DOKTORANDSKÉ DNY 2025

sborník workshopu doktorandů FJFI
oboru Matematické inženýrství

21. a 28. listopadu 2025

P. Ambrož, Z. Masáková (editoři)

Doktorandské dny 2025
sborník workshopu doktorandů FJFI oboru Matematické inženýrství

P. Ambrož, Z. Masáková (editoři)
Kontakt petr.ambroz@fjfi.cvut.cz / 770 127 206

Vydalo České vysoké učení technické v Praze
Zpracovala Fakulta jaderná a fyzikálně inženýrská

Počet stran 98, Vydání 1.

Seznam příspěvků

Předmluva

Doktorandské dny se na Fakultě jaderné a fyzikálně inženýrské ČVUT v Praze staly zavedenou tradicí, která pravidelně spojuje studenty doktorských programů Matematické inženýrství a nově také Aplikovaná informatika a Kvantové technologie. Tyto programy jsou realizovány ve spolupráci kateder matematiky, fyziky a softwarového inženýrství. Účastníci během workshopu představují výsledky své vědecké práce, které pokrývají široké spektrum témat v oblasti aplikované matematiky.

V roce 2025 se koná již dvacátý ročník této akce, a to ve dnech 21. a 28. listopadu v prostorách FJFI.

Sborník, který právě držíte v rukou, obsahuje jak plné verze studentských příspěvků, tak i abstrakty doplněné odkazy na články publikované (či zaslané k publikaci) v odborných časopisech nebo konferenčních sbornících.

Poděkování za finanční podporu patří katedře matematiky FJFI a také grantu Studentské vědecké konference SVK 32/25/F4.

Editoři

Mutual Information Estimation and Information Bottleneck in Neural Networks

Vladislav Belov
vladislav.belov@cvut.cz

study programme: Mathematical Engineering
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Radek Mařík, Department of Telecommunication Engineering
Faculty of Electrical Engineering, CTU in Prague

Abstract. Understanding how information propagates within neural networks is crucial to both theory and practice in Machine Learning. Mutual Information (MI) may serve as a tool for this analysis, but estimating it reliably in high-dimensional, noisy settings remains a challenge. First, we compare select MI estimators, with a focus on the stability of MINE and its conditional variant, CMINE. Our experiments demonstrate that while MINE yields consistent results under noise and increasing dimensionality, CMINE is more sensitive to spurious dependencies, with dimensionality reduction providing partial but lossy mitigation. Analysis of information flow across network layers further reveals trajectories consistent with the Information Bottleneck hypothesis, where deeper layers compress redundant input while aligning with targets.

Keywords: Information Bottleneck, Manifold Learning, Mutual Information

Abstrakt. Porozumění, jak se informace šíří v rámci neuronových sítí, je zásadní jak pro teorii, tak pro praxi strojového učení. Vzájemná informace (MI) může sloužit jako nástroj pro tuto analýzu, avšak její spolehlivý odhad ve vysoce dimenzionálních a šumem zatížených prostředích zůstává výzvou. Nejprve porovnáváme vybrané odhady MI se zaměřením na stabilitu metody MINE a její podmíněné varianty CMINE. Naše experimenty ukazují, že zatímco MINE poskytuje konzistentní výsledky i při růstu dimenzionality a přítomnosti šumu, CMINE je citlivější na falešné závislosti, přičemž redukce dimenzionality přináší částečné, ale přitom ztrátové zlepšení. Analýza toku informací napříč vrstvami sítě dále odhaluje trajektorie odpovídající hypotéze Information Bottleneck, kde hlubší vrstvy komprimují redundantní vstup a současně se více slaďují s cílovými proměnnými.

Klíčová slova: informační hrdlo, varietní učení, vzájemná informace

1 Introduction

In many complex systems, ranging from biological networks to social interactions and computational architectures, understanding how information is transmitted across successive stages is fundamental for both scientific and practical standpoints. Quantifying this transmission reveals hidden dependencies and uncovers potential bottlenecks that affect the effectiveness of communication of relevant signals. Within the context of Machine Learning, the question of how information evolves within a model is directly tied to issues of transparency and trustworthiness. Tracing information flow across layers provides one with a route for explaining and interpreting model behavior at both training

and inference times. This connection has fueled growing interest in information-theoretic approaches to explainable AI and Information-Bottleneck-based loss functions.

A central tool in this analysis is *Mutual Information* (MI), which provides a rigorous measure of statistical dependence between random variables. Applied to neural networks, MI allows us to examine how much information about the input or target is preserved, lost, or transformed as data propagates through the architecture. Such an analysis can clarify the internal decision-making process of models and also highlight points where signals are distorted or lost. The invariance of MI under smooth, invertible transformations, along with its interpretation in meaningful units, makes it an attractive metric for performing such tasks. At the same time, these benefits are weakened by numerous challenges: MI is rarely tractable in high-dimensional settings, and accurately estimating it from finite samples remains a long-standing open problem [14]. Addressing these challenges is essential for turning information-theoretic analysis into a reliable tool for modern Machine Learning.

In this paper, we make the following contributions: (i) a systematic comparison of state-of-the-art Mutual Information estimation techniques when applied to measuring information flow across neural network layers; (ii) evaluation of the stability and reliability of these estimators under varying sample sizes, dimensionalities, and noise conditions; (iii) highlight how estimator choice influences conclusions about model interpretability and robustness; (iv) empiric validation of findings on both synthetic benchmarks and real-world datasets. Our results clarify the trade-offs among competing methods and provide practical guidelines for researchers aiming to use information-theoretic tools in the analysis of Neural Network systems.

Our work is straightforwardly subdivided into sections. We discuss the related work in Section 2 to contextualize our research. In Section 3, we discuss the analyzed methods from the mathematical standpoint. Section 4 outlines the experimental setup we propose, followed by Section 5, where the results are discussed. The work is concluded by Section 6, where we review the results and propose the next steps.

2 Related Work

In this section, we discuss the literature available on information flow in neural network models. The authors of [7] estimate the information flow in feedforward neural models from the standpoint of Mutual Information between inputs and hidden layer outputs. Moreover, the paper verifies that the connection between MI reduction throughout training is driven by the progressive geometric clustering of representations of samples from the same class. In [7], the MI value is estimated using the classical approach of binning, which is limiting when one is concerned with high-dimensional representations, e.g., in Computer Vision or Natural Language Processing domains. A work that aims to explain the information propagation through deep feedforward neural networks is also presented in [20]. A lower bound for MI between model inputs and hidden layers is derived from layer weights and biases through the mean-field analysis. In addition, the authors demonstrate that parameter initializations known to be optimal for training are superior from the MI perspective. In [23], information flow is studied from the deep Restricted Boltzmann Machines (RBM) perspective. The authors link the inputs with regularization

utilizing Mutual Information, deriving the upper bound. For instance, they show that values corresponding to regularization, such as weight matrix norms and input variances, can be directly related to the MI of an RBM layer. The authors of both articles [20, 23] do not compute the MI values directly nor study the propagation of information concerning the output labels, which is one of the key differences in our work.

Estimation of Mutual Information

Authors of [4] comprehensively review some existing Mutual Information Estimators from different perspectives, such as data sparsity and dimensionality, long-tail nature of real-world distributions, or robustness to diffeomorphisms. In [17], another experimental study of neural MI estimators is carried out. Using random vectors that model power spectrum features across multiple dimensions, they demonstrate that the reviewed estimators [13, 8, 22, 1, 19] exhibit bias or variance in the estimated value at high expected MI values, leading to errors. They also propose a heuristic that combines the methods, highlighting the strengths of the underlying algorithm. Nonetheless, the authors state that the selection of parameters for the algorithm must be tailored based on the characteristics of the underlying data. Therefore, the generalization capabilities of the approach are not yet sufficiently explored.

One of the key challenges in estimating MI from real-world samples is the limited availability of data. In [9], it is shown that the MI estimated for N samples cannot be greater than $\mathcal{O}(\log N)$ while estimating a distribution-free high-confidence MI lower bound. Breaking through this limit is possible, e.g., by means of smoothing out the contrastive estimation (smoothed InfoNCE [24]) or estimating the MI employing a flow model (specifically, the Real NVP architecture [5, 17]). Usage of normalizing flows for estimating MI is further explored in [2], where the authors propose employing learnable diffeomorphisms to achieve tractable MI in the latent space.

Since we focus on analyzing the propagation of information within neural networks, plain Mutual Information can be considered too simplistic, as each subsequent layer of the neural network in question depends on its previous inputs. By computing Mutual Information flowing from one layer to another, one does not explicitly take into account previous layers. However, it is possible to be done through conditional Mutual Information (CMI) [3]. Authors of [11, 12] define an algorithm designated as CMINE (Conditional MINE), which is also a neural estimator, such as MINE [1]. The estimator is based on the idea of using a classifier differentiating between samples appearing in two distributions: appearing in the nominator and the denominator of a conditional MI definition. The outputs of such a classifier are served to an estimator, which is then used to compute values of CMI. Unfortunately, CMINE inherits the practical challenges of MINE, including potential instability during training, high variance of gradient estimates, and sensitivity to the choice of network architecture and optimization strategy.

In this work, we focused on MINE and CMINE.

Information Flow Measurement and Information Bottleneck

Applied to neural networks, the *Information Bottleneck* (IB) perspective suggests that deeper layers filter out unnecessary input information, retaining only task-relevant fea-

tures. This approach has been influential in explaining generalization, robustness, and the dynamics of training in deep models. However, the empirical validity and precise mechanisms by which neural networks approximate IB remain debatable [18, 16]. For instance, in [15], the authors analyze the information plane dynamics of intermediate neural-network representations based on the Hilbert-Schmidt independence criterion (HSIC), proposing HSIC as an alternative to Mutual Information. They compare layer-wise and end-to-end training approaches, where the difference is defined by how the objective function and respective gradients are computed and applied. Through HSIC, the end-to-end approach is argued to efficiently propagate input information, resulting in an Information Bottleneck representation at the final layer.

3 Modeling Methodology

Background

In this work, the Information Flow is measured through Mutual Information (MI) [3]:

$$\begin{aligned} I(X;Y) &= \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dxdy \\ &= D_{KL}\left(f(x,y)||f(x)f(y)\right), \end{aligned} \tag{1}$$

where $f(x,y)$ is the joint density of two random variables X and Y, and D_{KL} is the Kullback-Leibler divergence (KL-divergence) between two densities. $I(X;Y)$ can also be computed via the difference of entropies $H(X) - H(X|Y)$ or $H(Y) - H(Y|X)$, leading to an interpretation that MI is a measure of uncertanty reduction for a random variable when values of another random variable are known.

It is necessary to note that in the computational setup, one does not work with continuous random variables but with their quantized representations. For such representations, MI between X and Y is the limit of the MI between X^Δ and Y^Δ [3]:[1]

$$I(X^\Delta;Y^\Delta) \approx I(X;Y). \tag{2}$$

Conditional Mutual Information (CMI) captures the residual dependence between X and Y given Z, providing a more nuanced view of information flow in structured systems, such as neural networks. For conditional Mutual Information, the definition incorporates a third distribution, of a random variable Z whose values are given and are used to reduce the uncertainty of X due to knowledge of Y [3]:

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z). \tag{3}$$

Estimating Mutual Information

Among the selected techniques that measure MI is the Mutual Information Neural Estimation [1] technique to estimate MI between multidimensional variables. Its foundation is

[1] Here, Δ stands for the quantization size, e.g., for an n-bit quantization, it can be defined as $\Delta = 2^{-n}$.

based on the dual representations of the KL-divergence, primarily the Donsker-Varadhan (DV) representation [6]:

$$D_{KL}\left(P||Q\right) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_P\left[T\right] - \log \mathbb{E}_Q\left[e^T\right], \tag{4}$$

where the supremum is taken over such functions T for which the expectations are finite, and $\Omega \subset \mathbb{R}^d$ is some compact domain, defining distributions P and Q.

The Mutual Information Neural Estimator (MINE) is then defined as

$$\hat{I}_n(X;Y) = \sup_{\theta\in\Theta} \mathbb{E}_{P_{XY}^{(n)}}\left[T_\theta\right] - \log \mathbb{E}_{P_X^{(n)}\times P_Y^{(n)}}\left[e^{T_\theta}\right]. \tag{5}$$

In (5), an optimization framework over the family of functions $T_\theta : \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ is outlined. These functions are parametrized over Θ employing neural networks. The estimation is based on the empirical distribution $P^{(n)}$, associated with n i.i.d. samples, corresponding to distribution P.

From (4), it follows that the proposed framework creates a lower bound for the ground-truth Mutual Information between X and Y: $I(X;Y) \geq \hat{I}_n(X;Y)$.

While Mutual Information Neural Estimation (MINE) [1] has become a widely used tool for approximating MI in high-dimensional settings, many applications in Machine Learning require conditioning on auxiliary variables. In the case of the Neural Network, layers are always preconditioned by their predecessors, leading us naturally to the notion of CMI.

The CMINE framework extends MINE by introducing a parameterized critic function $T_\theta(x, y, z)$, which incorporates the conditioning variable z into the estimation. By leveraging the Donsker–Varadhan representation of the Kullback–Leibler divergence, CMINE approximates CMI through the following objective:

$$I_\theta(X;Y \mid Z) \approx \mathbb{E}_{P_{XYZ}^{(n)}}[T_\theta(x,y,z)] - \log\, \mathbb{E}_{P_{X|Z}^{(n)}\, P_{YZ}^{(n)}}\left[\exp\left(T_\theta(x,y,z)\right)\right]. \tag{6}$$

Here, the expectations are estimated via samples drawn from the joint distribution $p(x, y, z)$ and the product $p(x \mid z)\, p(y, z)$. As in MINE, the critic network is trained by maximizing this lower bound with respect to θ, yielding a data-driven approximation of conditional dependence.

Information Bottleneck

The Information Bottleneck principle [21] provides a theoretical framework for understanding representation learning in neural networks. It formalizes the trade-off between compression and prediction: an internal representation $T(X)$ of the input X should preserve more information about the target variable Y, while discarding irrelevant details of X. In practice, it means seeking representations that achieve a balance between $I(T;X)$ (how much the representation retains from the input) and $I(T;Y)$ (how much it informs about the output).

4 Experimental Setup

This section describes the experiments designed to evaluate the methodology introduced in Section 3. Our primary goal is to compare the performance and stability of different Mutual Information estimators, with a focus on MINE and CMINE, in the context of MI estimation through neural networks. For CMINE, we focus on the LDR estimate [11, 12], as the existing implementation of DV and NWJ [12] provided less stable and mathematically ill-conditioned results, converging towards negative MI values. Additionally, for stability evaluations, to allow for a fair comparison between MINE and CMINE, we set $Z = \text{const}$, which leads CMINE to compute the plain Mutual Information. To ensure both interpretability and controlled conditions, we consider the real-world dataset: the Iris benchmark.

We measure the Mutual Information between inputs, intermediate neural representations, and the class labels. Performance is evaluated along the following axes: (i) estimator stability under noise and dimensionality changes, where we embed the dataset into higher dimensions and add Gaussian noise with varying standard deviation; moreover, we attempt nonlinear mapping of embedded data onto a 2-dimensional space via UMAP [10] to reduce noise and observe how MI estimation performance is affected; (ii) consistency of information flow profiles across network layers from the Information Bottleneck perspective.

5 Experimental Results

Stability of MI Estimators

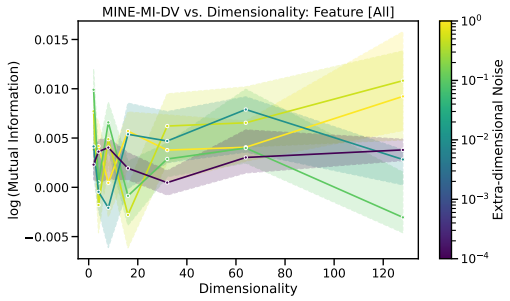
The first set of experiments focuses on the stability of MI estimators. Figure 1 compares the behavior of MINE (top row) and CMINE (bottom row) in estimating Mutual Information under varying dimensionalities and extra-dimensional noise, both with and without dimensionality reduction (UMAP). The experiment reveals differences between MINE and CMINE when estimating Mutual Information across varying dimensionalities and noise levels. MINE exhibited only minor fluctuations. Without dimensionality reduction, estimates remained stable even with higher values of dimensionality and the addition of noise. With UMAP, values tended to decrease the more noise and dimensions were added, suggesting a loss of information with dimensionality reduction. In contrast, CMINE showcased a tendency to see more information from the white noise added to extra dimensions without UMAP. With UMAP and increasing dimensionality and noise, MI estimates demonstrated a reversed order in terms of magnitude when compared to the experiment without dimensionality reduction. The results suggest that MINE provides a more reliable and interpretable estimator in noisy, high-dimensional settings, while CMINE struggles under such conditions. Nonetheless, applying dimensionality reduction helps reduce noise and stabilize predictions, even though some information may be lost during the process.

Information Flow in Neural Layers

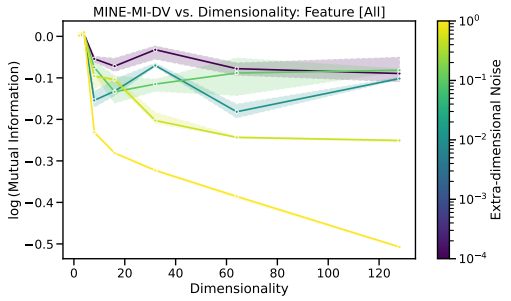
To investigate how information is transformed across network depth, we measured Mutual Information between inputs X, hidden representations T, and target labels Y for a feedforward neural network. For each hidden layer, we estimated $I(X;T)$ and $I(T;Y)$, using MINE, as it proved to be more stable in the general setup, and visualized them jointly in the information plane (Figure 2). The results reveal a characteristic trend that is consistent with the Information Bottleneck hypothesis, where early layers display relatively high $I(X;T)$ and lower $I(T;Y)$, indicating that the representations retain a large amount of raw input information but exhibit weaker predictive power. As the network deepens, the representations move along a trajectory toward lower $I(X;T)$ and higher $I(T;Y)$, indicating gradual compression of irrelevant features and increased alignment with the classification task.

These findings support the view that neural networks implement an implicit Information Bottleneck, where redundant details are discarded in favor of features relevant for generalization.

MINE

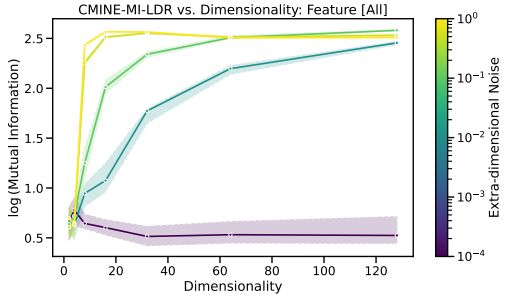


(a) Without UMAP

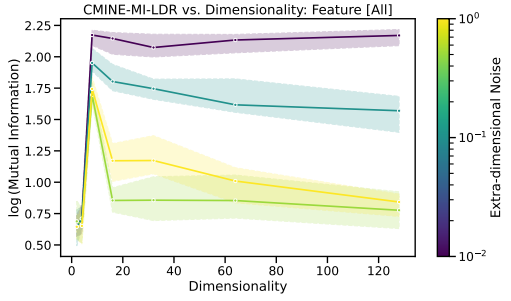


(b) With UMAP

CMINE



(c) Without UMAP



(d) With UMAP

Figure 1: Comparison of the Mutual Information estimators MINE and CMINE across increasing dimensionalities and varying noise levels within those dimensions. Each visual shows how estimation quality changes as additional noisy features are introduced. To further assess robustness, dimensionality reduction techniques are applied before estimation, allowing us to verify the stability of both estimators under high-dimensional and noisy conditions. The results highlight differences in sensitivity between MINE and CMINE.

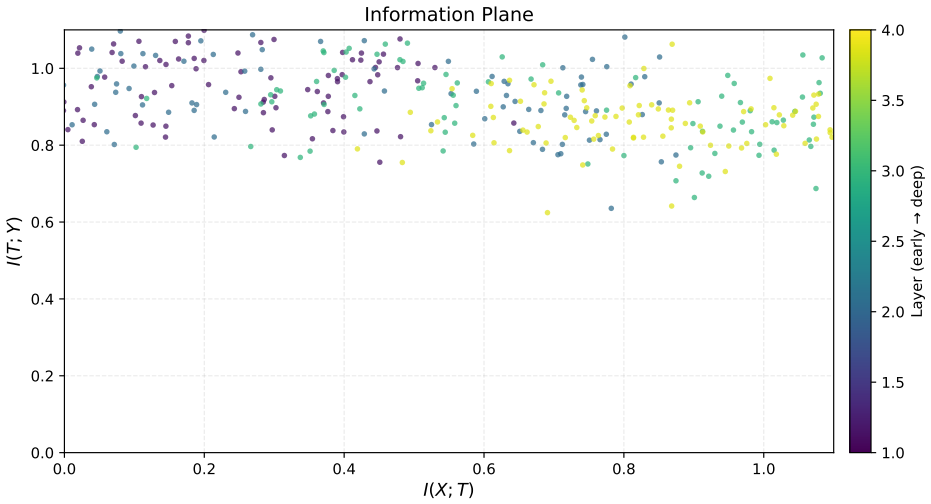


Figure 2: Information plane across network depth. Each colored cluster corresponds to one hidden layer, showing a progression from high $I(X;T)$ and low $I(T;Y)$ in early layers toward lower $I(X;T)$ and higher $I(T;Y)$ in deeper layers. The curved trend indicates information-bottleneck behavior, where representations compress input detail while concentrating task-relevant information.

