementary if for every sentence φ

$$T \vdash \varphi \quad \Longleftrightarrow \quad T \vdash \sigma(\varphi).$$

Fix finite systems $\Psi = \{\sigma_1, \ldots, \sigma_m\}$, $\Phi = \{f_1, \ldots, f_n\}$.

Definition 5 (A.R. Yeshkeyev)**.** Let T be a theory in a first-order language L, and let

$$\Psi = \{\sigma_1, \ldots, \sigma_m\}$$

be a finite family of T-elementary syntactic endomorphisms.

Define recursively a sequence of theories:

$$T[0] = T,$$
$$T[k+1] = \mathrm{Th}\Big(\bigcup_{j=1}^{m} \sigma_j(T[k])\Big), \quad k \geq 0.$$

The theory T is called syntactically fractal with respect to Ψ if there exists $n \geq 1$ such that

$$T[n] = T.$$

Definition 6 (A.R. Yeshkeyev)**.** A model $M \models T$ is fractal relative to Φ if

$$M = \bigcup_{i=1}^{n} f_i[M],$$

where each f_i is an elementary endomorphism of M, and

$$n = 1 \Rightarrow f_1 \text{ is not an automorphism.}$$

Define inductively:

$$T[0] = T,$$
$$T[k+1] = \bigcup_{j=1}^{m} \sigma_j(T[k]).$$

Definition 7 (A.R. Yeshkeyev)**.** The sequence $(T[k])_{k\geq 0}$ is the fractal corridor generated by Ψ. A formula φ is fractally provable if $\exists k\ (\varphi \in T[k])$.

Let $\mathbf{M}(T) = \{M \models T\}$. Define the quadruple $\mathcal{F} = (T, \mathbf{M}(T), \Psi, \Phi)$.

We say that syntactic and semantic fractality are coordinated if there exists a mapping

$$\theta : \{1, \ldots, m\} \to \{1, \ldots, n\}$$

such that for every $M \models T$, every formula φ, and every tuple $\bar{a} \in M$,

$$M \models \varphi(\bar{a}) \Rightarrow M \models \sigma_j(\varphi)\big(f_{\theta(j)}(\bar{a})\big).$$

Let

$$\mathcal{F}_{\Psi,\Phi}(T) = \{M \models T \mid M \text{ fractal relative to } \Phi\}.$$

Definition 8 (2)**.** Let X be a set and $\mathcal{F} \subseteq \mathcal{P}(X)$. The pair $(X, \mathcal{F})$ is called a set system. We say that $A \subseteq X$ is shattered by $\mathcal{F}$ if for every $S \subseteq A$ there is $F \in \mathcal{F}$ such that $F \cap A = S$. A family $\mathcal{F}$ is said to be a VC-class on X if there is some $n < \omega$ such that no subset of X of size n is shattered by $\mathcal{F}$. In this case the VC-dimension of $\mathcal{F}$, denoted by $\mathrm{VC}(\mathcal{F})$, is the smallest integer n such that no subset of X of size $n+1$ is shattered by $\mathcal{F}$. If no such n exists, we write $\mathrm{VC}(\mathcal{F}) = \infty$.

Theorem 9 (2)**.** *The following are equivalent for a hypothesis class H with domain X: (1) H has finite VC dimension. (2) H is PAC learnable (PAC stands for Probably Approximately Correct).*

Definition 10 (A.R. Yeshkeyev)**.** A Jonsson theory T is a J-o-minimal theory if it is T^*-o-minimal.

Definition 11. Let T be a fractal J-o-minimal theory and $\phi(x, y)$ be some $\exists$-formula (with a partition of the free variables). Say that $\phi(x, y)$ has the J-independence property (or JIP) if there are in M, $M \in \mathcal{F}$ (where $\mathcal{F} \subseteq K(T)$ is the theory T), sequences $(a_i)_{i<\omega}$ and $(b_s)_{s\subseteq\omega}$ such that $\phi(a_i, b_s)$ holds iff $i \in s$. Say that ϕ is $JNIP$ if not. The theory T is $JNIP$ if all formulas are $JNIP$.