6 Conclusion

In this work, the role of Mutual Information estimation in analyzing Information Flow through Neural Networks was examined, with an emphasis on the reliability and stability of different estimation techniques, as well as the Information Bottleneck observability through layers. Building on the methodology proposed in Section 3, the experiments (Sections 4-5) compared MINE and its conditional extension, CMINE. The results reveal notable stability differences: MINE maintains more consistent estimates under high-dimensional noise, while CMINE is more susceptible to spurious dependencies, especially in settings without dimensionality reduction. Although UMAP helps in reducing the variance of estimators, it comes at the cost of losing some information, highlighting the trade-off between stability and representation quality.

Our investigation of information flow in classification tasks further revealed how information transforms across layers, following trajectories in the information plane consistent with the Information Bottleneck hypothesis. Specifically, early layers preserve a larger portion of input variability, while deeper layers compress redundant details and align representations with predictive targets.

The findings suggest that while MINE provides a more reliable baseline in noisy, high-dimensional environments, conditional extensions such as CMINE require careful treatment to avoid instability. Future work should extend these analyses to larger-scale architectures and datasets. Moreover, it should explore alternative variational bounds or more stable training approaches for conditional Mutual Information estimation: it is expected to represent the information flow in a neural network setup more comprehensively.

References

[1] I. Belghazi, S. Rajeswar, A. Baratin, R. D. Hjelm, and A. C. Courville. *MINE: mutual information neural estimation.* CoRR **abs/1801.04062** (2018).

[2] I. Butakov, A. Tolmachev, S. Malanchuk, A. Neopryatnaya, and A. Frolov. Mutual information estimation via normalizing flows, (2024).

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, Hoboken, N.J, second edition, (April 2005).

[4] P. Czyż, F. Grabowski, J. E. Vogt, N. Beerenwinkel, and A. Marx. Beyond normal: on the evaluation of mutual information estimators. In 'Proceedings of the 37th International Conference on Neural Information Processing Systems', NIPS '23, Red Hook, NY, USA, (2024). Curran Associates Inc.

[5] L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using real nvp, (2017).

[6] M. D. Donsker and S. R. S. Varadhan. *Asymptotic evaluation of certain markov process expectations for large time. iv.* Communications on Pure and Applied Mathematics **36** (March 1983), 183–212.

[7] Z. Goldfeld, E. v. d. Berg, K. Greenewald, I. Melnyk, N. Nguyen, B. Kingsbury, and Y. Polyanskiy. *Estimating Information Flow in Deep Neural Networks.* (2018).

[8] M. Gutmann and A. Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In 'Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics', Y. W. Teh and M. Titterington, (eds.), volume 9 of *Proceedings of Machine Learning Research*, 297–304, Chia Laguna Resort, Sardinia, Italy, (13–15 May 2010). PMLR.

[9] D. McAllester and K. Stratos. Formal limitations on the measurement of mutual information, (2020).

[10] L. McInnes, J. Healy, and S. Astels. *hdbscan: Hierarchical density based clustering.* The Journal of Open Source Software **2** (03 2017).

[11] S. Molavipour, G. Bassi, and M. Skoglund. *Conditional Mutual Information Neural Estimator.* (2019).

[12] S. Molavipour, G. Bassi, and M. Skoglund. *Neural Estimators for Conditional Mutual Information Using Nearest Neighbors Sampling.* IEEE Trans. Signal Process., vol. 69, pp. 766-780, Jan. 2021 **69** (June 2020), 766–780.

[13] X. Nguyen, M. J. Wainwright, and M. Jordan. Estimating divergence functionals and the likelihood ratio by penalized convex risk minimization. In 'Advances in Neural Information Processing Systems', J. Platt, D. Koller, Y. Singer, and S. Roweis, (eds.), volume 20. Curran Associates, Inc., (2007).

[14] L. Paninski. *Estimation of Entropy and Mutual Information.* Neural Computation **15** (June 2003), 1191–1253.

[15] K. Sakamoto and I. Sato. End-to-End Training Induces Information Bottleneck through Layer-Role Differentiation: A Comparative Analysis with Layer-wise Training, (2024).

[16] A. M. Saxe, Y. Bansal, J. Dapello, M. Advani, A. Kolchinsky, B. D. Tracey, and D. D. Cox. *On the information bottleneck theory of deep learning.* Journal of Statistical Mechanics: Theory and Experiment **2019** (December 2019), 124020.

[17] D. Shin and H. S. Kim. *An experimental study of neural estimators of the mutual information between random vectors modeling power spectrum features.* EURASIP Journal on Advances in Signal Processing **2024** (Jan 2024), 2.

[18] R. Shwartz-Ziv and N. Tishby. *Opening the Black Box of Deep Neural Networks via Information.* (March 2017).

[19] J. Song and S. Ermon. Understanding the limitations of variational mutual information estimators, (2020).

[20] J. Tanner and G. Ughi. *Mutual Information of Neural Network Initialisations: Mean Field Approximations.* (2021).

[21] N. Tishby, F. C. Pereira, and W. Bialek. *The information bottleneck method.* (April 2000).

[22] A. van den Oord, Y. Li, and O. Vinyals. Representation learning with contrastive predictive coding, (2019).

[23] M. Vera, L. Rey Vega, and P. Piantanida. *Information flow in Deep Restricted Boltzmann Machines: An analysis of mutual information between inputs and outputs.* Neurocomputing **507** (October 2022), 235–246.

[24] X. Wang, A. Al-Bashabsheh, C. Zhao, and C. Chan. *Smoothed noise contrastive mutual information neural estimation.* Journal of the Franklin Institute **360** (2023), 12415–12435.

Geometry Optimization of Idealized Total Cavopulmonary Connection Using a CFD-Based Framework*

Jan Bureš
buresj11@fjfi.cvut.cz

study programme: Mathematical Engineering
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Radek Fučík, Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. A computational fluid dynamics-based framework for optimizing 3D geometry in an idealized total cavopulmonary connection (TCPC) is presented. The TCPC is a surgical procedure designed to treat congenital heart defects involving a single functional ventricle. The presented custom optimization framework integrates Python-based geometry generation, lattice Boltzmann method (LBM) simulations, and gradient-free optimization algorithms, including Nelder–Mead and the Mesh Adaptive Direct Search methods. The three optimization steps – generation of parameterized 3D geometry, simulation of incompressible Newtonian fluid flow with a rigid wall, and evaluation of objective functions – are executed automatically. The massively parallel implementation of LBM on GPUs allows the use of a spatial resolution suitable for optimizing the flow metrics sensitive to the actual resolution, such as the turbulent kinetic energy or near-wall shear rate. A simplified, parameterized model of the TCPC geometry was used to test the framework, demonstrating its feasibility and effectiveness. While this study focuses on idealized geometries with simplified assumptions, the results provide a foundation for extending the framework to patient-specific data and more complex physiological scenarios. This work represents a step in applying computational optimization to cardiovascular surgery, with the potential to improve clinical outcomes and patient-specific treatment planning.

Keywords: mathematical optimization, modeling of blood flow, lattice Boltzmann method (LBM), shape optimization, total cavopulmonary connection (TCPC)

Abstrakt. Představujeme rámec pro optimalizaci 3D geometrie idealizovaného úplného kavopulmonálního spojení (TCPC) využívající výpočetní dynamiku tekutin. TCPC je chirurgický postup určený k léčbě vrozených srdečních vad s jednokomorovou cirkulací. Navržený optimalizační rámec integruje generování geometrie v Pythonu, numerické simulace proudění pomocí mřížkové Boltzmannovy metody (LBM) a bezgradientní optimalizační algoritmy, jmenovitě Nelderovu–Meadovu metodu a Mesh Adaptive Direct Search. Tři kroky optimalizace – generování parametrizované 3D geometrie, simulace proudění nestlačitelné newtonovské tekutiny stěnami a vyhodnocení účelových funkcí – jsou prováděny automaticky. Paralelní implementace LBM na GPU umožňuje použít prostorové rozlišení vhodné pro optimalizaci metrik proudění citlivých na zvolené rozlišení, jako je turbulentní kinetická energie nebo smyková rychlost v blízkosti stěn. K

*The work has been supportedby the Ministry of Education, Youth and Sports of the Czech Republic under the Inter-Excellence grant no. LUAUS25049 and by the project SGS23/188/OHK4/3T/14 of the Student Grant Agency of the Czech Technical University in Prague.

otestování rámce byl použit zjednodušený parametrizovaný model geometrie TCPC, na kterém byla demostrována proveditelnost a účinnost postupu. Ačkoli se studie zaměřuje na idealizované geometrie se zjednodušenými předpoklady, výsledky poskytují základ pro rozšíření rámce na specifická data pacientů a složitější fyziologické scénáře. Tato práce představuje krok směrem k aplikaci metod optimalizace v kardiochirurgii s potenciálem vylepšit klinické výsledky a plánování léčby přizpůsobené pacientovi.

Klíčová slova: matematická optimalizace, modelování proudění krve, mřížková Boltzmannova metoda (LBM), optimalizace tvarů, úplné kavopulmonální spojení (TCPC)

Full paper: J. Bureš, R. Fučík, R. Chabiniok. *Geometry Optimization of Idealized Total Cavopulmonary Connection Using a CFD-Based Framework.* In: Functional Imaging and Modeling of the Heart 2025; 397–406.

Variation on the Theme of Jarzynski's Inequality: Applications to Non-equilibrium Chemical Systems

Dani R. Castellanos
castedan@cvut.cz

study programme: Quantum Technologies
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Petr Jizba, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. This study explores the Jarzynski inequality (JI), a fundamental result in non-equilibrium statistical mechanics that connects equilibrium free energy with non-equilibrium work statistics. We systematically investigate JI across different thermodynamic regimes, from linear response to far-from-equilibrium scenarios, with particular emphasis on isothermal-isobaric and isothermal-isochoric processes in chemical systems. The theoretical framework establishes connections between JI, the maximum work theorem, and entropy production principles. Our analytical developments are complemented by numerical investigations of the Brusselator model, a paradigmatic chemical oscillator system exhibiting complex non-equilibrium dynamics. Through extensive simulations analyzing multiple trajectories with varying winding numbers, we rigorously verify that work statistics obey JI even in strongly non-equilibrium conditions. The work provides new methodological approaches for studying non-equilibrium thermodynamics in chemical and biological systems, demonstrating the robustness and versatility of JI across diverse thermodynamic conditions.

Keywords: Jarzynski inequality, non-equilibrium thermodynamics, Brusselator model, chemical oscillations

1 Introduction

The Jarzynski equality (JE), introduced by Christopher Jarzynski in 1997, represents one of the most profound results in contemporary non-equilibrium statistical physics. This remarkable equality establishes a fundamental bridge between equilibrium thermodynamics and non-equilibrium dynamics, enabling the determination of equilibrium free energy differences from non-equilibrium work measurements. The JE has been experimentally verified across diverse physical systems, including single biomolecules, mesoscopic mechanical systems, and quantum systems, demonstrating its universal character.

The original Jarzynski equality states:

$$\langle e^{-\beta W} \rangle_\gamma = e^{-\beta \Delta F} \tag{1}$$

where $\beta = (k_B T)^{-1}$ is the inverse temperature (with $k_B = 1$ in natural units), W denotes the work performed on the system during a non-equilibrium process, ΔF represents the

equilibrium free energy difference between initial and final states, and $\langle \cdots \rangle_\gamma$ indicates an average over an ensemble of process realizations.

Application of Jensen's inequality, which states that $\langle e^x \rangle \geq e^{\langle x \rangle}$ for any random variable x, to equation (1) yields the Jarzynski inequality (JI):

$$\langle W \rangle_\gamma \geq \Delta F \tag{2}$$

This inequality, while mathematically simpler than the equality, possesses broader applicability and provides crucial insights into the thermodynamic constraints governing non-equilibrium processes. The saturation of JI occurs only in the quasi-static limit, where all process realizations become identical and reversible.

The relationship between JI and the classical maximum work theorem (MWT) warrants careful examination. While both inequalities express similar thermodynamic bounds, their foundational principles differ significantly. MWT emerges from Clausius' inequality within macroscopic thermodynamics, whereas JI derives from microscopic Hamiltonian dynamics and statistical mechanics. This distinction becomes particularly important when considering systems with strong fluctuations or those operating far from equilibrium.

This paper presents a comprehensive investigation of JI, with particular emphasis on chemical systems operating under various thermodynamic constraints. Our work addresses several key aspects:

- Theoretical development of JI variants for different thermodynamic processes (isothermal-isobaric and isothermal-isochoric)
- Analytical connections between JI, entropy production, and fluctuation theorems
- Numerical verification of JI in the Brusselator model, a prototypical chemical oscillator system
- Methodological developments for analyzing work statistics in complex non-equilibrium systems

The Brusselator model, introduced by Prigogine and colleagues, provides an ideal test bed for investigating JI in far-from-equilibrium conditions due to its rich dynamical behavior, including stable limit cycles and complex oscillation patterns.

2 Theoretical Framework

2.1 Maximum Work Theorem: Thermodynamic Foundations

The maximum work theorem represents a cornerstone of equilibrium thermodynamics, establishing fundamental limits on energy conversion processes. Consider an isolated system comprising three subsystems: the system of interest, an environment at constant temperature T_0 and pressure p_0, and a work source. The system undergoes a process from initial state X to final state Y, potentially involving irreversible transformations.

The first law of thermodynamics for the system reads:

$$\Delta U = -W_{\text{nm}} - p_0 \Delta V - \Delta Q_0 \tag{3}$$

where W_{nm} represents non-mechanical work, $p_0 \Delta V$ denotes mechanical work against the environment, and ΔQ_0 is the heat exchange with the environment.

The second law, expressed through Clausius' inequality, provides:

$$\Delta S + \Delta S_0 + \Delta S' \geq 0 \tag{4}$$

where S, S_0, and S' represent the entropies of the system, environment, and work source, respectively. For the work source, $\Delta S' = 0$, while for the environment, $\Delta S_0 = \Delta Q_0 / T_0$.

Combining these relations yields the fundamental inequality:

$$W_{\text{nm}} \leq -\Delta(U - T_0 S + p_0 V) = -\Delta G \tag{5}$$

where G denotes the Gibbs free energy. This result establishes that the maximum non-mechanical work extractable from a system undergoing a process between equilibrium states equals the decrease in Gibbs free energy, with equality holding only for reversible processes.

For processes at constant temperature and arbitrary pressure, the appropriate potential becomes the Helmholtz free energy $F = U - TS$, leading to:

$$\tilde{W} \leq -\Delta F \tag{6}$$

where $\tilde{W} = W_{\text{nm}} + p_0 \Delta V$ represents the total work performed on the system.

2.2 Path Integral Derivation of Jarzynski's Equality

The derivation of Jarzynski's equality within the path integral formalism provides deep insights into its statistical mechanical foundations. We consider a quantum system coupled to a thermal environment and subject to external driving through a time-dependent parameter $\lambda(t)$.

The Feynman-Vernon influence functional approach allows us to describe the reduced dynamics of the system after integrating out environmental degrees of freedom. The reduced density matrix evolves as:

$$\rho_r(x, x'; t) = \int dy d\bar{y}\, \mathcal{J}_r(x, x', t | y, \bar{y}, 0) \rho_S(y, \bar{y}; 0) \tag{7}$$

where the evolution kernel incorporates the influence functional:

$$\mathcal{J}_r(x, x', t | y, \bar{y}, 0) = \int_{x(0)=y}^{x(t)=x} \mathcal{D}x \int_{\bar{x}(0)=\bar{y}}^{\bar{x}(t)=x'} \mathcal{D}\bar{x}\, e^{\frac{i}{\hbar}(S_S[x] - S_S[\bar{x}] + S_{IA}[x, \bar{x}])} \tag{8}$$

The influence action $S_{IA}[x, \bar{x}]$ encapsulates all environmental effects on the system dynamics. For a system initially in thermal equilibrium at inverse temperature β, the initial density matrix is:

$$\rho_S(y, \bar{y}; 0) = \frac{1}{Z_0} \langle y | e^{-\beta \hat{H}_S} | \bar{y} \rangle \tag{9}$$

The work performed during a process driven by parameter $\lambda(t)$ is defined as:

$$W = \int_0^\tau dt \, \frac{\partial H(\lambda(t))}{\partial \lambda} \dot{\lambda}(t) \tag{10}$$

The average of the exponential work over all trajectories yields:

$$\langle e^{-\beta W} \rangle_\gamma = \int \mathcal{D}z \, \mathcal{P}[z, \lambda] e^{-\beta W[z]} \tag{11}$$

Under the assumptions of Markovian dynamics, detailed balance, and initial thermal equilibrium, this expression simplifies dramatically:

$$\langle e^{-\beta W} \rangle_\gamma = \frac{Z(\lambda(\tau))}{Z(\lambda(0))} = e^{-\beta \Delta F} \tag{12}$$

This elegant result demonstrates that the exponentially averaged work depends only on equilibrium partition functions, independent of the specific non-equilibrium path taken by the system.

2.3 Connections and Distinctions Between JI and MWT

While Jarzynski's inequality and the maximum work theorem share formal similarities, their physical interpretations and domains of applicability differ significantly:

Table 1: Comparison between Maximum Work Theorem and Jarzynski's Inequality

Maximum Work Theorem	**Jarzynski's Inequality**
Derived from Clausius' inequality	Derived from microscopic dynamics
Applies to individual macroscopic processes	Applies to ensemble averages
Valid for systems with well-defined temperature	Valid for fluctuating microscopic systems
Based on thermodynamic entropy	Based on statistical mechanics
Limited to near-equilibrium for specific bounds	Valid arbitrarily far from equilibrium

These distinctions become particularly important when considering chemical systems with strong fluctuations or those operating far from thermodynamic equilibrium.

3 Chemical Systems in Linear Response Regime

3.1 Entropy Production in Chemical Networks

For chemical reaction networks, the second law of thermodynamics finds expression through the entropy production rate. Considering a system exchanging heat with a reservoir at temperature T, the entropy balance reads:

$$\frac{dS_s}{dt} = \frac{dS_e}{dt} + \frac{dS_i}{dt} \tag{13}$$

where S_s represents system entropy, S_e denotes entropy exchange with the environment, and S_i signifies entropy production from irreversible processes.

The entropy production rate satisfies:

$$\frac{dS_i}{dt} \geq 0 \tag{14}$$

with equality holding only for reversible processes.

For chemical reactions, the central quantity is the chemical affinity A, defined through the De Donder relation:

$$A = -\sum_k \gamma_k \mu_k \tag{15}$$

where γ_k represent stoichiometric coefficients (negative for reactants, positive for products) and μ_k denote chemical potentials.

The relationship between affinity and Gibbs free energy emerges naturally:

$$\left(\frac{\partial G}{\partial \xi}\right)_{p,T} = \sum_k \gamma_k \mu_k = -A(\xi) \tag{16}$$

where ξ represents the extent of reaction.

The entropy production rate for a single chemical reaction becomes:

$$\frac{dS_i}{dt} = \frac{A}{T}\frac{d\xi}{dt} \geq 0 \tag{17}$$

For networks comprising multiple reactions, this generalizes to:

$$\frac{dS_i}{dt} = \sum_n J_n X_n \geq 0 \tag{18}$$

where $J_n = d\xi_n/dt$ represent reaction fluxes and $X_n = A_n/T$ denote thermodynamic forces.

3.2 Fluctuation Theory and Linear Response

Near equilibrium, the Landau-Lifshitz theory of fluctuations provides a powerful framework for analyzing thermodynamic fluctuations. For small deviations from equilibrium, the entropy production associated with a fluctuation becomes:

$$\Delta S_i = \frac{1}{2T}\left(\frac{\partial A}{\partial \xi}\right)_e (\xi - \xi_e)^2 \tag{19}$$

The probability distribution for fluctuations follows the Gaussian form:

$$p(X) = \sqrt{\frac{b}{2\pi}} e^{-\frac{1}{2}bX^2} \tag{20}$$

where $X = \xi - \xi_e$ and $b = -T^{-1}(\partial A/\partial \xi)_e > 0$.