Lemma 12. *If theory T is a fractal J-o-minimal theory, then T has a $JNIP$-formula.*

Theorem 13. *Every fractal J-o-minimal Jonsson theory is a $JNIP$-theory.*

Funding: This research is funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23489523).

Keywords: fractal subsets, fractal model, syntactic fractality, fractal corridor, fractal completeness, fractal compactness.

2020 Mathematics Subject Classification: 35Q79, 35K05, 35K20

References

[1] YeshkeyevA.R. *Theories and their models*,Karaganda, Kazakhstan: izd. KarU [in Russian], Volume 1,2. (2024)

[1]Vapnik, V. N., & Chervonenkis, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16(2), 264–280. (Reprinted from *Teor. Veroyatnost. i Primenen.*).

On Besov's Problem

M.A. ZAIGHUM

Heriot Watt International Faculty, K. Zhubanov Aktobe Regional University, Aktobe, Kazakhstan
Joint Work with
Erlan Nursultanov, Durvudkhan Suragan
m.asadzaighum@hw.ac.uk

This talk aims to offer a thorough introduction to grand Besov spaces, a natural and robust extension of the classical Besov scale that integrates the grand Lebesgue spaces framework. These generalized spaces excel at capturing subtle borderline phenomena in harmonic analysis, partial differential equations, and approximation theory.

The presentation focuses on the zero-smoothness case ($s = 0$). In this critical setting, we deliver a positive solution to Besov's problem, addressing embeddings and characterizations of zero-smoothness Besov spaces in connection with Lorentz spaces. Our method also enhances several classical embedding results.

2020 Mathematics Subject Classification: 42B15; 42B35

Balance constrained sparse group lasso for network time series

N. ZAKIYEVA[1], A. OTEMISSOV[2]

[1]*Institute of Mathematics and Mathematical Modeling, Almaty, Kazakhstan*
[2]*Nazarbayev University, Astana, Kazakhstan*

[1]nazgul.zakiyeva@gmail.com, [2]aotemissov@nu.edu.kz

We propose an Augmented Lagrangian Coordinate Descent (ALCD) based algorithm for solving the constrained sparse group lasso problem in balanced network time series models. Network autoregression with balance constraint of lag 1 is writt

$$\begin{aligned} \mathbf{Y} &= \mathbf{XB} + \mathbf{E}, \\ s.t. \quad & \mathbf{XB1}_N = \boldsymbol{G}, \end{aligned} \tag{1}$$

where

- $\mathbf{Y}$ is a T $\times$ N matrix of the observations;
- $\mathbf{X}$ is a T $\times$ N design matrix;
- The parameter matrix $\mathbf{B}$ is a N $\times$ N matrix with unknown parameters $\mathrm{b}_{j,i}$;
- $\boldsymbol{G}$ is a known vector in the constraint;
- Finally, $\mathbf{1}_N$ is a unit vector and $\boldsymbol{E}$ is a $T \times N$ matrix of white noise errors.

Sparse-Group Lasso penalty is imposed in $\boldsymbol{B}$, see Simon et al.(2013).

$$\min_{\boldsymbol{B}} \left\{ \frac{1}{2T} \|\boldsymbol{Y} - \boldsymbol{XB}\|_F^2 + \lambda_1 \sum_{\substack{j=1,\dots,N \\ i=1,\dots,N \\ i \neq j}} |b_{j,i}| + \lambda_2 \sum_{j=1}^{N} \|\boldsymbol{B}_j\|_2 \right\} \quad \text{s.t.} \quad \boldsymbol{XB1}_N = \boldsymbol{G}, \tag{2}$$

where $||\boldsymbol{A}||_F = \sqrt{\sum_{j,i} a_{j,i}^2}$ is the Frobenius norm, $\boldsymbol{B}_j$ is the $j-$th row of the off-diagonal of the weighted adjacency matrix corresponding to the network effect of node j. Let $\lambda_1 \geq 0$ be the Lasso tuning parameter for the individual parameter $b_{j,i}$, where $j \neq i$. Let $\lambda_2 = \sqrt{N}\lambda_1 \geq 0$ be the tuning parameter for group $\boldsymbol{B}_j$. The constrained sparse group lasso extends the standard sparse group lasso by incorporating linear equality constraints, allowing prior structural information, such as balance conditions to be imposed directly on the coefficient space. The proposed algorithm combines coordinate descent with an augmented Lagrangian framework, yielding a flexible optimization procedure that simultaneously accommodates sparsity, group structure, and linear restrictions. We test our method on i) simulated data where vary the sample size, dimensionality, sparsity pattern

and tuning parameters, and on ii) real world gas network time series dataset. We compare our method with the classical convex solvers in CVX package and different formulations of models with lasso that does not consider the balance constraint in estimation. For the simulated data, the proposed method achieves higher numerical precision and improved support recovery while enforcing constraints to numerical accuracy. It highlights the computational advantage of ALCD for smaller models over CVX. For gas network dataset, the ALCD delivers improved predictive performance and balance relative to unconstrained lasso based models.

Funding: The work for this article has been supported by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23483953).

Keywords: sparse group lasso, augmented Lagrangian, coordinate descent, multivariate time series models.

Combinatorial State-Space and Empirical Game Tree Complexity of Togyzkumalak: A Billion-Game Analysis

A. ZEINULLA

Nazarbayev University, Astana, Kazakhstan

ansar.zeinulla@nu.edu.kz

Togyzkumalak is a two-player board game belonging to the mancala family [4]. In contrast to standard board games, Togyzkumalak features 162 non-differentiable pieces (kumalaks), 9 regular playing pits and a score-keeping *Kazan* for each of the players and a highly asymmetric rule of *Tuzdyk* (sanctuary creation). Despite its deep strategic complexity, the computational boundaries of the game remain unknown. In this work, a rigorous mathematical evaluation of the game's State-Space Complexity and a massive-scale empirical analysis of its Game Tree Complexity (GTC) are presented.

Proposition 1. *Since standard captures strictly require an opponent's pit to reach an even number of kumalaks, Kazan sums in the absence of a Tuzdyk are mathematically constrained to even parity. The exact theoretical active state-space complexity of Togyzkumalak, bounded strictly by terminal score limits ($K < 82$), even-parity capture mechanics, Tuzdyk reachability constraints ($K \geq 3$), and rotational equivalence, is precisely* $15,097,418,467,704,714,090,211,147 \approx 1.51 \times 10^{25}$.

First, to mathematically formalize this space, let $E = \{x \in 2\mathbb{Z} \mid 0 \leq x \leq 80\}$ represent the set of valid even-parity Kazan states, $T = \{x \in \mathbb{Z} \mid 3 \leq x \leq 81\}$ represent the valid Tuzdyk Kazan states governed by the 3-piece reachability constraint, and let $R(k_1, k_2) = 162 - k_1 - k_2$ denote the remaining pieces on the board as a function of the iterators. By synthesizing combinatorial distributions over active bins and rotational symmetry, the unified state-space is formulated as:

$$
\begin{aligned}
Z_{unique} = \frac{1}{2}\Bigg[& \underbrace{\sum_{k_1 \in E} \sum_{k_2 \in E} \binom{R(k_1,k_2)+17}{17}}_{\text{0-Tuzdyks}} \\
& + \underbrace{8\left(\sum_{k_1 \in T} \sum_{k_2 \in E} \binom{R(k_1,k_2)+16}{16} + \sum_{k_1 \in E} \sum_{k_2 \in T} \binom{R(k_1,k_2)+16}{16} \right)}_{\text{1-Tuzdyk}} \\
& + \underbrace{56 \sum_{k_1 \in T} \sum_{k_2 \in T} \binom{R(k_1,k_2)+15}{15}}_{\text{2-Tuzdyks}} + \underbrace{\sum_{k \in E} \binom{89-k}{8}}_{\text{Self-symmetrical states}} \Bigg]
\end{aligned}
\tag{1}
$$