For multi-reaction systems, the covariance matrix of fluctuations relates to the Hessian of the Gibbs free energy:

$$\langle X_i X_k \rangle = (b^{-1})_{ik}, \quad b_{ik} = \frac{1}{T}\left(\frac{\partial^2 G}{\partial \xi_i \partial \xi_k}\right)_{p,T} \tag{21}$$

Within this linear response framework, we can derive JI-type inequalities:

$$\Delta G_{p,T} \leq \langle W_{\mathrm{nm}} \rangle_{\mathcal{C}} \tag{22}$$

$$\Delta F_T \leq \langle W \rangle_{\mathcal{C}} \tag{23}$$

These results demonstrate that JI emerges naturally from fluctuation theory in the linear response regime, providing a bridge between microscopic fluctuations and macroscopic thermodynamic bounds.

4 Far-from-Equilibrium Chemical Systems

4.1 Extended Thermodynamic Formulation

For chemical systems operating far from equilibrium, standard equilibrium thermodynamics requires careful extension. While thermodynamic potentials like Gibbs and Helmholtz free energies are strictly defined only for equilibrium states, their utility extends to non-equilibrium situations through appropriate generalizations.

Consider a generic chemical reaction at constant temperature and pressure. The differential of Gibbs free energy reads:

$$dG_{T,p} = \left(\frac{\partial G}{\partial \xi}\right)_{T,p} d\xi \tag{24}$$

The temporal evolution follows:

$$\frac{dG_{T,p}}{dt} = \left(\frac{\partial G}{\partial \xi}\right)_{T,p} \frac{d\xi}{dt} \tag{25}$$

This relationship enables the computation of free energy changes from reaction kinetics, even when the system remains far from equilibrium throughout the process.

4.2 JI for Isothermal-Isobaric Processes

For chemical processes under constant temperature and pressure conditions, we develop a specialized formulation of JI. Starting from the fundamental thermodynamic identity:

$$A = -\left(\frac{\partial F}{\partial \xi}\right)_{T,p} - p\left(\frac{\partial V}{\partial \xi}\right)_{T,p} \tag{26}$$

Integration from initial to final reaction extents yields:

$$\int_{\xi_e}^{\xi} \left(\frac{\partial G}{\partial \xi}\right)_{T,p} d\xi \leq -\int_{\xi_e}^{\xi} \left(\frac{\partial F}{\partial \xi}\right)_{T,p} d\xi - p\int_{\xi_e}^{\xi} \left(\frac{\partial V}{\partial \xi}\right)_{T,p} d\xi \tag{27}$$

Rearranging terms reveals the JI structure:

$$\langle W \rangle_{T,p} \geq \langle \Delta F \rangle_{T,p} \tag{28}$$

This result demonstrates that JI maintains its validity for isothermal-isobaric chemical processes, with the work bound now involving both mechanical and non-mechanical contributions.

4.3 JI for Isothermal-Isochoric Processes

For systems at constant temperature and volume, the Helmholtz free energy provides the appropriate thermodynamic potential. The fundamental differential reads:

$$dF = -SdT - pdV - \delta W_{\text{nm}} \tag{29}$$

Under isothermal-isochoric conditions ($dT = 0$, $dV = 0$), this simplifies to:

$$dF = -\delta W_{\text{nm}} \tag{30}$$

Since spontaneous evolution decreases Helmholtz free energy ($dF \leq 0$), we obtain:

$$dF \leq -\delta W_{\text{nm}} \tag{31}$$

Integration along the reaction path yields:

$$\Delta F_{T,p} \leq \langle W \rangle_{T,p} \tag{32}$$

This formulation extends JI to constant-volume chemical processes, demonstrating its broad applicability across different thermodynamic constraints.

5 Brusselator Model: A Case Study

5.1 Model Description and Dynamical Properties

The Brusselator model, introduced by Prigogine and colleagues, represents a paradigmatic example of chemical systems exhibiting non-equilibrium self-organization and oscillatory behavior. The model comprises four elementary reactions:

$$A \rightarrow X \tag{33}$$
$$B + X \rightarrow Y + D \tag{34}$$
$$2X + Y \rightarrow 3X \tag{35}$$
$$X \rightarrow E \tag{36}$$

Species A and B represent constant-concentration reactants, D and E denote products, while X and Y are intermediate species whose dynamics generate the characteristic oscillatory behavior.

The kinetic equations for intermediate concentrations read:

$$\frac{dx}{dt} = A + x^2y - (B+1)x \tag{37}$$

$$\frac{dy}{dt} = Bx - x^2y \tag{38}$$

The autocatalytic step ($2X + Y \rightarrow 3X$) introduces essential non-linearity, while the system's openness (through constant A and B concentrations) enables sustained non-equilibrium behavior.

The Brusselator exhibits rich dynamical phases:

- Stable fixed point for $B < 1 + A^2$
- Hopf bifurcation at $B = 1 + A^2$
- Stable limit cycle for $B > 1 + A^2$

For our investigations, we focus on the oscillatory regime ($B > 1 + A^2$), where the system displays persistent periodic behavior far from thermodynamic equilibrium.

5.2 Numerical Methods and Work Computation

We implemented numerical simulations of the Brusselator dynamics using fourth-order Runge-Kutta integration with adaptive time stepping. To investigate JI, we analyzed ensembles of trajectories connecting specific states within the limit cycle.

The chemical affinity for the autocatalytic step was approximated as:

$$A(t) = -RT \ln\left(\frac{x(t)}{y(t)}\right) \tag{39}$$

where R denotes the gas constant and T represents temperature.

The reaction velocity was computed from the extent of reaction ξ, related to concentration changes through stoichiometry. The work performed along each trajectory was calculated as:

$$W = \int_{t_1}^{t_2} A(t)v(t)dt \tag{40}$$

where the integration extends from initial state A (local maximum in concentration) to final state B (subsequent local minimum).

We analyzed trajectories with different winding numbers around the limit cycle, corresponding to paths completing different numbers of oscillations before reaching the target state. This approach enables investigation of JI across diverse non-equilibrium pathways.

5.3 Statistical Analysis and Weighted Averaging

To account for the different probabilities associated with various trajectories, we employed a weighted averaging scheme based on the Riemann zeta function:

$$\langle W \rangle = \frac{\sum_{k=1}^{N} \frac{1}{k^s} W_k}{\sum_{k=1}^{N} \frac{1}{k^s}} \tag{41}$$

The parameter s modulates the weight distribution across trajectories, with larger s values emphasizing lower-index (more probable) paths. Optimization of s ensured rapid convergence of the work average while maintaining physical relevance.

We computed the free energy difference ΔF from equilibrium considerations, recognizing that for periodic systems, ΔF represents the reversible work required to transition between states A and B.

5.4 Results and JI Verification

Our numerical investigations yielded comprehensive verification of JI across diverse conditions:

- For all parameter sets examined, the inequality $\langle W \rangle \geq \Delta F$ held rigorously
- The work distribution exhibited significant broadening in the far-from-equilibrium regime
- Trajectories with higher winding numbers generally displayed larger work values
- The weighted averaging scheme provided stable convergence of work statistics

These results establish that JI maintains its validity even in strongly non-equilibrium chemical systems exhibiting complex oscillatory dynamics, providing crucial support for its general applicability.

6 Conclusions

This investigation has established several key results regarding the Jarzynski inequality and its applications to chemical systems: The Jarzynski inequality emerges as a powerful and versatile tool for analyzing non-equilibrium processes across diverse thermodynamic conditions. Its robustness in complex chemical systems suggests broad applicability to biological energy transduction, synthetic chemistry, and materials science, providing fundamental insights into the thermodynamic constraints governing non-equilibrium transformations.

References

[1] C. Jarzynski. *Nonequilibrium equality for free energy differences*. Phys. Rev. Lett. 78, 2690 (1997).

[2] C. Jarzynski. *Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach*. Phys. Rev. E 56, 5018 (1997).

[3] H.B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. John Wiley & Sons (1985).

[4] I. Prigogine. *Time, structure, and fluctuations*. Science 201, 777-785 (1978).

[5] L.D. Landau, E.M. Lifshitz. *Statistical Physics: Volume 5*. Elsevier (2013).

[6] R.P. Feynman, F.L. Vernon. *The theory of a general quantum system interacting with a linear dissipative system*. Ann. Phys. 24, 118-173 (1963).

[7] I. Prigogine. *Thermodynamics of Irreversible Processes*. Thomas (1955).

[8] L. Onsager. *Reciprocal relations in irreversible processes*. Phys. Rev. 37, 405 (1931).

[9] J. Liphardt et al. *Equilibrium information from nonequilibrium measurements*. Science 296, 1832-1835 (2002).

Shattering Triples with Six Permutations*

Daniela Černá†
daniela.cerna@cvut.cz

study programme: Mathematical Engineering
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jan Volec, Department of Theoretical Computer Science
Faculty of Information Technology, CTU in Prague

Abstract. Given six permutations $(\pi_1, \pi_2, \ldots, \pi_6)$ of the same ground set $[n]$ and a subset $X \subseteq [n]$ of size 3, we say that the triple X is shattered by $(\pi_1, \pi_2, \ldots, \pi_6)$ if every relative order of X is induced by exactly one of the π_i's. A natural extremal question is to determine the maximum number of triples of $[n]$ that six permutations can shatter. Using the flag algebra method, we prove that no six-tuple shatters more than $\frac{1}{2}\binom{n}{3} + O(n^2)$ triples. On the other hand, for every n, we construct six permutations of $[n]$ that shatter at least $\frac{482}{975}\binom{n}{3} \sim 0.4944\binom{n}{3}$ triples. These results improve the previously known bounds of Johnson and Wickes.

Keywords: flag algebra, permutation, shattering

Abstrakt. Nechť $(\pi_1, \pi_2, \ldots, \pi_6)$ je šestice permutací množiny $[n]$ a $X \subseteq [n]$ je podmnožina velikosti 3. Říkáme, že X je roztříštěná (*shattered*) permutacemi $(\pi_1, \pi_2, \ldots, \pi_6)$, pokud je každé možné vzájemné uspořádání prvků z X realizováno právě jednou z permutací π_i.

Přirozenou extremální otázkou je určit maximální počet trojic z množiny $[n]$, které může nějaká šestice permutací roztříštit. Pomocí metody *flag algeber* ukážeme, že žádná šestice permutací nemůže roztříštit více než $\frac{1}{2}\binom{n}{3} + O(n^2)$ trojic. Navíc pro každé n existuje konstrukce šesti permutací množiny $[n]$, které roztříští alespoň $\frac{482}{975}\binom{n}{3} \sim 0.4944\binom{n}{3}$ trojic.

Tyto výsledky vylepšují dříve známé odhady Johnson a Wickes.

Klíčová slova: flag abgebra, permutace, shattering

1 Introduction

The concept of shattering appears in different parts of combinatorics. In the settings of sets, we say that a finite set X is *shattered* by a family $\mathcal{F}$ if the intersections $\{X \cap F : F \in \mathcal{F}\}$ generate all possible subsets of X. There is a number of natural questions connected to this setting. For example, if we are given a family $\mathcal{F}$, what is the maximal size of X that can be shattered by this family? The answer to this question is called the VC dimension of $\mathcal{F}$. Other question is, given integers n and k, determine the size of a smallest family of subsets of $[n]$ such that every k-element subset of $[n]$ is shattered. In 1973, Kleitman and Spencer [3] proved that for any fixed k this number is between $d_1 \cdot 2^k \log n$ and $d_2 \cdot k2^k \log n$ for some positive reals d_1 and d_2.

*This work has been supported by the grant GAČR 23-06815M.

†This is a result of a joint work with: A. Cliffton, B. Kielak, J. Volec.

Shattering has also been studied in the setting of permutations. A k-element subset $X \subseteq [n]$ is shattered by a family of permutations $\mathcal{S} \subseteq S_n$ if the number of different relative orderings of X by the elements of $\mathcal{S}$ is $k!$. In 1971, Spencer [7] showed the existence of a family $\mathcal{S} \subseteq S_n$ of size $\frac{k}{\log_2(k!/(k!-1))} \log_2 n$ such that every k-element subset of $[n]$ is shattered. On the other hand, Füredi [1] proved that such a family must have size at least $\frac{(k-1)!}{2} \log_2 n$. Radhakrishnan [5] improved the bound of Füredi by a factor of approximately $2/\log_2 e$, and Tarui [8] gave an explicit construction for the case $k = 3$ of size $(2 + o(1)) \log_2 n$.

In this work, we focus on a multiplicity version of this problem, where we consider families $\mathcal{S} \subseteq S_n$ of size $k!$ and the aim is to maximize the number of shattered k-element subsets of $[n]$. For the case $k = 3$, i.e., when we shatter triples of $[n]$ with 6 permutations, Johnson and Wickes [2] proved the existence of permutations of $[n]$ that shatter at least $\frac{17}{42}\binom{n}{3}$ triples. On the other hand, they proved that if $n \geq 8$ then no $\mathcal{S} \subseteq S_n$ of size 6 can shatter more than $\frac{47}{60}\binom{n}{3}$ triples. Our results are the following improvements of both of the aforementioned bounds.

Theorem 1. *Every six permutations of $[n]$ shatter at most $\frac{1}{2}\binom{n}{3} + O(n^2)$ triples.*

Proposition 2. *For every $n \geq 3$, there exist six permutations of $[n]$ that shatter at least $\frac{482}{975}\binom{n}{3}$ triples.*

2 Definitions and notation

We start with the definition of a permutation pattern.

Definition 1. *Let $\pi \in S_n$ be a permutation, and $X = \{x_1, x_2, \ldots, x_k\} \subseteq [n]$ such that $x_1 < x_2 < \ldots < x_k$. The* pattern in π induced by X *is the permutation $\tau \in S_k$ satisfying*

$$\tau(i) < \tau(j) \Leftrightarrow \pi(x_i) < \pi(x_j)\,.$$

Example 1. *For example, let $\pi = (1, 3, 2, 5, 4) \in S_5$ and $X = \{2, 4, 5\}$. To obtain the pattern in π induced by X, we need to rewrite the values $(\pi(x_1), \pi(x_2), \pi(x_3)) = (3, 5, 4)$ to get a permutation from S_3 while keeping all inequalities intact. In this setting, the pattern is $\tau = (1, 3, 2)$.*

Definition 2. *We say that a family $\mathcal{S} \subseteq S_n$* shatters *a k-element subset $X \subseteq [n]$ if and only if $\{\tau$ pattern in π induced by $X : \pi \in \mathcal{S}\} = S_k$.*

It follows from the definition, that for $|\mathcal{S}| < |S_k| = k!$, there is no shattering. Using the notions defined, there are two natural questions we might ask. The first question regarding the minimum size of $\mathcal{S} \subset S_n$ that shatters every k-tuple from $[n]$ has been already well explored. Combining the results from [1] and [7], we know that when n grows while $k \geq 3$ is fixed

$$\frac{(k-1)!}{2} \log_2 n \leq |\mathcal{S}| \leq \frac{k}{\log_2 (k!/(k!-1))} \log_2 n.$$

Note that $\frac{k}{\log_2(k!/(k!-1))} \sim (k+1)! \ln 2$ for $n \to \infty$.

In this work, we focus on the second question – what is the maximum number of k-tuples that can be shattered if we fix the size of $\mathcal{S}$ to be the least possible that allows shattering, i.e. when $|\mathcal{S}| = k!$. This notion is captured by the following definition.

Definition 3. *For $\mathcal{S} \subseteq S_n$, we define $F_k(\mathcal{S})$ to be the number of k-element subsets of $[n]$ shattered by $\mathcal{S}$. For every $n \geq k$, we set*

$$F_k(n) := \max_{\substack{\mathcal{S}\subseteq S_n \\ |\mathcal{S}|=k!}} F_k(\mathcal{S}) \quad \textit{and} \quad f_k(n) := \frac{F_k(n)}{\binom{n}{k}} .$$

It is clear that $f_k(k) = 1$. A standard double counting argument also shows that $f_k(n)$ is nonincreasing as a function of n as can be seen in the next lemma.

Lemma 3. *Let $k \geq 2$. Then the function $f_k(\cdot)$ is non-increasing, i.e. $\forall n \geq k : f_k(n+1) \leq f_k(n)$.*

Proof. Let $\mathcal{S} \subset \mathcal{S}_{n+1}$ be a family of permutations attaining $f_k(n+1)$. For each $i \in [n+1]$, define $\hat{\mathcal{S}}_i$ to be the family consisting of

$$\{\pi_i \text{ pattern in } \pi \text{ induced by } [n+1] \setminus \{i\} : \pi \in \mathcal{S}\}.$$

Specially, each $\hat{\mathcal{S}}_i \subset S_n$ shatters at most $F_k(n)$ k-tuples. Now, if we count the number of k-tuples not containing i shattered by $\mathcal{S}$, it is the same as counting the number of k-tuples shattered by $\hat{\mathcal{S}}_i$. If we sum up all of the values, we are counting each k-tuple $n+1-k$-times. Therefore,

$$F_k(\mathcal{S}) = \frac{1}{(n+1-k)} \sum_{i=1}^{n+1} F_k(\hat{\mathcal{S}}_i) \leq \frac{1}{(n+1-k)} \sum_{i=1}^{n+1} F_k(n) = \frac{n+1}{n+1-k} F_k(n).$$

Dividing both sides by $\binom{n+1}{k}$ yields

$$f_k(n+1) = \frac{F_k(\mathcal{S})}{\binom{n+1}{k}} = \frac{F_k(\mathcal{S})(n+1-k)!k!}{(n+1)!} \leq F_k(n)\frac{(n-k)!k!}{n!} = f_k(n).$$

□

Thus the limit $\lim_{n\to\infty} f_k(n)$ exists. For brevity, we write $c_k := \lim_{n\to\infty} f_k(n)$.

Remark. *We note that our definition of the pattern in π induced by X differs from the one in [2], and if $\mathcal{S} = (\pi_1, \ldots, \pi_{k!})$ is a family that attains $F_k(n)$ in our definition, then $\mathcal{S}' = (\pi_1^{-1}, \ldots, \pi_{k!}^{-1})$ is the corresponding solution in their setting. In particular, the extremal questions are the same.*

Remark. *We may always assume that the first permutation is identity (or any other fixed permutation).*

Proof. Let $\tau' \in S_n$ be the desired permutation, and $\mathcal{S} = \{\pi_1, \pi_2, \ldots, \pi_\ell\}$ be any family of permutations from S_n. Set $\tau = \pi_1^{-1} \circ \tau'$, and $\mathcal{S}' := \{\pi_i \circ \tau : \pi_i \in \mathcal{S}\}$. Then $\mathcal{S}$ shatters $X = \{x_1, x_2, \ldots, x_k\}$ if and only if $\mathcal{S}'$ shatters $\tau^{-1}(X) = \{\tau^{-1}(x_1), \tau^{-1}(x_2), \ldots, \tau^{-1}(x_k)\}$. Therefore, $\mathcal{S}$ and $\mathcal{S}'$ shatter the same number of k-tuples. □

3 Upper bound

In this section, we show that every six permutations of $[n]$ shatter at most $\frac{1}{2}\binom{n}{3} + O(n^2)$ triples.

We use a versatile framework of Razborov [6] called *flag algebras* to show that $c_3 \leq \frac{1}{2}$. From the remark, we may assume that the first permutation is identity. Therefore, we use flag algebras with five-tuples of permutations, and count the number of triples ordered by those permutations in all the five possible ways that are not monotone increasing.

We apply flag algebras to convergent sequences of five-tuples of permutations, each tuple having permutations of the same order n. The substructure counting is given by picking k elements from $[n]$ uniformly at random, and in each permutation considering the probability distribution on the $k!$ patterns of order k. Let $\mathcal{F}_3$ be the set all of possible outcomes for $k = 3$. Clearly $|\mathcal{F}_3| = (3!)^5 = 7776$. Next, let $\mathcal{H} \subseteq \mathcal{F}_3$ be the set of all 120 five-tuples corresponding to the different orderings of $S_3 \setminus \{\mathrm{id}\}$. Note that the normalized count of shattered triples is equal to the sum of the densities in $\mathcal{H}$.

Our main technical result is the following:

Proposition 4. *There are pattern-density expressions $a_1, a_2, \ldots, a_5$ such that*

$$\left[\!\!\left[\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix}^T \begin{pmatrix} 3 & -1 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 & 1 \\ -1 & -1 & 3 & 1 & -1 \\ -1 & -1 & 1 & 3 & -1 \\ -1 & 1 & -1 & -1 & 3 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} \right]\!\!\right] = \sum_{F \in \mathcal{F}_3} d_F \cdot F \,,$$

where $d_F = -1$ for every $F \in \mathcal{H}$ and $d_F \in \left\{-\frac{1}{3}, 0, \frac{1}{3}, 1\right\}$ for every $F \in \mathcal{F} \setminus \mathcal{H}$.