To compute the Empirical Game Tree Complexity (GTC), mathematical estimations fall short due to the vastly dynamic branching factor and the absence of historical game databases comparable to chess. To resolve this, a highly optimized, high-throughput C++ state generator was engineered to execute **1,000,000,000 (one billion)** uniform random playouts. By playing purely random legal moves without heuristic bias, we extracted the absolute statistical baseline of the game from exactly 124,474,599,634 observed board transitions.

Claim 2. *The dynamic branching factor b of Togyzkumalak decays significantly as the game progresses: from* $b = 8.36$ *in the opening phase (half-moves 1–20) to* $b = 6.09$ *in the endgame (half-moves 61+). Across all 124.4 billion positions, the global average branching factor stabilizes strictly at* $b_{avg} \approx 7.00$*. The empirical game length (d) strongly converges to an average of 124.47 half-moves, with a recorded maximum of 874 half-moves. While expert human games may exhibit different optimal path lengths, this uniform random baseline establishes the definitive mathematical average of the unpruned game tree.*

Using Shannon's classical formulation [2], we establish the exact Empirical Game Tree Complexity of Togyzkumalak:

$$GTC = b_{avg}^{d} = 7.00^{124.47} \approx 10^{105.19} \tag{2}$$

The dataset of one billion games reveals several key heuristics of the game's mathematical structure, independent of human strategy:

- **First-Mover Advantage:** Uniform random playouts resulted in a 53.14% win rate for the First Player, compared to 44.68% for the Second Player and a 2.18% draw rate.
- **The Tuzdyk Dominance:** The data mathematically quantifies the sanctuary (Tuzdyk) as the ultimate heuristic feature. When only the First player holds a Tuzdyk, their win ratio reaches 81.3%, whereas when only the Second player holds one, their win ratio is 72.8%. This 8.5% discrepancy demonstrates how the First-Mover Advantage synergizes with sanctuary creation, forcing the Second player to play from behind even when acquiring a Tuzdyk.
- **Positional Distribution:** Tuzdyk placement frequencies strictly decrease from the 1st to the 8th pit. However, the win ratio conversely increases from the 1st to the 8th pit for both players, proving mathematically that deeper sanctuaries are statistically harder to establish but yield a higher strategic payoff.

Conclusion 3. *Both the state-space and the empirical game tree of Togyzkumalak are significantly larger than those of solved combinatorial games like Checkers (state-space* 10^{18}*, GTC* 10^{31}*) [1] and Awari (state-space* 10^{12}*, GTC* 10^{32}*) [1]. The high computational complexity—proven as* 10^{25} *for the active state-space and* 10^{105} *for the game tree—demonstrates that traditional Minimax algorithms relying on brute-force Alpha-Beta pruning will be fundamentally insufficient for solving the game. However, the precise metrics and powerful material heuristics (such as the asymmetric Tuzdyk win probability) established in this billion-game analysis provide the critical mathematical foundation required to design state-evaluation functions for modern Zero-knowledge Deep Reinforcement Learning architectures [3].*

Keywords: Togyzkumalak, State-space complexity, Game tree complexity, Monte Carlo simulations, Artificial Intelligence, Combinatorial game theory.

References

[1] Allis, L. V. (1994). *Searching for solutions in games and artificial intelligence.* [Doctoral Thesis, Maastricht University]. Rijksuniversiteit Limburg. https://doi.org/10.26481/dis.19940923la

[4] World Togyzqumalaq Federation. (2026). *Laws of Togyzqumalaq: Official reference text and chapter guide.* Official website. `https://9qumalaq.world/#laws`

Capture ceilings for budget-constrained early warning in education

I.B. ZEYNALOV

Baku State University, Baku, Azerbaijan
ismayil.zeynalov.bakhish@bsu.edu.az

Predictive models that identify learners at risk of failing are judged almost universally by the area under their ROC curve [1]. What an institution obtains, however, is not an area but a number of learners reached: support capacity is finite, so only a fraction t of the cohort can be contacted. We ask what that fraction, with the quality of the ranking, implies for the share of at-risk learners actually reached.

Let p be the prevalence of at-risk learners and let the fraction t with the worst predicted scores be contacted. Write R for the *capture*, the share of at-risk learners inside the contacted group.

Proposition 1. *For every ranking rule,* $R \le \min\{1,\, t/p\}$.

Indeed a cohort of size n contains pn at-risk learners while only tn are contacted, so at most $\min\{tn, pn\}$ are reached. The bound uses no property of the model: with a 10% budget and 20% prevalence, no predictor, however accurate, reaches more than half of them.

To see how far a model falls below that ceiling, adopt the binormal model [2]: scores of at-risk learners are $N(\mu, 1)$, those of the others $N(0, 1)$, so the area under the curve is $A = \Phi(\mu/\sqrt{2})$ and $\mu = \sqrt{2}\,\Phi^{-1}(A)$. The threshold c contacting exactly the fraction t solves $p\,[1 - \Phi(c - \mu)] + (1 - p)[1 - \Phi(c)] = t$, whence

$$R(A, p, t) = 1 - \Phi(c - \mu), \qquad E(A, p, t) = \frac{R(A, p, t)}{\min\{1, t/p\}} \in [0, 1], \tag{1}$$

E being the efficiency with which the budget is used. Capacity limits have been treated as a loss of realized net benefit [4]; (1) instead gives capture in closed form and inverts into the capacity a planner must buy.

The trade-off between a better model and more capacity is then explicit. At $p = 0.20$ and $t = 0.05$, improving the model from $A = 0.75$ to $A = 0.80$ raises capture from 0.154 to 0.180, whereas five extra budget points raise it to 0.268: capacity is worth 4.4 times the model gain. The ratio falls to 2.1 at $t = 0.10$ and 1.1 at $t = 0.20$, so model quality dominates only once the budget approaches the prevalence.

On the Open University Learning Analytics dataset [3] we rank the 2498 learners of its largest presentation by their click volume in the first four weeks ($p = 0.594$, $A = 0.752$). Predicted and observed capture agree with a mean absolute error of 0.017 over budgets $t \in \{0.05, 0.10, 0.20, 0.30\}$; reaching half of the at-risk learners requires contacting 36.5% of the cohort. Capture at the deployable budget, not the area under the curve, is what an early-warning system should be asked to report.

Keywords: early-warning systems, ROC analysis, binormal model, capacity planning, learning analytics, educational data mining.

2020 Mathematics Subject Classification: 62P25, 62H30, 68T05

References

[1] Hanley J.A., McNeil B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, **143** (1982), 29–36.

[2] Green D.M., Swets J.A. *Signal Detection Theory and Psychophysics*, Wiley, New York (1966).

[3] Kuzilek J., Hlosta M., Zdrahal Z. Open University Learning Analytics dataset, *Scientific Data*, **4** (2017), 170171.

[4] Singh K., Shah N.H., Vickers A.J. Assessing the net benefit of machine learning models in the presence of resource constraints, *J. Am. Med. Inform. Assoc.*, **30**:4 (2023), 668–673.

Decomposing prompt-format sensitivity: extraction artifacts, reasoning suppression, and surface form

I.B. ZEYNALOV

Baku State University, Baku, Azerbaijan

ismayil.zeynalov.bakhish@bsu.edu.az

Language models are reported to be strikingly sensitive to meaning-preserving changes of prompt format [1,2]; a recent counter-position holds that much of this is produced by the evaluation machinery rather than by the models [3]. We separate the claims experimentally and find that what is reported as one phenomenon is three, of very unequal size.