Since the 5×5 matrix is positive definite and $d_F \leq 1$ for all $\mathcal{F} \setminus \mathcal{H}$, we derive the following asymptotic inequality for five-tuples of permutations:

$$\sum_{F \in \mathcal{F}_3 \setminus \mathcal{H}} F - \sum_{H \in \mathcal{H}} H \geq 0 \,.$$

Combining this with trivial equation

$$\sum_{F \in \mathcal{F}_3 \setminus \mathcal{H}} F + \sum_{H \in \mathcal{H}} H = \sum_{F \in \mathcal{F}_3} F = 1 \,,$$

we see that $c_3 = \sum_{H \in \mathcal{H}} H \leq \frac{1}{2}$. Moreover, when applying the established "limit" proof to a tuple of permutations on $[n]$, all the involved inequalities are valid up to an error of the order $O\left(\frac{1}{n}\right)$. Therefore, $F_3(n) \leq \frac{1}{2}\binom{n}{3} + O\left(n^2\right)$.

4 Lower bound and results for small n

If we want to find the lower bound for $f_k(n)$, it is enough to find the lower bound for c_k since $f_k(n) \geq c_k$ for all n. The following theorem from [2] allows us to blow-up contructions for small n to obtain the bound for the limit.

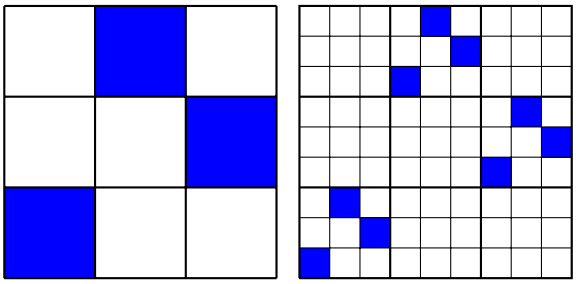


Figure 1: One iteration of blow-up of a permutation matrix (colored square symbolizes 1 at the position) starting with permutation matrix of $\pi = (1, 3, 2)$ and obtaining permutation matrix of $\pi^{(2)} = (1, 3, 2, 7, 9, 8, 4, 6, 5)$.

Theorem 5 ([2]). *Suppose $\mathcal{S}$ is a family that shatters $\alpha\binom{N}{k}$ k-tuples from $[N]$, then for all $n \geq N$*

$$f_k(n) \geq \frac{\alpha(N-1)!}{(N-k)!(N^{k-1}-1)}. \tag{1}$$

Proof. To prove this theorem, we need to first define the blow-up of the permutation $\pi \in \mathcal{S}_N$. Let $M \in \{0,1\}^{N,N}$ be the permutation matrix of π, meaning that $M_{i,j} = 1$ if $\pi(i) = j$ and it is 0 otherwise. One iteration of the blow-up means obtaining matrix $M^{(2)} \in \{0,1\}^{N^2,N^2}$ by replacing each 1 in M by a copy of M and each 0 of M by $N \times N$ zero matrix, see Figure 1. This is equivalent to saying $M^{(2)} = M \otimes M$, where $\otimes$ denotes the Kronecker product.

Let us iterate this blow-up r times. Let $\pi^{(r)}$ be the permutation defined by $M^{(r)}$ (matrix of size $N^r \times N^r$) and $X = \{x_1, x_2, \ldots, x_k\} \subset [N^r]$.

For all $j \in \{0, 1, \ldots, r-1\}$; let $x_i^{(j)} = \lfloor \frac{x_i - 1}{N^{r-1-j}} \rfloor \mod N$. This is equivalent to saying

$$x_i - 1 = (x_i^{(0)}, x_i^{(1)}, \ldots, x_i^{(r-2)}, x_i^{(r-1)})_N,$$

where $(\cdot)_N$ denotes the base N expansion of $x_i - 1$.

Suppose there exists ℓ such that for all $i \neq j$: $x_i^{(\ell)} \neq x_j^{(\ell)}$ while $x_i^{(k)} = x_j^{(k)}$ for all $k < \ell$. In this case, we can determine whether X is shattered using only the matrix M and the coordinates $\{x_1^{(\ell)}, x_2^{(\ell)}, \ldots, x_k^{(\ell)}\}$. For a fixed ℓ, there are exactly

$$N^\ell N(N-1)(N-2)\ldots(N-k+1)(N^{r-1-\ell})^k = N^{(r-1)k+(1-k)\ell}\frac{N!}{(N-k)!}$$

such k-tuples. Moreover, an α-fraction of these define a shattered k-tuple by the assumption.

Therefore,

$$\begin{aligned} f_k(N^r) &\geq \sum_{\ell=0}^{r-1} \frac{\alpha}{N^{rk}} N^{(r-1)k+(1-k)\ell}\frac{N!}{(N-k)!} = \frac{\alpha N!}{(N-k)!N^k}\frac{1-\frac{1}{N^{(k-1)r}}}{1-\frac{1}{N^{k-1}}} \\ &= \frac{\alpha(N-1)!}{(N-k)!(N^{k-1}-1)}\left(1-\frac{1}{N^{(k-1)r}}\right). \end{aligned}$$

Letting $r \to +\infty$ and using Lemma 3 gives us the result. □

n	$F_3(n)$	$f_3(n)$	lower bound on c_3	
5	8	$\frac{4}{5}$	$\frac{2}{5}$	$\doteq 0.4$
6	16	$\frac{4}{5}$	$\frac{16}{35}$	$\doteq 0.4571$
7	26	$\frac{26}{35}$	$\frac{13}{28}$	$\doteq 0.4643$
8	40	$\frac{5}{7}$	$\frac{10}{21}$	$\doteq 0.4762$
9	57	$\frac{19}{28}$	$\frac{19}{40}$	$\doteq 0.475$

Table 1: The extremal values for $5 \leq n \leq 9$ and the resulting bound on c_3.

$\pi_1 = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26)$
$\pi_2 = (3, 2, 1, 22, 23, 21, 25, 24, 26, 11, 10, 17, 16, 15, 14, 13, 12, 20, 19, 18, 4, 6, 5, 8, 9, 7)$
$\pi_3 = (19, 20, 18, 22, 21, 23, 9, 8, 7, 14, 17, 12, 16, 10, 13, 11, 15, 2, 1, 3, 5, 4, 6, 26, 25, 24)$
$\pi_4 = (19, 18, 20, 6, 5, 4, 25, 26, 24, 17, 14, 15, 11, 16, 12, 13, 10, 2, 3, 1, 23, 22, 21, 8, 7, 9)$
$\pi_5 = (18, 21, 20, 16, 11, 14, 6, 10, 7, 4, 25, 24, 2, 23, 1, 26, 3, 17, 22, 19, 13, 15, 12, 9, 5, 8)$
$\pi_6 = (21, 18, 19, 11, 16, 13, 9, 5, 8, 25, 4, 1, 23, 3, 26, 2, 24, 22, 17, 20, 15, 12, 14, 6, 10, 7)$

Table 2: Six permutations of [26] that shatter 1446 triples out of 2600.

Specially in our setting, the theorem gives us

$$c_3 \geq f_3(N)\frac{N-2}{N+1}.$$

First, we determined the extremal values for $n \leq 8$ as well as all extremal configurations using an exhaustive computer search.

Using the double counting argument in the proof of Lemma 3, we can see that any configuration for $n = 9$ with at least 58 triples shattered contains at least one subconfiguration for $n' = 8$ with either 39 or 40 shattered triples:

$$58 \leq F_3(\mathcal{S}) \leq \frac{1}{6}\sum_{i=1}^{9} F_3(\hat{\mathcal{S}}_i) \quad \Longrightarrow \quad 38 < \frac{1}{9}\sum_{i=1}^{9} F_3(\hat{\mathcal{S}}_i).$$

Since it possible to generate all the configurations for n' with at least 39 shattered triples, inspecting all their possible extensions by one point reveals that $F_3(9) = 57$. These results, together with the corresponding lower bound on c_3 from Theorem 5, are in Table 1.

Using a heuristic computer search for larger permutations, we have found six permutations of [26] that shatter 1446 triples out of 2600; see Table 2. In particular, it holds that $f_3(26) \geq \frac{723}{1300}$ and $c_3 \geq \frac{482}{975} \sim 0.4944$ by Theorem 5.

5 Related problems

While studying the original problem, we have solved two closely related problems, where we impose some additional symmetry assumptions on the six-tuple $(\pi_1, \pi_2, \ldots, \pi_6)$.

Firstly, we consider the problem of maximum number of shattered triples, where $\pi_1 = \mathrm{Id}$ and $\pi_2 = \pi_{\text{decreasing}}$. Secondly, we restrict the considered six-tuples even further and assume $\pi_1 = \mathrm{Id}, \pi_2 = \pi_{decreasing}$ and $\pi_4 = \pi_{decreasing} \circ \pi_3$, $\pi_6 = \pi_{decreasing} \circ \pi_5$.

Problem 1. *What is the maximal density of shattered triples by the family $\mathcal{S}$ of the form $\mathcal{S} = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\} \subset S_n$, where $\pi_1 = (1, 2, 3, \ldots, n)$ and $\pi_2 = (n, n-1, \ldots, 2, 1)$?*

Problem 2. *What is the maximal density of shattered triples by the family $\mathcal{S}$ of the form $\mathcal{S} = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\} \subset S_n$, where $\pi_1 = (1, 2, 3, \ldots, n)$, and $\pi_{2i} = (n, n-1, \ldots, 2, 1) \circ \pi_{2i-1}$ for $i \in [3]$?*

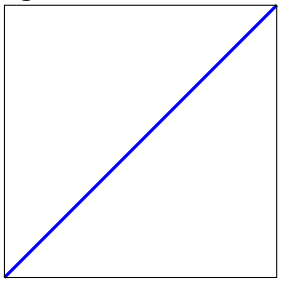
5.1 Problem 2

In this setting, we want to find maximal density of shattered triples when we have π_1, π_2, π_3, and π_4, π_5, π_6 are the "mirror images". In this setting, mirror image means the application of the decreasing permutation in a finite setting, or flipping along the axis $y = \frac{1}{2}$ for the permutons.

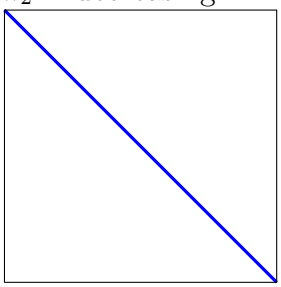
Proposition 6. *The maximal density of shattered triples in Problem 2 is $\frac{1}{4} + o(1)$.*

The flag-algebra framework yields that the density of shattered triples is at most $\frac{1}{4} + o(1)$. In the rest of this section, we will give a construction that attains $\frac{1}{4}$ in the limit. The extremal set $\mathcal{S} = \{\pi_1, \pi_2, \pi_3, \ldots, \pi_6\}$ contains permutations, whose permutation matrices have following structure, where T means that the permutation is on the top branch, B means the bottom branch:

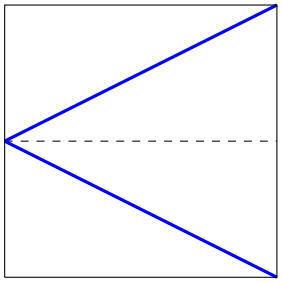
$\pi_1 = \mathrm{Id}$:



$\pi_2 = \text{decreasing}$:



2×

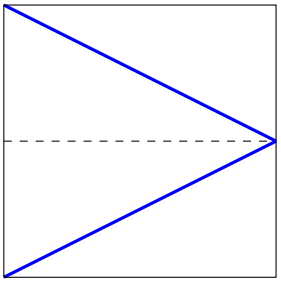


π_3: T for $x = 0 \mod 2$,

π_4: T for $x = 1 \mod 2$.

- TBT $\to (2, 1, 3)$
- BTB $\to (2, 3, 1)$

2×



π_5: T for $x = 0 \mod 2$,

π_6: T for $x = 1 \mod 2$.

- TBT $\to (3, 1, 2)$
- BTB $\to (1, 3, 2)$

Any triple from $[n]$ induces pattern $(1, 2, 3)$ in π_1 and pattern $(3, 2, 1)$ in π_2. For example, the triple attaining TBT in π_3 induces the pattern $(2, 1, 3)$.

The permutations were chosen in such a way, that π_3 and π_5 are always on the same branch while π_4 and π_6 are on the opposite branch.

For any triple attaining positions TBT in π_3, which happens for $\left(\frac{1}{8} - o(1)\right)\binom{n}{3}$ triples, we get TBT in π_5 and BTB in π_4 and π_6. Specially, we get all four non-monotone patterns from S_3 and such triple is shattered. By similar argument, it can be shown that a triple is shattered also for BTB in π_3, which happens with the same probability.

Therefore, at least $\left(\frac{1}{4} - o(1)\right)\binom{n}{3}$ triples are shattered proving Proposition 6.

5.2 Problem 1

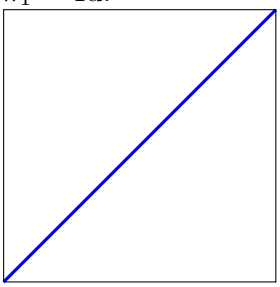
In this setting, we need to find 4 permutations on $[n]$ such that many triples induce the non-monotone patterns, i.e. we need to see $\{(1,3,2),(2,1,3),(2,3,1),(3,1,2)\}$.

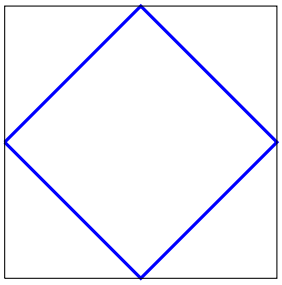
Proposition 7. *The maximal density of shattered triples in Problem 1 is* $\frac{1}{3} + o(1)$.

The Flag-algebra framework yields that the density of shattered triples in this setting is at most $\frac{1}{3} + o(1)$. In the rest of the section, we will show the contruction attaining $\frac{1}{3} - o(1)$.

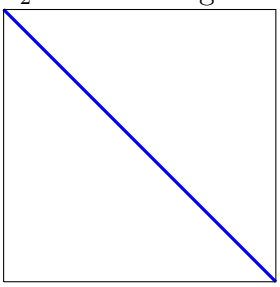
The extremal set $\mathcal{S} = \{\pi_1, \pi_2, \pi_3, \ldots, \pi_6\}$ contains permutations, whose permutation matrices have following structure, where T means that the permutation is on the top branch, B means the bottom branch:

$\pi_1 =$ Id:





$\pi_2 =$ decreasing:



π_3: T for $x = 0, 1, 2 \mod 6$

π_4: T for $x = 0, 3, 4 \mod 6$

π_5: T for $x = 1, 3, 5 \mod 6$

π_6: T for $x = 2, 4, 5 \mod 6$

Now, looking at a triple, there are four possible groups of patterns we might see depending on the relative positions of the x coordinates. The pattern groups are depicted in Figure 2. We will model the groups using bars of different lengths modelling the distances between the top and bottom branch. The actual pattern is then obtained by choosing the top point or the bottom point from the bar depending on whether the permutation chooses top branch or the bottom branch.

For example, the first pattern appears whenever the first two coordinates are in the left half and the third coordinate is closer to the middle than the second one. The last pattern is obtained analogously. Note that each of the groups appear for $\left(\frac{1}{4} + o(1)\right)\binom{n}{3}$ triples.

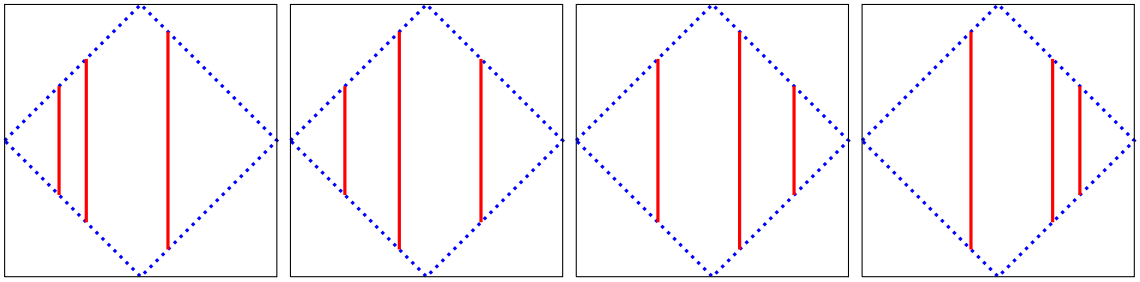


Figure 2: Four groups of patterns depending on the relative positions of the x coordinates of the triple.

Let us now look at the second group. In this case, the middle branch is the longest while the left one is the shortest. The possible patterns that can be obtained by choosing either the top of the bar (T) or the bottom of the bar (B) are

- TTT, BTT $\to$ 132,
- TTB, BTB $\to$ 231,
- TBT, BBT $\to$ 213,
- TBB, BBB $\to$ 312.

We can see that it does not matter whether the first point is T or B. To get all four non-monotone patterns, we need to have all four options {TT, TB, BT, BB} happening in the last two bars. In total, there are 36 possible options (all of equal density in the limit) for the values mod 6 of the coordinates of x_2 and x_3, and exactly 24 of them are favourable. Namely, they are

$$\{(0,1),(1,0),(0,2),(2,0),(0,3),(3,0),(0,4),(4,0),(1,2),(2,1),(1,3),(3,1),\\ (1,5),(5,1),(2,4),(4,2),(2,5),(5,2),(3,4),(4,3),(3,5),(5,3),(4,5),(5,4)\}.$$

Therefore, the six-tuple shatters $\frac{24}{36} - o(1) = \frac{2}{3} - o(1)$ of triples that are in this group.

Analogously, the six-tuple shatters $\frac{2}{3} - o(1)$ of the triples that fall into the third group. Putting these together, we can see that this family of permutations shatters at least

$$2 \times \frac{1}{4} \times \frac{2}{3} - o(1) = \frac{1}{3} - o(1) \text{ triples.}$$

6 Conclusion

In this work, we have shown that $c_3 \in \left[\frac{482}{975}, \frac{1}{2}\right]$. We conjecture that the upper bound is correct.

Conjecture. *For every n, there exist six permutations of $[n]$ that shatter at least $\frac{1}{2}\binom{n}{3}$ triples.*

We have also determined the exact values of $f_3(n)$ for $n \in \{6, 7, 8, 9\}$ and solved two auxiliary problems closely related to the original one.

We believe it would be interesting to have a better understanding of c_k for $k \geq 4$. Using the result of Spencer [7] and the monotonicity of $f_k(n)$, Johnson and Wickes [2] observed that $c_k < 1$. On the other hand, Levenshtein [4] constructed $k!$ permutations of $[k+1]$ that shatter all the k-element subsets; the proof can be also found in [2]. Therefore, the iterated blow-up of the Levenshtein construction yields $c_k \geq \frac{k!}{(k+1)^{k-1}-1}$. So far, we are not able to say anything substantially better than these two bounds.

References

[1] Z. Füredi. *Scrambling permutations and entropy of hypergraphs.* Random Structures and Algorithms **8** (1996), 97–104.

[2] R. Johnson and B. Wickes. *Shattering k-sets with permutations.* Order **41** (2023), 419–436.

[3] D. Kleitman and J. Spencer. *Families of k-independent sets.* Discrete Mathematics **6** (1973), 255–262.

[4] V. I. Levenshtein. *Perfect codes in the metric of deletions and insertions.* Discrete Mathematics and Applications **2** (1992), 241–258.

[5] J. Radhakrishnan. *A note on scrambling permutations.* Random Structures and Algorithms **22** (2003), 435–439.

[6] A. A. Razborov. *Flag algebras.* Journal of Symbolic Logic **72** (2007), 1239–1282.

[7] J. Spencer. *Minimal scrambling sets of simple orders.* Acta Mathematica Academiae Scientiarum Hungaricae **22** (1971), 349–353.

[8] J. Tarui. *On the minimum number of completely 3-scrambling permutations.* Discrete Mathematics **308** (2008), 1350–1354.

Benchmarking and Transfer Learning for Hyperparameter Optimization of Graph Neural Networks

Marek Dědič
marek.dedic@fjfi.cvut.cz

study programme: Mathematical Engineering
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:

Martin Holeňa, Department of Artifical Intelligence
Institute of Computer Science, CAS

Lukáš Bajer, Cisco Systems, Inc.