Design. We generate 150 multiple-choice items programmatically: contamination-free by construction, reproducible from a seed, with exact keys and named-error distractors. Each item is posed under 6 semantically equivalent formats (markers `A)`, `A.`, `(A)`, `[A]`, `Option A:`; options inline or one per line; answer instruction before or after the question), with identical content and option order and greedy decoding. Four systems are evaluated, an open-weight 7B model run locally [4] and three proprietary tiers: 5,400 scored responses.

One: the harness. Under the common shortcut of capping generation at a few tokens and reading the first option letter that appears, the open model looks catastrophically format-sensitive: a spread of 25.3 points, 48% of items format-dependent. It is an illusion. In 600 of 900 responses (66.7%) no option letter was emitted at all, the model having begun to reason and been truncated; the parser assigned each a definite option regardless, collapsing all 600 onto the letter that first occurs inside words such as "To **d**etermine". Exactly 21.3% of these phantom answers scored correct, the chance rate.

Two: the instruction. With the answer read from a fixed final-answer syntax and a budget under which nothing is truncated, a 52.7-point spread survives in the open model, and it does not follow the markers. The two formats whose instruction demands a bare letter yield 2 to 3 words of output and 47.3% and 49.3% accuracy; the four whose answer instruction follows the question yield about 117 words of working and 98.7% to 100%. Within the suppressing format the few responses that nevertheless showed working were all correct (3/3) against 46.3%: the format decides whether the model computes, not what it knows.

Three: surface form. Holding harness and instruction fixed, marker style and layout account for 1.3 points of spread and 1.1% mean churn. On the proprietary tiers even that vanishes: 99.3% to 100% under every format, mean churn 0.2%, 0.0% and 0.0% over 900 responses each, the two strongest choosing the identical option for every item under every format; a deliberately harder generated set was answered 30 out of 30 by the strongest tier.

Implication. Prompt-format sensitivity should be reported only after harness and instruction semantics are separated out: benchmarks should state the generation budget, the extraction rule, the parse rate, the share of responses with no extractable answer, and whether a format suppresses intermediate computation. Surface form proper matters little, and at frontier capability not at all.

Keywords: large language models, benchmark evaluation, prompt sensitivity, answer extraction, chain-of-thought, evaluation protocol.

2020 Mathematics Subject Classification: 68T50, 68T05, 62P99

References

[1] Sclar M., Choi Y., Tsvetkov Y., Suhr A. Quantifying language models' sensitivity to spurious features in prompt design, *ICLR* (2024), arXiv:2310.11324.

[2] Mizrahi M., Kaplan G., Malkin D., Dror R., Shahaf D., Stanovsky G. State of what art? A call for multi-prompt LLM evaluation, *Trans. Assoc. Comput. Linguistics*, **12** (2024), 933–949.

[3] Hua A., Tang K., Gu C., Gu J., Wong E., Qin Y. Flaw or artifact? Rethinking prompt sensitivity in evaluating LLMs, arXiv:2509.01790 (2025).

[4] Bai S., Chen K., Liu X. et al. Qwen2.5-VL technical report, arXiv:2502.13923 (2025).

From score reliability to decision reproducibility in AI-based educational early warning

I.B. ZEYNALOV

Baku State University, Baku, Azerbaijan

ismayil.zeynalov.bakhish@bsu.edu.az

An artificial-intelligence early-warning system in education does not deliver a score; it delivers a *decision*, the list of learners to be contacted. Whether that list survives an equally valid re-measurement is a property of the decision rule rather than of the score, and it matters most where data per learner are scarce and every intervention has a cost. Prediction instability has been studied as an artefact of model development [4]; here it arises from measurement error in the score itself.

Let a learner's risk score be $X = T + E$ with independent true and error parts and reliability $\rho = \operatorname{Var} T / \operatorname{Var} X$, and let two parallel occasions yield (X_1, X_2), standard bivariate normal with correlation ρ under a Gaussian score distribution. A system flags the fraction q with the lowest scores, that is, $X \le c$ with $c = \Phi^{-1}(q)$. The chance-corrected reproducibility of the flag is

$$\kappa(\rho, q) = \frac{2\,\Phi_2(c, c; \rho) + 1 - 2\Phi(c) - p_e}{1 - p_e}, \qquad p_e = q^2 + (1-q)^2, \tag{1}$$

with $\Phi_2(\cdot, \cdot; \rho)$ the standard bivariate normal distribution function.