Abstract. Machine learning on graphs is an ever-more popular approach to learning on structured data. However, a critical element for successful model deployment is hyperparameter optimization (HPO), as the choice of parameters (such as learning rate or hidden dimensions) can drastically impact performance. Although classical HPO methods like grid and random search are popular for simplicity, sophisticated strategies such as Gaussian-process Bayesian optimization (GP) and the Tree-structured Parzen Estimator (TPE) promise greater efficiency by adaptively modeling the search space. Despite extensive benchmarking in standard domains (like image and text tasks), there has been no systematic study detailing how these HPO techniques perform specifically on general graph-structured problems.

To address this gap, this work presents two main contributions: first, we conduct the first systematic HPO benchmark on Graph Neural Networks (GNNs), evaluating the convergence speed and efficacy of five classical HPO methods (grid search, random search, GP, TPE, and Sobol Quasi-Monte Carlo) for tuning core GraphSAGE hyperparameters across ten standard graph datasets that vary in scale and structural properties. Second, we evaluate a meta-learning approach for hyperparameter transfer by introducing a cross-dataset Random Forest surrogate model. This model learns the interaction between dataset properties (e.g., node count, homophily) and hyperparameter performance by pooling past HPO results. This accumulated meta-knowledge is then used to "warm-start" the optimization process on new, previously unseen datasets, thereby reducing the number of costly model evaluations required to reach peak performance.

Our comprehensive empirical evaluation yielded several key findings regarding the classical HPO algorithms. We found that GP and TPE consistently dominated the benchmark, securing the highest score on seven of the ten graph datasets and never ranking lower than third. Classical exhaustive methods (grid and random search) remained robust fallbacks, typically trailing the Bayesian leaders by approximately one percentage point. Sobol QMC, however, underperformed, finishing last on six datasets, suggesting that its theoretical uniformity advantage diminishes rapidly when applied to our eight-hyperparameter search space which requires crucial feedback for effective tuning.

Furthermore, the meta-surrogate model demonstrated significant gains in efficiency compared to the strong GP baseline. By leveraging transfer knowledge, the meta-surrogate reached

its peak performance in fewer trials than GP on several tasks, confirming a warm-start advantage. The method also achieved similar or superior final scores, particularly on the largest or most structurally distinct graphs (ArXiv, DBLP, Squirrel), and generally exhibited reduced variability across seeds, indicating more stable sampling of high-quality configurations.

Keywords: Graph Neural Networks, Hyperparameter Optimization, Meta-Learning, Sequential Model-Based Optimization

Abstrakt. Strojové učení na grafech je populárním přístupem pro učení na strukturovaných datech. Kritickým prvkem pro úspěšné nasazení modelu je však optimalizace hyperparametrů (HPO), jelikož volba hyperparametrů (jako je learning rate nebo šířka skrytých vrstev) může dramaticky ovlivnit výkon. Ačkoli jsou klasické HPO metody jako grid search a random search populární pro svou jednoduchost, sofistikovanější strategie jako Bayesovská optimalizace s Gaussovskými procesy (GP) a Tree-structured Parzen Estimator (TPE) slibují vyšší efektivitu díky adaptivnímu modelování prohledávaného prostoru. Navzdory rozsáhlým srovnávacím studiím ve standardních oblastech (jako jsou úlohy se zpracováním obrazu a textu) dosud neexistuje žádná systematická studie podrobně popisující, jak si tyto HPO techniky vedou specificky u obecných problémů s grafovou strukturou.

Abychom tuto mezeru zaplnili, přináší tato práce dva hlavní příspěvky: zaprvé představujeme první systematickou srovnávací studii (benchmark) HPO pro grafové neuronové sítě (GNN), kde srovnáváme rychlost konvergence a účinnost pěti klasických HPO metod (grid search, random search, GP, TPE a Sobolova kvazi-Monte Carlo metoda) pro optimalizaci klíčových hyperparametrů modelu GraphSAGE na deseti standardních grafových datasetech, které se liší rozsahem a strukturálními vlastnostmi. Za druhé představujeme přístup pomocí mezidatasetového surogátního modelu typu Náhodný les (Random Forest). Tento model se učí interakci mezi vlastnostmi datasetu (např. počet uzlů, homofilie) a hyperparametry tím, že shromažďuje výsledky z minulých běhů. Tato nashromážděná meta-data jsou poté použita k „teplému startu“ (warm-start) optimalizačního procesu na nových, dosud neviděných datasetech, čímž se snižuje počet nákladných evaluací modelu potřebných k dosažení optimálního výkonu.

Naše empirické vyhodnocení přineslo několik klíčových zjištění týkajících se klasických HPO algoritmů. Zjistili jsme, že metody GP a TPE konzistentně vedly ve srovnání, dosáhly nejvyššího skóre na sedmi z deseti grafových datasetů a nikdy se neumístily hůře než na třetím místě. Klasické metody (grid search a random search) zůstaly robustními záložními řešeními, obvykle zaostávající za bayesovskými metodami přibližně o jeden procentní bod. Sobolova QMC metoda si však vedla špatně a skončila poslední na šesti datasetech, což naznačuje, že její teoretická výhoda rovnoměrnosti se rychle snižuje při aplikaci na náš prohledávaný prostor s osmi hyperparametry, který vyžaduje klíčovou zpětnou vazbu pro efektivní ladění.

Navrhovaný model dále prokázal výrazné zvýšení efektivity ve srovnání se silnou výchozí metodou GP. Využitím přenesených znalostí dosáhl model svého maximálního výkonu v menším počtu pokusů než GP u několika úloh, což potvrdilo výhodu „teplého startu“. Metoda také dosáhla podobných nebo lepších konečných výsledků, zejména na největších nebo strukturálně nejodlišnějších grafech (ArXiv, DBLP, Squirrel), a obecně vykazovala sníženou variabilitu napříč různými inicializacemi, což naznačuje stabilnější výběr vysoce kvalitních konfigurací.

Klíčová slova: Grafové neuronové sítě, Optimalizace hyperparametrů, Meta-učení, Sequential Model-Based Optimization

Full paper: M. Dědič, M. Bělohlávek, and M. Holeňa. *Benchmarking and Transfer Learning for Hyperparameter Optimization of Graph Neural Networks.* Unpublished manuscript, (2025).

WARD: Weather-Aware Road Surface Condition Monitoring Dataset

Soňa Drocárová
esnidalova@utia.cas.cz

study programme: Mathematical Engineering
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Adam Novozámský, Department of Image Processing
Institute of Information Theory and Automation, CAS

Abstract. Road surface condition (RSC) monitoring is essential for enhancing vehicle safety and accident prevention. This study investigates the application of computer vision techniques for real-time sensing of road surface conditions. We introduce a novel dataset named WARD (Weather-Aware Road Dataset), a comprehensive collection of almost 55 000 images collected in real-world driving scenarios across diverse seasonal and weather conditions, designed to advance RSC detection, now available for download. We thoroughly evaluate state-of-the-art computer vision models, specifically MobileNet and EfficientNet, on both the WARD and publicly available RoadSaW datasets, providing insights into their classification performance. MobileNet exhibited superior classification and inference speed results, processing images at up to 30 fps on an affordable GPU. To improve real-time efficiency, we employ temporal smoothing through moving window aggregation. Our findings validate the potential of non-contact, camera-based RSC monitoring, showcasing its practicality and cost-effectiveness compared to other sensors.

Keywords: road surface condition detection, non-contact detection, autonomous driving, computer vision, deep learning, convolutional neural network

Abstrakt. Monitorovanie stavu povrchu vozovky je kľúčové pre zvyšovanie bezpečnosti cestnej premávky a predchádzanie nehodám. Táto práca sa zaoberá aplikáciou techník počítačového videnia na snímanie stavu povrchu vozovky v reálnom čase. Predstavujeme nový dataset s názvom WARD (Weather-Aware Road Dataset), ktorý obsahuje takmer 55 000 snímok zozbieraných v reálnych podmienkach počas rôznych ročných období a za rôzneho počasia. Dataset bol vytvorený s cieľom posunúť vpred úlohu detekcie stavu povrchu vozovky a je dostupný pre verejnosť. Dôkladne hodnotíme moderné modely počítačového videnia, konkrétne MobileNet a EfficientNet, na datasete WARD ako aj na verejne dostupnom datasete RoadSaW, pričom prezentujeme ich klasifikačnú úspešnosť. MobileNet dosiahol lepšie výsledky v presnosti klasifikácie aj v rýchlosti predikcie, keď dokázal spracovávať snímky rýchlosťou až 30 fps na cenovo dostupnom GPU. Na zvýšenie efektívnosti v reálnom čase využívame časové vyhladzovanie prostredníctvom agregácie časovo pohyblivým oknom. Naše výsledky potvrdzujú potenciál bezkontaktného monitorovania stavu povrchu vozovky pomocou kamery a zdôrazňujú jeho praktickosť a nákladovú efektívnosť v porovnaní s alternatívnymi senzormi.

Kľúčové slová: detekcia stavu povrchu vozovky, bezkontaktná detekcia, autonómne riadenie, počítačové videnie, hlboké učenie, konvolučné neurónové siete

Full paper: WARD: Weather-Aware Road Surface Condition Monitoring Dataset. In 'Proceedings of the ICIAP 2025 – International Conference on Image Analysis and Processing, Rome, Italy, Springer'.

Two-Dimensional Schrödinger Operators with Non-Local Singular Potentials

Lukáš Heriban
heribluk@cvut.cz

studiní program: Matematické inženýrství
Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

školitel: Matěj Tušek, Katedra matematiky
Fakulta jaderná a fyzikálně inženýrská, ČVUT v Praze

Abstract. Consider a bounded subset of $\mathbb{R}^2$ denoted by Ω_+ with Lipschitz smooth boundary Σ. Our main goal is to investigate self-adjoint extensions of the free Schrödinger operator restricted to functions vanishing along the boundary Σ. Such an operator is symmetric with infinite deficiency indices, and thus the family of possible self-adjoint extensions is huge. So far, mathematicians mainly focused on extensions given by the formal expressions $-\Delta+\alpha\delta_\Sigma$ or $-\Delta+\alpha\delta'_\Sigma$, where δ_Σ stands for the single-layer distribution supported on Σ, known as δ-shell and δ'-shell interactions, respectively. We refer to such interactions as *local* ones.

We investigate new self-adjoint realizations of the Schrödinger operator in $L^2(\mathbb{R}^2)$, where a subclass of them can be formally written as $-\Delta+\alpha\,|\delta_\Sigma\rangle\langle\delta_\Sigma|$, and is strictly different from the local δ-shell interaction. Because of the nature of the formal projection, we call these *non-local δ-shell interactions*. This family of non-local interactions can be generalized to the class of self-adjoint Schrödinger operators with transmission conditions along the curve Σ with a compact self-adjoint operator

$$B : L^2(\Sigma;\mathbb{C}^2) \to L^2(\Sigma;\mathbb{C}^2)$$

as the transmission parameter capturing the nature of the interaction. These conditions make use of the Dirichlet traces of the functions belonging to the operator domains, together with the Dirichlet traces of their Wirtinger derivatives $\partial = (\partial_1 + i\partial_2)/2$.

Such operators can be formally written as

$$-\Delta+\sum_{n=1}^{+\infty} b_n\Big(|\varphi_n^1\delta_\Sigma\rangle\langle\varphi_n^1\delta_\Sigma| + |\varphi_n^1\delta_\Sigma\rangle\langle\partial(\varphi_n^2\delta_\Sigma)| + |\partial(\varphi_n^2\delta_\Sigma)\rangle\langle\varphi_n^1\delta_\Sigma| + |\partial(\varphi_n^2\delta_\Sigma)\rangle\langle\partial(\varphi_n^2\delta_\Sigma)|\Big),$$

where $\{(b_n,(\varphi_n^1,\varphi_n^2))\}_{n\in\mathbb{N}}$ are the eigenpairs of B with orthonormal eigenfunctions. Furthermore, using a boundary triple framework, we investigate several spectral properties and reveal how these models naturally connect to self-adjoint Dirac operators with non-local δ-shell interactions through the non-relativistic limit.

Keywords: Schrödinger operator, Non-local singular interactions, Extension theory, Boundary triple, Non-relativistic limit

Full paper: L. Heriban, M. Holzmann, C. Stelzer-Landauer, G. Stenzel, M. Tušek, *Two-dimensional Schrödinger operators with non-local singular potentials.* Journal of Mathematical Analysis and Applications 549 (2), 2025, 129498.

Abstrakt. Nechť $\Omega_+ \subset \mathbb{R}^2$ je omezená oblast s Lipschitzovsky hladkou hranicí Σ. Naším cílem je studium samosdružených rozšíření volného Schrödingerova operátoru, omezeného na funkce mizící na hranici Σ. Tento operátor je symetrický s nekonečnými indexy defektu, a tudíž množina jeho samosdružených rozšíření je obrovská. Dosavadní práce se soustředily zejména na samosdružená rozšíření formálně popsaná výrazy $-\Delta + \alpha\delta_\Sigma$ či $-\Delta + \alpha\delta'_\Sigma$, kde δ_Σ značí jednoduchou vrstvu na Σ, tj. δ-skořápkovou a δ'-skořápkovou interakci. Tyto interakce označujeme jako *lokální*.

V této práci zkoumáme nové samosdružené realizace Schrödingerova operátoru v prostoru $L^2(\mathbb{R}^2)$. Podtřída těchto operátorů má formální tvar $-\Delta + \alpha\,|\delta_\Sigma\rangle\langle\delta_\Sigma|$ a zásadně se liší od lokální δ-skořápkové interakce. Díky nelokální povaze formální poruchy tento typ operátoru nazýváme *nelokální δ-skořápková interakce*. Ty lze obecně popsat jako Schrödingerovy operátory s přechodovou podmínkou podél Σ, kde se v podmínce objevuje parametr v podobě kompaktního samosdruženého operátoru

$$B : L^2(\Sigma;\mathbb{C}^2) \to L^2(\Sigma;\mathbb{C}^2).$$

Podmínky využívají Dirichletovy stopy funkcí z definičních oborů operátorů a zároveň Dirichletovy stopy jejich Wirtingerových derivací $\partial = (\partial_1 + i\partial_2)/2$.

Formálně lze tyto operátory zapsat ve tvaru

$$-\Delta + \sum_{n=1}^{+\infty} b_n \Big(|\varphi_n^1\delta_\Sigma\rangle\langle\varphi_n^1\delta_\Sigma| + |\varphi_n^1\delta_\Sigma\rangle\langle\partial(\varphi_n^2\delta_\Sigma)| + |\partial(\varphi_n^2\delta_\Sigma)\rangle\langle\varphi_n^1\delta_\Sigma| + |\partial(\varphi_n^2\delta_\Sigma)\rangle\langle\partial(\varphi_n^2\delta_\Sigma)| \Big),$$

kde $\{(b_n, (\varphi_n^1, \varphi_n^2))\}_{n\in\mathbb{N}}$ jsou vlastní páry operátoru B s ortonormálními vlastními funkcemi. Pomocí teorie hraničních trojic dále zkoumáme spektrální vlastnosti těchto operátorů a ukazujeme, že se přirozeně pojí se samosdruženými Diracovými operátory s nelokálními δ-skořápkovými interakcemi v nerelativistické limitě.

Klíčová slova: Schrödingerův operátor, nelokální singulární interakce, teorie samosdružených rozšíření, hraniční trojice, nerelativistická limita

Plná verze: L. Heriban, M. Holzmann, C. Stelzer-Landauer, G. Stenzel, M. Tušek, *Two-dimensional Schrödinger operators with non-local singular potentials.* Journal of Mathematical Analysis and Applications 549 (2), 2025, 129498.

Marcus Cross Relation in the Space of H-atom Abstraction Reactions Boosted Through Off-Diagonal Thermodynamics*

Jan Kovář
kovarj29@fjfi.cvut.cz

study programme: Mathematical Engineering
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Radek Fučík, Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. In this paper, we revisit and extend the Marcus cross relation (MCR) [2] in the context of hydrogen-atom abstraction (HAA) reactions, which encompass both proton-coupled electron transfer (PCET) and hydrogen-atom transfer (HAT) processes [3]. These reactions are of fundamental importance in biological systems and organic synthesis, and their kinetics are often interpreted within the framework of the canonical MCR, originally developed to describe electron transfer. The relation links the rate constant of a mixed HAA reaction between two distinct reactants to the rate constants of two corresponding self-exchange reactions and to the thermodynamic driving force of the mixed process.

Using a dataset of 307 mixed HAA reactions and 109 associated self-exchange reactions, we conducted a systematic computational analysis of the validity and limitations of the canonical MCR. Our results demonstrate that, beyond the standard thermodynamic driving force, two additional parameters—termed off-diagonal thermodynamic factors—must be considered to achieve quantitative predictive accuracy. These factors, asynchronicity and frustration, arise from the imbalance and interplay of the electron- and proton-transfer components of the reaction [1]. Of these, asynchronicity is identified as the dominant contributor, while frustration provides a secondary refinement.

Incorporating off-diagonal thermodynamics significantly enhances the applicability of the MCR, with the most pronounced improvement observed in HAT reactions compared to PCET-like HAA processes. Furthermore, our model predicts the existence of a pseudoinverted region, in which more exergonic reactions paradoxically exhibit higher free-energy barriers, a phenomenon that cannot be explained by the original Marcus theory alone.

Thus, the present study refines the theoretical framework for understanding HAA kinetics, highlighting the necessity of extending the MCR with off-diagonal thermodynamic terms. These results underscore the importance of considering asynchronicity and frustration in mechanistic analyses of PCET and HAT processes and suggest future directions for integrating these effects into predictive models of reactivity.

Keywords: Marcus cross relation, Predictive models of reactivity, Hydrogen-atom abstraction, Off-diagonal thermodynamics

Abstrakt. V této práci rozšiřujeme Marcus cross relation (MCR) [2] v kontextu reakcí štěpení vodíkové vazby z anglického hydrogen-atom abstraction (HAA), které zahrnují jak proton-

*This work has been supported by the grant No. GA24-11247S

coupled electron transfer (PCET), tak hydrogen-atom transfer (HAT) procesy [3]. Tyto reakce hrají zásadní roli v biologických systémech i v organické syntéze a jejich kinetika je často interpretována v rámci MCR, původně vyvinuté pro popis přenosu elektronu. Tato relace spojuje rychlostní konstantu smíšené HAA reakce mezi dvěma různými reaktanty s rychlostními konstantami odpovídajících self-exchange[1] reakcí a s termodynamickou hnací silou smíšeného procesu.

Na základě souboru 307 smíšených HAA reakcí a 109 odpovídajících self-exchange reakcí jsme provedli systematickou výpočetní analýzu platnosti a omezení původní MCR. Naše výsledky ukazují, že kromě standardní termodynamické hnací síly je nutné zohlednit dva dodatečné parametry – tzv. mimodiagonální termodynamické faktory – aby bylo dosaženo kvantitativně přesné predikce. Tyto faktory, asynchronicita a frustrace, vycházejí z nerovnováhy a vzájemného působení komponent přenosu elektronu a protonu [1]. Z nich byla jako dominantní určena asynchronicita, zatímco frustrace poskytuje sekundární zpřesnění.

Začlenění mimodiagonální termodynamiky významně rozšiřuje použitelnost MCR, přičemž nejvýraznější zlepšení bylo pozorováno u HAT reakcí ve srovnání s PCET HAA procesy. Náš model dále předpovídá existenci tzv. pseudoinvertované oblasti, v níž více exergonické reakce paradoxně vykazují vyšší bariéry volné energie – jev, který nelze vysvětlit původní MCR teorií.

Tato studie tedy zpřesňuje teoretický rámec pro porozumění kinetice HAA a ukazuje nutnost rozšířit MCR o mimodiagonální termodynamické členy. Výsledky zdůrazňují význam zohlednění asynchronicity a frustrace v analýzách PCET a HAT procesů a naznačují budoucí směry pro integraci těchto efektů do prediktivních modelů reaktivity.

Klíčová slova: Marcus cross relation, Prediktivní modely reaktivity, Štěpení vodíkové vazby, Mimodiagonální termodynamika

Full paper: J. Kovář, E. Andris, Z. Wojdyla, J. S. Gopinath, J. Klinkovský, R. Fučík, M. Srnec. *Marcus cross relation in the space of H-atom abstraction reactions boosted through off-diagonal thermodynamics.* J Chem Phys 2025; 163(000000).
`https://doi.org/10.1063/5.028906`

References

[1] D. Bím, M. Maldonado-Domínguez, L. Rulíšek, and M. Srnec. *Beyond the classical thermodynamic contributions to hydrogen atom abstraction reactivity.* Proceedings of the National Academy of Sciences **115** (2018), E10287–E10294.

[2] R. A. Marcus and N. Sutin. *Electron transfers in chemistry and biology.* Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics **811** (1985), 265–322.

[3] D. Usharani, D. C. Lacy, A. Borovik, and S. Shaik. *Dichotomous hydrogen atom transfer vs proton-coupled electron transfer during activation of x–h bonds (x= c, n, o) by nonheme iron–oxo complexes of variable basicity.* Journal of the American Chemical Society **135** (2013), 17090–17104.

[1] Reakce, při které jsou oba reaktanty totožné.

Existence and Asymptotic Behaviour of Weakly Bounded States*

David Kramár
kramada1@fjfi.cvut.cz

study programme: Mathematical Engineering
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:

Dirk Hundertmark, Department of Mathematics
Karlsruhe Institute of Technology

Michal Jex, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstrakt. Představíme IMS lokalizační formuli, významného nástroje v teorii Schrödingerových operátorů a matematické fyzice. Formule umožňuje rozklad Laplaceova operátoru pomocí systémů hladkých cutoff funkcí a poskytuje tak efektivní nástroj pro studium spektrálních vlastností a lokalizace vlastních funkcí, zejména v situacích, kdy se vlastní hodnoty blíží prahu esenciálního spektra. Jsou popsány základní myšlenky a intuitivní význam IMS formule, dále její aplikace v analýze esenciálního spektra a při dokazování neexistence vázaných stavů. Jsou diskutovány možnosti zobecnění této metody na širší třídy operátorů, například magnetické Schrödingerovy operátory nebo Diracovy operátory.

Klíčová slova: Schrödingerův operátor, Spektrum, IMS lokalizační formule

Abstract. We present the IMS localization formula, an important tool in the theory of Schrödinger operators and mathematical physics. The formula enables the decomposition of the Laplacian by means of smooth cutoff functions, thus providing an effective tool for studying spectral properties and eigenfunction localization, particularly when eigenvalues approach the threshold of the essential spectrum. We explain the main ideas and the intuitive meaning of the IMS formula, discusses its applications in essential spectrum analysis and in proving the absence of bound states, and outlines possible extensions to broader classes of operators, such as magnetic Schrödinger operators and Dirac operators.

Keywords: Schrödinger operator, Spectrum, IMS localization formula

Introduction

One of the central role of mathematical physics in the quantum mechanics is to provide understanding of the spectral properties of Schrödinger-type Hamiltonians. These operators combine the Laplacian, which describes kinetic energy, with a potential term that governs interactions. A natural challenge arises: question of the existence and behavior of

*This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS25/165/OHK4/3T/14

the bound states, more precisely, spatial localization of the particles - quantum systems, in general. Mathematically speaking, the rate of decay of the corresponding eigenfunction is studied [4]. Although formulation of the problem is not complicated, its analysis, both qualitative and quantitative, can be challenging as the eigenvalues tend to the threshold of the essential spectrum. Indeed, standard approaches such as Agmon's method [1] rely on the safety distance from the essential spectrum. The *IMS formula* provides controllability without a necessary need of the aprior knowledge of the eigenvalue-essential-spectrum-treshold distance [3, 6]. It expresses the Laplacian applied to a state as a sum of localized contributions, plus a correction term that measures the error caused by localization.

The abbreviation *IMS* comes from the names Ismagilov–Morgan–Simon, three mathematicians who developed and popularized the localization identity. R. S. Ismagilov (1968) appears to have been the first to formulate a version of this decomposition for quadratic forms in his work on spectral theory [7]. About a decade later, J. D. Morgan and B. Simon used and promoted the formula in a systematic way in their investigations of Schrödinger operators. In particular, Simon's influential monographs and review articles highlighted the formula as an essential tool for understanding essential spectra and scattering [9].

Because of this double origin, the localization identity became universally known as the *IMS formula*. Today, it is regarded as a standard result in functional analysis and mathematical physics, with applications ranging from few-body quantum mechanics to spectral geometry.

The Formula

Let us introduce the formula in its simplest and most widely used setting. Consider Schrödinger operator

$$H = -\Delta + V(x) \qquad \text{on } L^2(\mathbb{R}^n),$$

where $V : \mathbb{R}^n \to \mathbb{R}$ is a multiplication function - potential, physically speaking. Operator H is understood as its self-adjoint realisation on domain $\mathrm{dom}(H)$. Now choose smooth cutoff functions $\chi_1, \ldots, \chi_N \in C^\infty(\mathbb{R}^n)$ such that they form a partition of unity in the sense

$$\sum_{j=1}^{N} \chi_j(x)^2 = 1 \quad \text{for all } x \in \mathbb{R}^n.$$

Then the *IMS localization formula* states that for every $\psi \in \mathrm{dom}(H)$,

$$\Re\,(\psi, H\psi) = \sum_{j=1}^{N} (\chi_j\psi, H\chi_j\psi) - \sum_{j=1}^{N} \left(\psi, |\nabla\chi_j|^2\psi\right).$$

If an eigenstate of H is chosen as ψ, the formula allows one to localize ψ into contributions supported in the spatial regions where cutoff functions χ_j are active. The price for that is paid in the form of negative corrections term involving $|\nabla\chi_j|^2$. This error depends only on the choice of the partition of unity, not on the potential or the state.

Intuitive Meaning

Physically, the IMS formula encodes the following idea: when we attempt to confine a quantum particle to smaller regions by multiplying the wave function with cutoff functions, we inject additional momentum because of the uncertainty principle. The term $-\sum_j (\psi, |\nabla\chi_j|^2\psi)$ measures precisely this "penalty" for localization.

Applications

Let us mention a few directions where IMS formula can be useful and find its utilization.

- **Essential spectrum analysis**
  In multi-particle quantum mechanics, one often wants to compare the spectrum of a Hamiltonian with that of its subsystems when particles are sent far apart. The IMS formula allows one to decompose the quadratic form into cluster contributions plus small error terms, providing a systematic way to prove HVZ-type theorems about the essential spectrum [6, 9].

- **Absence of bound states**
  In one-particle systems, the formula helps establish whether negative energy eigenvalues can exist. By localizing the Hamiltonian into regions where the potential is weak or decaying, one can derive contradictions that show the nonexistence of bound states under certain conditions [8].

Generalizations

Even though the IMS formula was first written for the Euclidean Laplacian, there are many directions where the extension of the principle can be possibly studied, we present the following

- **Magnetic Schrödinger operators**
  Although localization of eigenvalues of magnetic Schrödinger operators

  $$H = (-i\nabla - A(x))^2 + V(x)$$

  is studied [5], different tools [2] are used and we are not aware of the direct generalization of the IMS formula for these.

- **Dirac operators**
  It is natural to ask whether there is an analogy regarding the localization results in the relativistic quantum mechanics. To our knowledge, there is no universal extension of the IMS formula for the Dirac operators. The main reasons which make the problem much more delicate are non-positivity of the Dirac and the nontrivial matrix sctructure [4].

- **Abstract operator theory**
  More abstractly, the IMS identity is a manifestation of a general idea: quadratic forms of differential operators admit decompositions under partitions of unity, with remainder terms controlled by the gradients of the cutoffs. This makes it possibly applicable to a broad class of self-adjoint operators defined via quadratic forms.

References

[1] S. Agmon. *Lectures on exponential decay of solutions of second-order elliptic equations: bounds on eigenfunctions for N-body Schrödinger operatos*. Mathematical Notes **29** (1982).

[2] K. Broderix, D. Hundertmark, and H. Leschke. *Continuity properties of Schrödinger semigroups with magnetic fields*. Reviews in Mathematical Physics - RMP **12** (2000), 181–225.

[3] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon. *Schrödinger operators with application to quantum mechanics and global geometry*. Springer-Verlag, Berlin, (1987).

[4] M. Griesemer. *Exponential decay and ionization thresholds in non-relativistic quantum electrodynamics*. Journal of Functional Analysis **210** (2004), 321–340.

[5] J. G. Hoskins, H. Quan, and S. Steinerberger. *Magnetic schrödinger operators and landscape functions*. (2022).

[6] D. Hundertmark, M. Jex, and M. Lange. *Quantum systems at the brink: Helium-type systems*. (2021).

[7] R. S. Ismagilov. *On the spectrum of some self-adjoint operators*. Soviet Math. Dokl. **9** (1968), 1363–1366.

[8] J. D. Morgan. *Schrödinger operators whose potentials have separated singularities*. Journal of Operator Theory **1** (1979), 109–115.

[9] M. Reed and B. Simon. *Methods of Modern Mathematical Physics, Vol. IV: Analysis of Operators*. Academic Press, (1978).

Generalized Eddington–Finkelstein Coordinates and Exact Vaidya-Type Solutions in Weyl Conformal Gravity*

Tereza Lehečková
lehecter@cvut.cz

study programme: Mathematical Engineering
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Petr Jizba, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. We study Vaidya-type solutions in Weyl conformal gravity (WCG) using Eddington–Finkelsteinlike coordinates. Our considerations focus on spherical as well as hyperbolic and planar symmetries. In particular, we find all vacuum dynamical solutions for the aforementioned symmetries. These are, in contrast to general relativity, structurally quite non-trivial. We provide a thorough analysis of their basic properties, such as, relation to other known WCG solutions, algebraic types, singularities, horizons, and symmetries. In the same vein, we also derive, classify, and discuss non-vacuum solutions with the Coulombic electric field and null dust. Other salient issues, such as the gauge equivalence of WCG solutions to Einstein spaces and the role of the Birkhoff–Riegert theorem, are also addressed.

Keywords: Weyl conformal gravity, Vaidya solution, black holes, Eddington–Finkelstein coordinates, modified gravity theories

Abstrakt. Zabýváme se studiem řešení typu Vaidya ve Weylově konformní gravitaci (WKG) za použití zobecněných Eddington-Finkelsteinových souřadnic. Tato volba umožňuje zároveň studovat jak sféricky, tak planárně a hyperbolicky symetrická řešení. Konkrétně nalezneme všechna taková vakuová dynamická řešení uvedených symetrií, která jsou ve WKG, na rozdíl od řešení v rámci obecné relativity netriviální. Poskytneme též analýzu jejich základních vlastností, především vztahu k jiných známým WKG řešením, algebraických typů, singularit, horizontů a symetrií. Stejným způsobem též odvodíme, klasifikujeme a diskutujeme nevakuová řešení s Coulombickým polem a „null dust“. Adresujeme též další podstatné otázky jako je kalibrační ekvivalence WKG řešení a Einsteinovských prostorů, a role Birkhoff–Riegertova teorému.

Klíčová slova: Weylova konformní gravitace, Vaidya řešení, černé díry, Eddington-Finkelsteinovy souřadnice, modifikované teorie gravitace

Full paper: P. Jizba and T. Lehečková: *Generalized Eddington–Finkelstein Coordinates and Exact Vaidya-Type Solutions in Weyl Conformal Gravity*, The European Physical Journal C, volume 85, article number 602, (2025).

*This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS25/163/OHK4/3T/14.

Symmetric Poisson Geometry, Totally Geodesic Foliations and Jacobi-Jordan Algebras*

Filip Moučka
mouckfil@cvut.cz

study programme: Mathematical Engineering
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:

Roberto Rubio Núñez, Department of Mathematics
Universitat Autònoma de Barcelona

Branislav Jurčo, Mathematical Institute
Faculty of Mathematics and Physics, Charles University

Jan Vysoký, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. We introduce symmetric Poisson structures as pairs consisting of a symmetric bivector field and a torsion-free connection satisfying an integrability condition analogous to that in usual Poisson geometry. Equivalent conditions in Poisson geometry have inequivalent analogues in symmetric Poisson geometry and we distinguish between symmetric and strong symmetric Poisson structures. We prove that symmetric Poisson structures correspond to locally geodesically invariant distributions together with a characteristic metric, whereas strong symmetric Poisson structures correspond to totally geodesic foliations together with a leaf metric and a leaf connection. We introduce, using the Patterson-Walker metric, a dynamics on the cotangent bundle and show its connection to symmetric Poisson geometry, the parallel transport equation and the Newtonian equation for conservative systems. Finally, we prove that linear symmetric Poisson structures are in correspondence with Jacobi-Jordan algebras, whereas strong symmetric correspond to those that are moreover associative.

Keywords: symmetric bivector fields, torsion-free connections, totally geodesic foliations, Jacobi-Jordan algebras, symmetric Cartan calculus, the Patterson-Walker metric

Abstrakt. Zavádíme symetrické Poissonovy struktury jako páry skládající se ze symetrického bivektorového pole a beztorzní konexe, které splňují podmínku integrability analogickou té ve standardní Poissonově geometrii. Ekvivalentní podmínky v Poissonově geometrii mají neekvivalentní analogie v symetrické Poissonově geometrii a rozlišujeme mezi symetrickými a silně symetrickými Poissonovými strukturami. Symetrické Poissonovy struktury odpovídají lokálně geodeticky invariantním distribucím spolu s charakteristickou metrikou, naopak silně symetrické Poissonovy struktury odpovídají totálně geodetickým foliacím spolu s listovou metrikou a listovou konexí. Pomocí Patterson-Walkerovy metriky zavádíme dynamiku na kotečném bandlu

*This work has been supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS25/163/OHK4/3T/14 and by the Spanish State Research Agency under the grant PID2022-137667NA-I00.

a ukazujeme její souvislost se symetrickou Poissonovou geometrií, rovnicí paralelního přenosu a Newtonovou rovnicí pro konzervativní systémy. Nakonec dokazujeme, že lineární symetrické Poissonovy struktury odpovídají Jacobi-Jordanovým algebrám, zatímco silně symetrické odpovídají těm, které jsou navíc asociativní.

Klíčová slova: symetrická bivektorová pole, beztorzní konexe, totálně geodetické foliace, Jacobi-Jordanovy algebry, symmetrický Cartanův kalkulus, Patterson-Walkerova metrika

Summary of the paper

Riemannian, or more generally (pseudo-)Riemannian, and symplectic geometry are among the most established areas in mathematics. They describe geometric structures: metrics and symplectic forms. These are, respectively, symmetric and skew-symmetric 2-forms with a non-degeneracy condition (and, in the case of symplectic structures, the extra integrability condition of being closed).

Symplectic geometry degenerates, or is extended, in two ways. First, by allowing the closed 2-form to be degenerate, which gives rise to a presymplectic structure. Second, by considering the inverse of a symplectic structure, that is, a skew-symmetric bivector field, and allowing it to be degenerate. The latter, together with an integrability condition generalizing closedness, is the starting point of Poisson geometry, a much richer theory than presymplectic geometry.

For a metric, its degeneration results in the study of symmetric bilinear forms with non-trivial kernel, which are well known. However, we have not found a study of degenerations of metrics as symmetric bivectors, that is, of what should be called symmetric Poisson geometry.

The aim of this paper is to overtake this study of symmetric bivector fields by bringing in the natural analogy with Poisson geometry, that is, establishing an integrability condition and describing its geometric and dynamical significance.

Using an analogue of the most common formulations of the integrability condition in Poisson geometry results in a void condition or the vanishing of the symmetric bivector field. We resort to symmetric Cartan calculus [2] to find suitable integrability conditions. There are two key points that deserve special mention.

First, symmetric Cartan calculus depends upon the choice of a connection on M. This is actually very natural if we recall that non-degenerate bivector fields are equivalent to (pseudo-)Riemannian metrics, whose features are two-fold:

metric-related	lengths and angles, geodesics as length minimizing curves, gradients, volume and integration, Hodge star
connection-related	curvature, parallel transport and geodesics, totally geodesic submanifolds, differentiation of tensor fields

This split is not apparent because of the existence and uniqueness of the Levi-Civita connection. However, when a (pseudo-)Riemannian metric degenerates as a symmetric bivector field, there is no Levi-Civita connection and we have to incorporate a compatible connection into the definition. The choice of a connection, which we can assume to be torsion-free, determines the symmetric derivative operator, the symmetric Lie derivative

and the symmetric bracket on vector fields, $[\ ,\ ]_s$, which we extend to the symmetric Schouten bracket on symmetric multivector fields $\mathfrak{X}^\bullet_{\mathrm{sym}}(M)$.

Second, for a pair (ϑ, ∇) consisting of $\vartheta \in \mathfrak{X}^2_{\mathrm{sym}}(M)$ and a torsion-free connection ∇, the integrability conditions that mirror the most common ones in Poisson geometry are

$$[\vartheta, \vartheta]_s = 0 \qquad \text{and} \qquad X_{\{f,\mathrm{g}\}} = [X_f, X_\mathrm{g}]_s$$

for $f, \mathrm{g} \in \mathcal{C}^\infty(M)$, where $\{f, \mathrm{g}\} := \vartheta(\mathrm{d}f, \mathrm{dg})$, $X_f := \vartheta(\mathrm{d}f)$, which is the gradient of f in this context, and $[\ ,\ ]_s$ also denotes the symmetric Schouten associated to ∇. Unlike in the usual Poisson case, these two conditions are not equivalent. The latter implies the former, thus we arrive at the following definition:

Definition. A **symmetric Poisson structure** is a pair (ϑ, ∇) such that $[\vartheta, \vartheta]_s = 0$. If, moreover, $X_{\{f,\mathrm{g}\}} = [X_f, X_\mathrm{g}]_s$, we call it a strong symmetric Poisson structure.

A good source of recognizable examples is provided by that non-degenerate symmetric Poisson structures correspond to the inverses of Killing tensors (see [2] and the references therein) together with the torsion-free connection they are Killing for, whereas non-degenerate strong symmetric Poisson structures are inverses of pseudo-Riemannian metric together with their Levi-Civita connection.

The next step is the geometric interpretation of the integrability conditions for a pair (ϑ, ∇). When regarded as a map $\vartheta : T^*M \to TM$, we have a distribution $\operatorname{im}\vartheta \subseteq TM$ and a $\mathcal{C}^\infty(M)$-module $\mathcal{F}_\vartheta := \vartheta(\Omega^1(M)) \subseteq \mathfrak{X}(M)$. In general, the distribution and module do not determine a partition of the manifold, as the condition $[\vartheta, \vartheta]_s = 0$ does not even imply Lie involutivity, but only that the characteristic module is preserved by the symmetric bracket, that is, $[\mathcal{F}_\vartheta, \mathcal{F}_\vartheta]_s \subseteq \mathcal{F}_\vartheta$. On the other hand, the stronger condition $X_{\{f,\mathrm{g}\}} = [X_f, X_\mathrm{g}]_s$ does imply the Lie involutivity of $\mathcal{F}_\vartheta$. This motivates the introduction of an a priori, intermediate class, involutive symmetric Poisson structures, those (ϑ, ∇) such that $\mathcal{F}_\vartheta$ is involutive for the Lie bracket. We thus have:

$$\left\{ \begin{array}{c} \text{strong symmetric} \\ \text{Poisson structures} \end{array} \right\} \subseteq \left\{ \begin{array}{c} \text{involutive symmetric} \\ \text{Poisson structures} \end{array} \right\} \subseteq \left\{ \begin{array}{c} \text{symmetric} \\ \text{Poisson structures} \end{array} \right\}.$$

The main questions are what is the geometric interpretation of these three classes of structures and whether these are different classes.

The distribution $\operatorname{im}\vartheta$ comes equipped, at each $m \in M$, with the canonical linear (pseudo-)Riemannian metric g_{ϑ_m} given, for $\alpha, \beta \in T^*_m M$, by

$$g_{\vartheta_m}(\vartheta(\alpha), \vartheta(\beta)) := \vartheta(\alpha, \beta).$$

For ∇-geodesics that are ϑ-admissible (those that admit a curve $a : I \to T^*M$ such that $\vartheta(a)$ is the velocity of the geodesic), we prove that the square of the speed, with respect to g_{ϑ_m}, is constant, thus extending a well-known fact for (pseudo-)Riemannian manifolds.

When (ϑ, ∇) is moreover involutive, we can describe what happens on a leaf N: the metrics g_{ϑ_m} extend to a (pseudo-)Riemannian metric g_N, and the connection ∇ restricts to a torsion-free connection ∇^N.

This leaf-wise structure allows us to prove our first main result, which should be compared with the symplectic foliation of usual Poisson geometry.

Theorem 1. *The characteristic partition of an involutive symmetric Poisson structure (ϑ, ∇) is totally geodesic. Moreover, on any leaf N, the pair (g_N^{-1}, ∇^N) is a non-degenerate symmetric Poisson structure. In addition, if (ϑ, ∇) is strong, ∇^N is the Levi-Civita connection of g_N, that is, (g_N^{-1}, ∇^N) is also strong.*

Poisson geometry originates from the study of Hamiltonian dynamics formulated in terms of the canonical Poisson bracket on $\mathcal{C}^\infty(T^*M)$, which comes from the inverse of the canonical symplectic form. The symmetric analogue of this symplectic form is the so-called Patterson-Walker metric g_∇, first introduced in [3], which depends upon the choice of a connection ∇ and can be understood in the framework of symmetric Cartan calculus, [2, Sec. 5]. By considering the inverse of g_∇ we obtain the symmetric Poisson bracket $\{\,,\}_\nabla$ on $\mathcal{C}^\infty(T^*M)$, which allows us to introduce an analogue of Hamiltonian dynamics, which we call Patterson-Walker dynamics. For a given $H \in \mathcal{C}^\infty(T^*M)$ we study the integral curves of $\{H, \}_\nabla$. By making different choices, we recover the parallel transport equation, the gradient extension of a dynamical system and the Newtonian equation for conservative systems. The study of this dynamics comes with two unexpected consequences. First, a relation between the symmetric Schouten bracket and the Patterson-Walker metric.