Proposition 1. *For the median rule $q = 1/2$, Sheppard's orthant identity $\Phi_2(0,0;\rho) = \frac{1}{4} + \frac{1}{2\pi}\arcsin\rho$ reduces (1) to $\kappa = \frac{2}{\pi}\arcsin\rho$.*

Proposition 2. *For fixed $\rho \in (0,1)$, both $\kappa(\rho,q)$ and the re-flag probability $\Phi_2(c,c;\rho)/q$ tend to 0 as $q \to 0$, and both decrease as the rule becomes more selective on $(0,1/2]$.*

Proposition 2 follows from the asymptotic independence of Gaussian tails [2]. If the score aggregates k exchangeable units (weeks of activity, quiz items) of single-unit reliability ρ_1, the Spearman–Brown relation $\rho_k = k\rho_1/(1+(k-1)\rho_1)$ with Proposition 1 inverts into a data budget: the median rule attains a target κ^* as soon as

$$k \;\geq\; \frac{\rho^*(1-\rho_1)}{\rho_1(1-\rho^*)}, \qquad \rho^* = \sin\Big(\frac{\pi\kappa^*}{2}\Big). \tag{2}$$

A score meeting the conventional target $\rho = 0.80$ therefore supports only $\kappa = 0.59$ under a median rule and $\kappa = 0.47$ when the most at-risk 5% are flagged, where merely half of the flagged learners would be flagged again; $\kappa^* = 0.6$ needs $\rho \geq 0.81$ at $q = 1/2$ but $\rho \geq 0.89$ at $q = 0.05$. The tighter the intervention budget, the more data each decision requires.

We validate (1) on the Open University Learning Analytics dataset [3]: weekly activity of the 2212 learners of its largest presentation, single-week reliability $\rho_1 = 0.43$. Split-half decisions from odd and from even weeks reproduce the prediction of (1) at the observed half-score correlation with a mean absolute error of 0.019 in κ across $q \in \{0.5, 0.3, 0.2, 0.1\}$; at $q = 0.2$ predicted and observed are 0.598 and 0.599. Rule (2) then prescribes 6, 11 and 26 weeks of data for $\kappa^* = 0.6, 0.7, 0.8$: a budget checkable before deployment.

Keywords: decision consistency, reliability, Gaussian orthant probability, Spearman–Brown formula, early-warning systems, learning analytics.

2020 Mathematics Subject Classification: 62H20, 62P25, 68T05

References

[1] Sheppard W.F. On the application of the theory of error to cases of normal distribution and normal correlation, *Philos. Trans. R. Soc. A*, **192** (1899), 101–167.

[2] Sibuya M. Bivariate extreme statistics, I, *Ann. Inst. Statist. Math.*, **11** (1960), 195–210.

[3] Kuzilek J., Hlosta M., Zdrahal Z. Open University Learning Analytics dataset, *Scientific Data*, **4** (2017), 170171.

[4] Riley R.D., Collins G.S. Stability of clinical prediction models developed using statistical or machine learning methods, *Biometrical Journal*, **65**:8 (2023), 2200302.

Toward Verified Intelligence: A Survey of AI for Mathematics

N. ZIBATKALIYEV

Desargues, Inc., Astana, Kazakhstan

znurmukhamed57@gmail.com

A survey of the current wave of AI-driven mathematics: gold-medal olympiad performance, decades-old open problems falling, and formal verification increasingly in the loop. The talk traces the arc from informal to formal proof, introducing Lean and Mathlib as the infrastructure behind this shift, and walks through the timeline of AI-and-formal-math breakthroughs to date. It then turns to where the field is actually stuck — the scarcity of high-quality formalized data, the unresolved gap between informal mathematical intuition and formal verifiability, autoformalization's faithfulness problem, and the difficulty of building benchmarks that measure real reasoning rather than memorized patterns — before closing with a view of where verified intelligence is heading next, from mathematics into code and science more broadly.