Proposition. *Let ∇ be a torsion-free connection on M. The vertical lift of a symmetric multivector field is an algebra isomorphism between the commutative algebras $(\mathfrak{X}^\bullet_{\mathrm{sym}}(M), [\,,\,]_s)$ and $(\mathcal{P}ol(T^*M), \{\,,\}_\nabla)$.*

Second, the geometrical description of symmetric Poisson structures as locally geodesically invariant distributions, which we introduce as those for which every geodesic whose velocity is on the distribution at a point, must have its velocity on the distribution in a neighbourhood. Our second main result is:

Theorem 2. *The characteristic distribution of a symmetric Poisson structure is locally geodesically invariant.*

Having this clear geometric interpretation in Theorems 1 and 2 is the most solid justification for our definition of (strong) symmetric Poisson structures formulated in terms of symmetric Cartan calculus.

We finish the paper by looking at examples of symmetric Poisson structures, including squares of autoparallel vector fields, as well as left-invariant symmetric Poisson structures. The most interesting class is that of linear symmetric Poisson structures on the dual of a vector space. By considering the linear function $\iota_u \in \mathcal{C}^\infty(V^*)$ for $u \in V$, we prove our third and last main result.

Theorem. *The assignment $\iota_{u\cdot v} := \{\iota_u, \iota_v\}$ gives a bijection*

$$\left\{ \begin{array}{c} \textit{linear (strong) symmetric Poisson} \\ \textit{structures } (\{\,,\}, \nabla^{\mathrm{Euc}}) \textit{ on } V^* \end{array} \right\} \overset{\sim}{\longleftrightarrow} \left\{ \begin{array}{c} \textit{(associative) Jacobi-Jordan} \\ \textit{algebra structures } \cdot \textit{ on } V \end{array} \right\}.$$

This is the analogue of the fact that linear Poisson structures on a vector space V^* correspond to Lie algebra structures on V. It establishes a bridge to Jacobi-Jordan algebras [1], which allows us to bring classification results in low dimensions and translate

geometric invariants into algebraic ones. In particular, we bring a normal form for non-strong linear symmetric Poisson structures in dimension 5 and find the first example of an involutive symmetric Poisson structure that is non-strong for any choice of connection. This also shows that symmetric, involutive symmetric and strong symmetric Poisson structures are indeed different classes.

In this work, we have laid the foundations of symmetric Poisson geometry from the Poisson viewpoint. A natural continuation is the study of symmetric Poisson maps, product structures, submanifolds and cohomology. However, there is an intriguing (pseudo-)Riemannian side too, where natural notions such as the *symmetric Poisson scalar curvature* $\mathcal{R}_{(\vartheta,\nabla)} := \mathrm{tr}(\vartheta \otimes \mathrm{Ric}_\nabla)$ and the *symmetric Poisson Laplacian* $\Delta_{(\vartheta,\nabla)} f := \mathrm{tr}(\vartheta \otimes \nabla \mathrm{d}f)$ can be defined, or a connection with sub-Riemannian geometry can be established. These directions will be explored in future work.

Full paper: F. Moučka, R. Rubio. Symmetric Poisson geometry, totally geodesic foliations and Jacobi-Jordan algebras. [arXiv:2508.15890], 2025.
Currently submitted for publication, undergoing peer review.

References

[1] Dietrich Burde and Alice Fialowski. Jacobi-Jordan algebras. *Linear Algebra Appl.*, 459:586–594, 2014.

[2] F. Moučka, R. Rubio. Symmetric Cartan calculus, the Patterson-Walker metric and symmetric cohomology. [arXiv:2501.12442], 2025.
Currently submitted for publication, undergoing peer review.

[3] E. M. Patterson and A. G. Walker. Riemann extensions. *Quart. J. Math. Oxford Ser.* (2), 3:19–28, 1952.

Applications of Constrained Curvature Flow in Plane*

Maneesh Narayanan
narayman@cvut.cz

study programme: Mathematical Engineering
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Michal Beneš, Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. This article studies the planar non-local curvature flow for closed Jordan curves, relevant to problems in natural sciences. The flow evolves by

$$V = -\kappa_\Gamma + I + \frac{\int_{\Gamma_t}(\kappa_\Gamma - I), d\Gamma}{\int_{\Gamma_t} d\Gamma}, \tag{1}$$

$$\Gamma_t\Big|_{t=0} = \Gamma_0, \tag{2}$$

where Γ_t is a smooth simple closed curve, V the normal velocity, $\mathbf{n}$ the outward unit normal, κ_Γ the curvature, and $F = F(t, x)$ the normal force. The motion law (1-2) and its applications are discussed in [2, 3, 5]. The motion law is treated via the **parametric method**, reformulated as a system of **degenerate parabolic partial differential equations**. Numerical solutions are obtained using the **finite volume method**, and solved numerically by **Runge–Kutta–Merson** scheme as in [2]. We examine the deformation of an initially circular curve under two representative force scenarios:

1. **Droplet falling under gravity:** where the force $I = I(t, \theta)$ is provided by the experimental data available in literature.
2. **Circular eukaryotic cell under time-dependent force:** where the term $I = I(t, \theta)$ mimics mechanical forces acting on the cell membrane.

Our simulations of droplet motion and eukaryotic cell deformation provide quantitative insights into **constrained geometric flows** and their implications for physical and biological systems.

Keywords: Curve evolution; Parametric method; Method of lines; Constrained geometric flows

Abstrakt. Tento článek se zabývá rovinným nelokálním pohybem uzavrenych Jordanovych krivek uzavřené Jordanovy křivky, který lze pozorovat v problematice přírodních věd. Pohyb je dán zákonem:

$$V = -\kappa_\Gamma + I + \frac{\int_{\Gamma_t}(\kappa_\Gamma - I), d\Gamma}{\int_{\Gamma_t} d\Gamma}, \tag{3}$$

*The research was partly supported by the project "Modeling, prediction, and control of processes in nature, industry, and medicine powered by high performance computing," No. SGS23/188/OHK4/3T/14 of the Student Grant Agency of the Czech Technical University in Prague.

$$\Gamma_t\Big|_{t=0} = \Gamma_0, \tag{4}$$

kde Γ_t je hladká jednoduchá uzavřená křivka, V je normálová rychlost, $\mathbf{n}$ je vnější jednotkový normálový vektor, κ_Γ je křivost a $F = F(t, x)$ je normálová síla. Pohybový zákon (3-4) a jeho aplikace jsou diskutovány v [2, 3, 5]. Pohybový zákon je řešen pomocí **parametrické metody**, přeformulované jako systém **degenerovaných parabolických parciálních diferenciálních rovnic**. Numerická řešení jsou získána **metodou konečných objemů** a numericky řešena **schématem Runge–Kutta–Merson** jako v [2]. Zkoumáme deformaci počátečně kruhové křivky ve dvou scénářích:

1. **Kapka padající vlivem gravitace:** kde síla $I(t, \theta)$ je daná experimentálními daty dostupnými v literatuře.
2. **Kruhová eukaryotická buňka pod vlivem síly závislé na čase:** kde člen $I(t, \theta)$ napodobuje mechanické síly působící na buněčnou membránu.

Naše simulace pohybu kapky a deformace eukaryotické buňky poskytují kvantitativní vhled do **omezených geometrických toků** a jejich důsledků pro fyzikální a biologické systémy.

Klíčová slova: Evoluce křivek; parametrická metoda; metoda přímek; vázané geometrické toky

Full paper: This contribution is based on the article [1] and the oral presentation in [4].

References

[1] M. Beneš and M. Narayanan. Evolution of space curves by parametric method with natural and uniform redistribution. In *ALGORITMY 2024, 22th Conference on Scientific Computing, High Tatra Mountains, Slovakia, March 15-20, 2024, Proceedings of contributed papers and posters*, 2024.

[2] M. Kolář, M. Beneš, and D. Ševčovič. Computational studies of conserved mean-curvature flow. *Math. Bohem.*, 139(4):677–684, 2014.

[3] M. Kolář, M. Beneš, and D. Ševčovič. Computational analysis of the conserved curvature driven flow for open curves in the plane. *Math. Comput. Simul.*, 126:1–13, 2016.

[4] M. Narayanan. Applications of constrained curvature flow in plane. Oral presentation at ACOMEN 2025, Ghent, Belgium. September 15–19, 2025.

[5] P. Pauš and M. Beneš. Direct approach to mean-curvature flow with topological changes. *Kybernetika*, 45:591–604, 2009.

Fast Texture-Based Analysis and Classification of 3D MRI Brain Scans*

Matěj Pokorný
pokorm20@fjfi.cvut.cz

study programme: Mathematical Engineering
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jaromír Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. Texture-based analysis of medical images is an increasingly important diagnostic tool. This paper presents a method for the construction and application of two- and three-dimensional Zernike polynomial-based translational-rotational-mirroring (TRM) image texture invariants, derived via the Fourier transform, for image texture analysis and classification tasks. We propose an approach for computing global characteristics from these invariants, enabling fast and efficient feature extraction. The method is evaluated on a binary classification task that distinguishes between cognitively normal (CN)individuals and patients with Alzheimer's Disease (AD), using axial MRI brain scans from the ADNI database. Multiple classifiers - LDA, QDA, k-NN, SVM, and ANN - are tested, with the best performance achieved using 3D Zernike invariants and a QDA classifier with regularization, reaching an accuracy of ACC = 0.925. The extension from 2D central slices and their invariants to 3D invariants of full brain scans yields a notable improvement in both accuracy and Matthews correlation coefficient. The presented method is computationally efficient and achieves results comparable to more complex state-of-the-art techniques. Moreover, its generality and speed make it applicable to a wide range of 2D and 3D image classification problems beyond biomedical imaging.

Keywords: Fourier Transform, Alzheimer's Disease, Zernike Polynomial, Magnetic Resonance Imaging, Texture Invariants, Texture Analysis, Image Classification

Abstrakt. Texturní analýza snímků v medicíně je diagnostický nástroj s rostoucím významem. Tento článek představuje metodu konstrukce dvou a tří dimenzionálních translačních, rotačních, zrcadlových (TRM) invariantů založených na Zernikeho polynomech, získaných pomocí Fourierovy transformace, a jejich využití pro analýzu textur a klasifikaci obrazů. Navrhujeme přístup pro výpočet globálních charakteristik těchto invariantů, což umožňuje rychlé a efektivní získávání vlastností obrazu. Metoda byla experimentálně otestována na binárním klasifikačním problému rozlišení kognitivně normálních (CN) pacientů od pacientů s Alzheimerovou chorobou (AD) s použitím axiálních snímků magnetické rezonance (MRI) z databáze ADNI. Bylo otestováno více klasifikátorů - LDA, QDA, k-NN, SVM a ANN - přičemž nejlepších výsledků, s přesností ACC = 0.925, bylo dosaženo s 3D Zernikeho invarianty a klasifikátorem QDA s využitím regularizace. Rozšířením z 2D centrálních řezů a jejich invarintů na 3D invarianty celých snímků mozku dochází k významnému zlepšení přesnosti a Matthewsova korelačního koeficientu klasifikace. Prezentovaná metoda je výpočetně efektivní a dosahuje výsledků, jež jsou srovnatelné s nejmodernějšími a komplexnějšími technikami. Dále, rychlost a obecnost této metody ji

*This work has been supported by the grant SGS23/190/OHK4/3T/14 of Czech Technical University in Prague.

umožňují aplikovat na široké spektrum klasifikačních problémů pro 2D a 3D obrazy mimo oblast snímků v biomedicíně.

Klíčová slova: Fourierova transformace, Alzheimerova choroba, Zernikeho polynom, magnetická rezonance, texturní invarianty, analýza textur, klasifikace obrazu

Full paper: M. Pokorný, J. Kukal. *Fast Texture-Based Analysis and Classification of 3D MRI Brain Scans.* In peer review process in Biomedical Signal Processing and Control.

Non-Self-Adjoint Dirac Operators on Graphs

Václav Růžek
ruzekva2@fjfi.cvut.cz

study programme: Mathematical Engineering
Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Matěj Tušek, Department of Mathematics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. Quantum graphs are a popular and widely studied model in mathematical physics. They typically concern the Laplace operator acting on functions defined on the edges of a metric graph. In our work, we focus on a relativistic operator of energy: the Dirac operator. Unlike the non-relativistic case, the Dirac operator is a first-order differential operator with matrix structure.

To properly define various realisations of the Dirac operator, it is necessary to specify their domain, typically by prescribing boundary or transmission conditions. We begin with the most restrictive conditions and then study all possible intermediate extensions of this symmetric *minimal* operator. Our analysis encompasses not only self-adjoint extensions but also more general non-self-adjoint operators. The central framework enabling an elegant description of these extensions is the theory of boundary triples.

We first introduce the operator on a line segment and a half-line, which serve as building blocks for graphs – the former for internal edges, the latter for external ones. The Hilbert space, Dirac operator, and boundary triple for the full graph are constructed as direct sums over these respective edgewise objects. While this direct sum is taken over edges, we also define a vertex space, decomposed as a direct sum over the vertices, to precisely capture the concept of *local* transmission conditions. These conditions, at each vertex, depend on the values (traces) of functions only at that vertex, not at others.

After extensive preparatory work, we adapt known general results concerning self-adjointness, spectral criteria, and resolvent formulas for the Dirac operator. Notably, we present – in what we believe to be the first instance – a proof of these criteria in the general non-self-adjoint setting for our chosen parametrisation of Dirac operators. As a demonstration, the spectral criterion is applied to an example of a star-shaped graph.

The final section is devoted to physically motivated transformations of the Hilbert space and their capacity to be symmetries of the Dirac operator. Specifically, we discuss a change of orientation (analogous to a parity transformation), which does not affect the spectrum; a form of time reversal, characterised as anti-linear and involutive, which reflects the spectrum with respect to the real axis; and charge conjugation, also anti-linear and involutive, which reflects the spectrum with respect to the imaginary axis. For each transformation, we provide equivalent conditions for the operator to be symmetric with respect to these transformations, formulated in terms of the properties of vertex transmission conditions.

Keywords: Dirac operator, non-self-adjoint operator, relativistic quantum graph

Abstrakt. Kvantové grafy jsou populárním a široce studovaným modelem matematické fyziky. Typicky se zabývají Laplaceovým operátorem působícím na funkce definované na hranách

metrického grafu. V našem článku se zaměřujeme ale na relativistický operátor energie: Diracův operátor. Na rozdíl od nerelativistického případu Diracův operátor je diferenciální operátor prvního řádu s maticovou strukturou.

Abychom korektně zavedli všeliké realizace Diracova operátoru je nutné specifikovat jejich definiční obor, typicky se tak děje předepsáním hraničních či přechodových podmínek. Začínáme od nejrestriktivnějších a posléze studujeme všechna mezilehlá rozšíření tohoto symetrického *minimálního* operátoru. Naše analýza se neomezuje pouze na samosdružená rozšíření, ale též uvažuje obecně nesamosdružená. Hlavním nástrojem umožňující elegantní popis těchto rozšíření je teorie „boundary triple".

Nejdříve zavedeme operátor na úsečce a polopřímce, které slouží jako stavební kameny grafu – úsečka odpovídá konečným hranám, polopřímka nekonečným. Hilbertův prostor, Diracův operátor a „boundary triple" na celém grafu jsou sestaveny jako direktní sumy svých komponent na hranách. Navíc definujeme „vertex space", který má strukturu direktní sumy přes vrcholy, sloužícího k přesnému zachycení pojmu *lokálních* přechodových podmínek. Tyto podmínky jsou takové, které v každém vrcholu závisí na hodnotách (stopách) funkcí pouze v tomto vrcholu a ne v žádném jiném.

Po rozsáhlé přípravě upravíme známé obecné výsledky o samosdruženosti, o spektrálních kritériích a o resolventní formuli pro případ Diracova operátoru. Zejména zde uvádíme – zdá se, že jako první – důkaz těchto spektrálních kritérií pro obecnou nesamosdruženou situaci námi zvolené parametrizace Diracova operátoru. Použití ukážeme na příkladě vyšetření spektra hvězdicovitého grafu.

Závěrečná sekce je věnována fyzikálně motivovaným transformacím Hilbertova prostoru a jejich schopnosti být symetriemi Diracova operátoru. Konkrétně se zabýváme změnou orientace (jako analogem inverze prostorových souřadnic), kterážto spektrum neovlivňuje; druhem časové inverze, která je antilineární a involutivní a která zrcadlí spektrum vzhledem k reálné ose; a nábojové konjugaci, též antilineární a involutivní, kterážto zrcadlí spektrum vzhledem k imaginární ose. Ke každé transformaci formulujeme ekvivalentní podmínku, kdy je operátor k transformaci symetrický, ve formě vlastností přechodových podmínek ve vrcholech.

Klíčová slova: Diracův operátor, nesamosdružený operátor, relativistický kvantový graf

Full paper: M. Holzmann, V. Růžek, and M. Tušek. *Non-self-adjoint Dirac operators on graphs.* Journal of Physics A: Mathematical and Theoretical **58** (apr 2025), 165205.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this abstract the author used Perplexity.ai in order to eliminate grammatical mistakes and avoid cumbersome formulations. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Threefold Nature of Graded Vector Bundles

Rudolf Šmolka*
smolkrud@cvut.cz

study programme: Mathematical Engineering
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jan Vysoký, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. Vector bundles form a backbone of modern differential geometry. The notion formalizes the idea of gluing a vector space to each point of a smooth manifold, allowing to describe vector space valued quantities on manifolds, such as vector fields, differential forms, differential operators, densities, etc.

Graded vector bundles over a given $\mathbb{Z}$-graded manifold can be defined in three different ways: certain sheaves of graded modules over the structure sheaf of the base graded manifold, finitely generated projective graded modules over the algebra of global functions on the base graded manifold, or locally trivial graded manifolds with a suitable linear structure. After a brief introduction to $\mathbb{Z}$-graded manifold theory we recall the first approach, which was examined in previous works. We then introduce the latter two approaches, which have not yet been fully treated in the literature in the context of $\mathbb{Z}$-graded manifolds. All three approaches are formalized as suitable categories, and all three categories are shown to be equivalent.

Keywords: Graded vector bundles, Graded manifolds, Graded modules, Serre–Swan theorem

Abstrakt. Vektorové fibrované prostory tvoří zásadní část moderní diferenciální geometrie. Formalizují koncept lepení vektorového prostoru ke každému bodu hladké variety a umožňují tak popis hladkých funkcí s hodnotami ve vektorových prostorech: vektorová pole, diferenciální formy, diferenciální operátory, hustoty, a podobné.

Gradovaný vektorový fibrovaný prostor nad $\mathbb{Z}$-gradovanou varietou může být definován třemi způsoby: jako snop gradovaných modulů nad strukturálním snopem bazické gradované variety, jako konečně generovaný projektivní gradovaný modul nad algebrou globálních funkcí bazické gradované variety, nebo jako lokálně triviální gradovaná varieta s vhodnou lineární strukturou. Po krátkém úvodu do světa $\mathbb{Z}$-gradovaných variet připomeneme první z přístupů, který byl již vyšetřen v jiných pracích. Dále zavedeme zmíněné další dva přístupy, které nebyly, alespoň v kategorii $\mathbb{Z}$-gradovaných variet, v literatuře doposud zpracovány. Všechny tři přístupy jsou formalizovány v podobě příslušné kategorie, a tyto kategorie se ukáží býti ekvivalentní.

Klíčová slova: Gradované vektorové fibrované prostory, gradované variety, gradované moduly, Serre-Swanova věta.

Full paper: R. Smolka and J. Vysoky. *Threefold nature of graded vector bundles.* J. Geom. Phys, Volume 216, 2025, 105557. DOI: 10.1016/j.geomphys.2025.105557.

*The author is grateful for the support of the grant GAČR 24-10031K and also for the support of the Grant Agency of the Czech Technical University in Prague, grant No. SGS25/163/OHK4/3T/14.

Central Splitting of C_2 Weight Lattice Transforms*

Vojtěch Teska
vojtech.teska@fjfi.cvut.cz

study programme: Mathematical Engineering
Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisors:

Jiří Hrivnák, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Lenka Motlochová, Department of Physics
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. The simple Lie algebra of type C_2 is characterized by its irreducible root system which determines its Weyl group. The corresponding weight and root lattices are invariant with respect to the Weyl group. The weight lattice is disjointly decomposed into cosets of the root lattice for which there exists a non-trivial admissible shift. Fragments of these (shifted) lattices located inside fundamental domains of generalized affine Weyl groups are used to construct label and point sets for discrete Fourier-like transforms of C–functions. Transform matrices are introduced and central splitting is described, relating the matrices and simplifying calculations.