Keywords: formal verification, Lean, autoformalization, AI for mathematics, mathematical reasoning benchmarks.

Index of Names

Abek A.N., 7
Abylayeva A., 44
Akhmanova D.M., 35
Aliyev M., 9
Askarov A., 13
Assanova A.T., 13
Assylbekov Y.M., 10

Babayeva G., 10
Baktybergen K., 11
Bassarov S.Zh., 12
Beisenbayeva K., 28
Bokayev N.A., 7, 21
Botaibekov A., 13
Bukanay N.U., 13

Dildabek G., 52

Esfahani A., 15

Fagnoni T., 34
Fazyl G.G., 56
Fazyl Z.G., 16

Garibli N., 18, 19
Gogatishvili A, 7, 21, 23

Huseynli F.S., 25
Huseynov E., 27

Igenbayev A., 41
Issayeva A.K., 60

Jarbulova K., 44
Jubatkanov K., 28
Jumadildayev M.A., 29

Kairolla M., 30
Kalidolday A.K., 31
Karshygina G, 23
Kenzhebekova G.S., 43
Kettebekov Z.K., 55
Khalmetov E., 49
Kim A., 33
Kingston P., 34
Koshzhanov R.A., 55
Kosmakova M.T., 35
Kovalenko A.A., 55
Kuanyshov N., 37
Kuanysh S.K., 36
Kuzeubayeva N.K., 21

Manarbek M., 39
Markhabatov N., 28
Markhabatov N.D., 16, 56
Mirzakulova A.E., 13
Molybaikyzy A., 39
Mosavi A., 41, 42
Mukhambetzhan M.A., 43
Mukhamedina A.I., 57
Muratbekov M., 44
Murzabekova G.Y., 45

Nursultanov E.D., 31

Omarov M., 49
Orazbayev A., 50
Orumbayeva N.T., 50
Ospanov K.N., 52
Otemissov A., 63

Popova N.V., 60

Ramazanov M., 49

Sabdenaliyev B., 42
Sadybekov M., 52
Satybaldina A.N., 54
Sobolev B., 11

Tleukhanova N.T., 12, 39
Tulenbaev K.M., 55
Turgumbayev M.Zh., 43
Tursungaliyev N.N., 56
Tussupov J., 28

Ussipbayev A.A., 57

Verbovskiy V.V., 59

Yeginbay I., 37
Yeshkeyev A.R., 60
Yeskabylova Zh.B., 52

Zaighum M.A, 50
Zaighum M.A., 62
Zakiyeva N., 63
Zeinulla A., 64
Zeynalov I.B., 66–68
Zhantassova B.B., 50
Zhardemova N., 15
Zharykov M.N., 55
Zibatkaliyev N., 70

Ünver T, 7, 21, 23

For notes

1ST INTERNATIONAL CONFERENCE AND WORKSHOP ON MATHEMATICS AND ARTIFICIAL INTELLIGENCE

This Book of Abstracts contains the abstracts of the scientific contributions presented at the 1st International Conference and Workshop on Mathematics and Artificial Intelligence, held on September 9–11, 2026, at Nazarbayev University, Astana, Kazakhstan.

The conference brings together researchers, practitioners and young scientists working at the intersection of modern mathematics and artificial intelligence. The contributions cover function spaces and harmonic analysis, operator theory, differential equations and inverse problems, mathematical logic, combinatorics and topology, as well as machine learning, large language models, intelligent decision support and AI for mathematics.

ORGANIZERS AND PARTNERS

ISSAI Nazarbayev University | Institute of Smart Systems and Artificial Intelligence

Nazarbayev University

School of Computing and Artificial Intelligence

NU Impact Foundation

ISBN 978-601-08-6470-2

Nazarbayev University, Astana

September 9–11, 2026