Keywords: Lie algebra, root system, Weyl group, invariant lattice, orbit function, central splitting

Abstrakt. Prostá Lieova algebra typu C_2 je charakterizována svým ireducibilním kořenovým systémem, který určuje její Weylovu grupu. Odpovídající váhová a kořenová mříž jsou invariantní vzhledem k této Weylově grupě. Váhová mříž je disjunktně rozložená do levých tříd kořenové mříže, pro kterou existuje netriviální přípustné posunutí. Části těchto (posunutých) mříží uvnitř fundamentálních oblastí zobecněných afinních Weylových grup jsou využity ke konstrukci množin vah a bodů pro diskrétní Fourierovy transformace C–funkcí. Jsou zavedeny matice těchto transformací a centrální rozklad, který popisuje vzájemný vztah matic a zjednodušuje výpočty.

Klíčová slova: Lieova algebra, kořenový systém, Weylova grupa, invariantní mříž, orbitová funkce, centrální rozklad

1 Introduction

This article describes the so-called central splitting mechanism for C–functions in the case of the Lie algebra C_2. The first two sections describe the necessary preliminaries, such as the explicit form of the root system and the associated lattices, relations between them and fundamental domains of generalized affine Weyl groups, as presented in [1, 2] and

*This work was supported by the student grant of the Grant Agency of the Czech Technical University in Prague, grant number SGS25/163/OHK4/3T/14.

[3]. The following section gives the definition and overview of properties of C–functions [4], followed by an explicit description of the needed label and point sets.

The next section is dedicated to discrete orthogonality properties of C–functions and their utilization for the definition of corresponding transform matrices which were introduced in [5]. The splitting matrices are defined in the final section, followed by the main result, which is analogous to the case of the root system A_2, examined in [6], followed by an explicit example. The conclusion mentions other cases of transform decompositions and discusses possible future areas of research.

2 The root system C_2 and associated lattices

Consider the Euclidean space $\mathbb{R}^2$ endowed with inner product $\langle \cdot, \cdot \rangle$. The crystallographic root system Π of type C_2 is formed by the action of the Weyl group W on the set $\Delta = \{\alpha_1, \alpha_2\}$ of simple roots. The simple roots are determined (up to isomorphism) by their angles and lengths,

$$\measuredangle(\alpha_1, \alpha_2) = \tfrac{3}{4}\pi, \quad ||\alpha_1|| = 1, \quad ||\alpha_2|| = \sqrt{2}. \tag{1}$$

The coroot of $\alpha \in \Pi$ is defined as

$$\alpha^\vee = \frac{2\alpha}{\langle \alpha, \alpha \rangle}. \tag{2}$$

The simple reflections $r_{1,2}$ associated with the simple roots $\alpha_{1,2}$ are determined by the equivalent formulas

$$r_i v = v - \langle v, \alpha_i^\vee \rangle \, \alpha_i = v - \langle v, \alpha_i \rangle \, \alpha_i^\vee, \quad \forall v \in \mathbb{R}^2, i = 1, 2. \tag{3}$$

The fundamental weights $\omega_{1,2}$ are defined by the duality relations

$$\left\langle \omega_i, \alpha_j^\vee \right\rangle = \delta_{ij}, \quad i, j = 1, 2 \tag{4}$$

and the fundamental coweights $\omega_{1,2}^\vee$ are determined by

$$\langle \omega_i^\vee, \alpha_j \rangle = \delta_{ij}, \quad i, j = 1, 2. \tag{5}$$

The simple roots, simple coroots, fundamental weights and fundamental coweights each form a basis of $\mathbb{R}^2$. In terms of the fundamental weights ω_1, ω_2 using the definition of coroots (2), fundamental weights (4) and fundamental coweights (5), we obtain:

$$\begin{array}{lll} \alpha_1 = 2\omega_1 - \omega_2, & \alpha_1^\vee = 4\omega_1 - 2\omega_2, & \omega_1^\vee = 2\omega_1, \\ \alpha_2 = -2\omega_1 + 2\omega_2, & \alpha_2^\vee = -2\omega_1 + 2\omega_2, & \omega_2^\vee = \omega_2. \end{array} \tag{6}$$

These equalities along with the duality relation (4) result in the calculation of inner products

$$\langle \omega_1, \omega_1 \rangle = \tfrac{1}{2}, \quad \langle \omega_1, \omega_2 \rangle = \tfrac{1}{2}, \quad \langle \omega_2, \omega_2 \rangle = 1. \tag{7}$$

The vectors described in (6) are depicted in Figure 1.

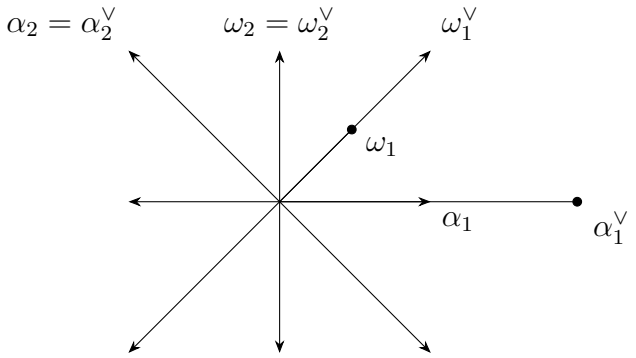


Figure 1: The arrows denote the vectors belonging to the root system Π of type C_2. The circles denote ω_1 and $\alpha_1^\vee$ which do not belong to Π.

A lattice in $\mathbb{R}^2$ is formed by integer linear combinations of vectors from a given basis of $\mathbb{R}^2$. The $\mathbb{Z}$–spans of the α– and $\alpha^\vee$–bases form the root and coroot lattice,

$$Q = \mathbb{Z}\alpha_1 \oplus \mathbb{Z}\alpha_2, \quad Q^\vee = \mathbb{Z}\alpha_1^\vee \oplus \mathbb{Z}\alpha_2^\vee. \tag{8}$$

The dual lattices of Q and $Q^\vee$ are defined as

$$Q^\perp = \{v \in \mathbb{R}^2 \mid \langle v, a\rangle \in \mathbb{Z}, \forall a \in Q\}, \quad (Q^\vee)^\perp = \{v \in \mathbb{R}^2 \mid \langle v, a\rangle \in \mathbb{Z}, \forall a \in Q^\vee\}. \tag{9}$$

Using the duality relations (4) and (5), the equalities

$$P = (Q^\vee)^\perp = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2, \quad P^\vee = Q^\perp = \mathbb{Z}\omega_1^\vee \oplus \mathbb{Z}\omega_2^\vee \tag{10}$$

are obtained. A lattice $A \subset \mathbb{R}^2$ is W-invariant if

$$w \cdot A \subset A, \quad \forall w \in W. \tag{11}$$

Using the crystallographic property of C_2, it can be checked that Q and $Q^\vee$ are both W-invariant. Moreover, if A is W-invariant, then its dual $A^\perp$ is W-invariant as well, hence P and $P^\vee$ are W-invariant.

For the root system Π of type C_2, the inclusions $Q \subseteq P$ and $Q^\vee \subseteq P^\vee$ hold [1]. Furthermore, the definition (2) and the normalization (1) result in the obvious inclusions $Q^\vee \subsetneq Q$ and $P^\vee \subsetneq P$. Finally, the explicit relations between the bases (6) for Π of type C_2 imply the equalities $Q = P^\vee$ and $2P = Q^\vee$, giving the final inclusion relationships between the four lattices

$$2P = Q^\vee \subsetneq Q = P^\vee \subsetneq P. \tag{12}$$

The orders of the quotient groups of lattices are given as absolute values of the determinants of the corresponding transition matrices,

$$|P/Q| = |P/P^\vee| = |Q/Q^\vee| = 2, \quad |P/Q^\vee| = 4. \tag{13}$$

The corresponding disjoint decomposition of the weight lattice P into Q–cosets is

$$P = Q \cup (\omega_1 + Q) \tag{14}$$

and the coweight lattice $P^\vee$ decomposes into

$$P^\vee = Q^\vee \cup (\omega_2^\vee + Q^\vee). \tag{15}$$

In the remainder of this article, we will use the abbreviated notation

$$v = a\omega_1 + b\omega_2 \equiv (a, b) \in \mathbb{R}^2 \tag{16}$$

for any vector expressed in the ω–basis. Using (6), the vector $v = (v_1, v_2)$ satisfies

$$v \in Q \Leftrightarrow v_1 \equiv 0 \mod 2, \quad v \in \omega_1 + Q \Leftrightarrow v_1 \equiv 1 \mod 2. \tag{17}$$

An admissible shift of Q is a vector $\zeta \in \mathbb{R}^2$ such that

$$w(\zeta + Q) \subseteq \zeta + Q, \quad \forall w \in W. \tag{18}$$

There exists an equivalence between admissible shifts $\zeta_{1,2}$ of Q,

$$\zeta_1 \sim \zeta_2 \quad \Leftrightarrow \quad \zeta_1 + Q = \zeta_2 + Q. \tag{19}$$

The zero shift 0 is obviously admissible, moreover any $q \in Q$ is equivalent to 0. The shifts by $q \in Q$ are called trivial. The vector ζ is an admissible shift of Q if and only if $\zeta - r_i\zeta \in Q$ for $i = 1, 2$, hence the simple reflection formulas (3) imply that ζ is admissible if and only if $\zeta \in P$. The coset decomposition (14) forces every non-trivial admissible shift of Q to be equivalent to ω_1. The notion of admissibility was introduced in [7] and [8].

3 The Weyl group and its extensions

The Weyl group W of a root system of type C_2 is generated by the reflections $r_{1,2}$ and contains 8 distinct elements. The orbit of a vector $v = (a, b) \in \mathbb{R}^2$ is obtained using the generators (3) as

$$W \cdot v \equiv W \cdot (a, b) = \{\pm(a, -a - b), \pm(a, b), \pm(a + 2b, -a - b), \pm(a + 2b, -b)\}. \tag{20}$$

For every $a \in A$, where A is a W–invariant lattice, the mapping $T(a) : \mathbb{R}^2 \to \mathbb{R}^2$ is defined as the translation by a, explicitly $T(a)v = v + a$ for all $v \in \mathbb{R}^2$. The generalized affine Weyl group W_A^{aff} is generated by the simple reflections $r_{1,2}$ and the shifts $T(a), a \in A$. For convenience, the group of shifts $\{T(a) \mid a \in A\}$ is identified with A. Since A is a normal subgroup of W_A^{aff}, the only shift contained in W is $T(0)$ and $W_A^{\text{aff}} = T(A)W$, it holds that

$$W_A^{\text{aff}} = A \rtimes W. \tag{21}$$

The fundamental domain of an affine Weyl group is a subset F_A of $\mathbb{R}^2$ such that

1. $W_A^{\text{aff}} \cdot F_A = \mathbb{R}^2$,
2. F_A contains at most one point from every orbit of W_A^{aff}.

Fundamental regions of the generalized affine Weyl groups using the lattices Q and $Q^\vee$ can be found in [3]. The fundamental region $F_Q = F_{P^\vee}$ is formed as the convex hull of the points $0, \omega_1, \frac{1}{2}\omega_2$, explicitly

$$F_Q = \{(x_1, x_2) \mid x_0 + x_1 + 2x_2 = 1, x_{0,1,2} \geq 0\} \tag{22}$$

and $F_{Q^\vee}$ is the convex hull of the vertices $0, \frac{1}{2}\omega_1^\vee, \omega_2^\vee$. After transforming into ω–basis, this set can be described as

$$F_{Q^\vee} = \{(y_1, y_2) \mid y_0 + y_1 + y_2 = 1, y_{0,1,2} \geq 0\}\,. \tag{23}$$

From these explicit descriptions of F_Q and $F_{Q^\vee}$, it is clear that

$$F_Q = F_{P^\vee} \subsetneq F_{Q^\vee}. \tag{24}$$

The group $W_Q^{\text{aff}} = W_{P^\vee}^{\text{aff}}$ can be formed as the semidirect product

$$W_Q^{\text{aff}} = W_{P^\vee}^{\text{aff}} = W_{Q^\vee}^{\text{aff}} \rtimes \Gamma^\vee, \quad \Gamma^\vee \cong Q/Q^\vee. \tag{25}$$

The group $\Gamma^\vee$ contains two elements

$$g_0 = \text{id}, \quad g_2^\vee = T(\omega_2^\vee) r_2 r_1 r_2 \tag{26}$$

which leave the fundamental domain $F_{Q^\vee}$ invariant [2],

$$\Gamma^\vee F_{Q^\vee} \subseteq F_{Q^\vee}. \tag{27}$$

The projection $\tau : W_Q^{\text{aff}} \to Q$, defined as $\tau(T(q)w) = q$ for every $q \in Q$ and $w \in W$, is used to define the homomorphism $\theta_{\omega_1} : W_Q^{\text{aff}} \to U(1)$ as

$$\theta_{\omega_1}(z) = e^{2\pi \mathrm{i} \langle \tau(z), \omega_1 \rangle}, \quad z = T(q)w \in W_Q^{\text{aff}}. \tag{28}$$

For each $x \in F_Q$, let $\text{Stab}_{\Gamma^\vee}(u)$ denote the stabilizer of x in the group $\Gamma^\vee$. The reduced fundamental domain $F_Q(\omega_1) \subsetneq F_Q$ includes only the points $x \in F_Q$ such that

$$\theta_{\omega_1}\left(\text{Stab}_{\Gamma^\vee}(x)\right) = \{1\}. \tag{29}$$

The inner products (7) and the algorithm for determining the stabilizers from [3] imply that

$$F_Q(\omega_1) = \{(x_1, x_2) \mid x_0 + x_1 + 2x_2 = 1, x_0 > 0, x_{1,2} \geq 0\}\,. \tag{30}$$

Using the symbol $\star$ in the triple $[x_0, x_1, x_2]$ of barycentric coordinates of $x \in F_Q$ from (22) to denote a non-zero element, the counting function $\varepsilon_Q : F_Q \to \mathbb{N}$ is defined as

$$\varepsilon_Q(x) = \begin{cases} 8 & x = [\star, \star, \star], \\ 4 & x = [0, \star, \star], [\star, 0, \star], [\star, \star, 0], \\ 2 & x = [0, 0, \star], \\ 1 & x = [0, \star, 0], [\star, 0, 0]. \end{cases} \tag{31}$$

Analogously, the coordinates $[y_0, y_1, y_2]$ of $y \in F_{Q^\vee}$ appearing in (23) are used to define the counting function

$$\varepsilon_{Q^\vee}(y) = \begin{cases} 8 & y = [\star, \star, \star], \\ 4 & y = [0, \star, \star], [\star, 0, \star], [\star, \star, 0], \\ 2 & y = [0, \star, 0], \\ 1 & y = [0, 0, \star], [\star, 0, 0]. \end{cases} \tag{32}$$

The values of ε_Q and $\varepsilon_{Q^\vee}$ correspond to the cardinalities of the orbits Wx and Wy on the tori $\mathbb{R}^2/Q$ and $\mathbb{R}^2/Q^\vee$ respectively (see [3] for details). The order of the stabilizer of $x \in F_Q$ with respect to the action of W_Q^{aff} will be denoted $h_Q(x)$. The number $h_Q(x)$ can be calculated using the orbit-stabilizer theorem as

$$h_Q(x) = \frac{8}{\varepsilon_Q(x)}, \quad x \in F_Q. \tag{33}$$

The function $d : F_{Q^\vee} \to \mathbb{N}$ denotes the order of $\mathrm{Stab}_{\Gamma^\vee}(y)$ for any $y \in F_{Q^\vee}$. The formula (26) implies that $g_2^\vee \cdot (y_1, y_2) = (y_1, y_2)$ if and only if $1 - y_1 = 2y_2$, hence $d(y)$ is given explicitly by

$$d(y) = \begin{cases} 2 & 2y_2 = 1 - y_1, \\ 1 & \text{otherwise.} \end{cases} \tag{34}$$

4 C–functions in the root system C_2

C–functions are complex functions which respect the symmetries of the generalized affine Weyl group. The C–function $\Phi_p^\sigma : \mathbb{R}^2 \to \mathbb{C}$ is defined as

$$\Phi_p^\sigma(x) = \sum_{w \in W} e^{2\pi \mathrm{i} \langle wp, x \rangle} \tag{35}$$

for any $p \in P$ [4]. Utilizing the abbreviated notation (16) for the vectors $p = (p_1, p_2) \in P$ and $v = (v_1, v_2) \in \mathbb{R}^2$, the explicit form of $\Phi_p(v)$ can be calculated as

$$\begin{aligned} \Phi_p(v) = \; 2\big[\; & \cos(\pi(p_2 v_1 - p_1 v_2)) + \cos(\pi(p_1 v_2 + p_2(v_1 + 2v_2))) \\ & + \cos(\pi(p_1(v_1 + v_2) + p_2 v_1)) + \cos(\pi(p_1(v_1 + v_2) + p_2(v_1 + 2v_2))\big] \end{aligned} \tag{36}$$

using (20) and the values of the inner products (7).

Let us list the pertinent properties of the C–functions:

1. symmetry with respect to $w \in W$:

$$\Phi_p(wv) = \Phi_{wp}(v) = \Phi(v), \quad \forall v \in \mathbb{R}^2; \tag{37}$$

2. if $A \subseteq P$ is a W–invariant lattice, then for any $a \in A, a^\perp \in A^\perp$ it holds that

$$\Phi_a(v + a^\perp) = \Phi_a(x), \quad \forall v \in \mathbb{R}^2, \tag{38}$$

   hence φ_a^σ is well-defined on $\mathbb{R}^2/A^\perp$;

3. for any $q \in Q$ and $v \in \mathbb{R}^2$, the following relations hold:

$$\Phi_q(g_2^\vee v) = \Phi(v), \quad \Phi_{\omega_1 + q}(g_2^\vee v) = \theta(g_2^\vee)\Phi_{\omega_1+q}(v) = -\Phi_{\omega_1+q}(v). \tag{39}$$

5 Point and label sets

For any $M \in \mathbb{N}$, the label set Λ_M is defined as the fragment of lattice P belonging to the reduced fundamental domain $F_Q(\omega_1)$ magnified by the factor M,

$$\Lambda_M = P \cap MF_Q(\omega_1). \tag{40}$$

Using (30), elements of Λ_M can be expressed in terms of the ω–basis as

$$\Lambda_M = \left\{(\lambda_1, \lambda_2) \mid \lambda_0 \in \mathbb{N}, \lambda_{1,2} \in \mathbb{Z}^{\geq 0}, \lambda_0 + \lambda_1 + 2\lambda_2 = M\right\}. \tag{41}$$

The decomposition (14) allows the grid Λ_M to be partitioned into the disjoint sets

$$\Lambda_M^{(0)} = Q \cap MF_Q(\omega_1), \quad \Lambda_M^{(1)} = (\omega_1 + Q) \cap MF_Q(\omega_1). \tag{42}$$

Moreover, the equivalence (17) combined with the explicit form (30) of the reduced fundamental domain $F_Q(\omega_1)$, results in the descriptions

$$\Lambda_M^{(k)} = \left\{(\lambda_1, \lambda_2) \mid \lambda_0 \in \mathbb{N}, \lambda_{1,2} \in \mathbb{Z}^{\geq 0}, \lambda_0 + \lambda_1 + 2\lambda_2 = M, \lambda_1 \equiv k \mod 2\right\}. \tag{43}$$

The point set F_M is defined as the fragment of the rescaled shifted lattice $\frac{1}{M}(\omega_1 + Q)$ inside the fundamental domain $F_{Q^\vee}$,

$$F_M = \frac{1}{M}(\omega_1 + Q) \cap F_{Q^\vee}. \tag{44}$$

Using the description (23) of $F_{Q^\vee}$ and the modular equation (17), F_M is given in terms of linear combinations of vectors from ω–basis as

$$F_M = \left\{\left(\tfrac{t_1}{M}, \tfrac{t_2}{M}\right) \mid t_{0,1,2} \in \mathbb{Z}^{\geq 0}, t_0 + t_1 + t_2 = M, t_1 \equiv 1 \mod 2\right\}. \tag{45}$$

Each of the label sets $\Lambda_M^{(0,1)}$ is associated with a splitting point set $F_M^{(0,1)}$, defined as

$$F_M^{(0)} = \frac{1}{M}(\omega_1 + Q) \cap F_Q, \quad F_M^{(1)} = \frac{1}{M}(\omega_1 + Q) \cap F_Q(\omega_1). \tag{46}$$

The descriptions (22) of F_Q and (30) of $F_Q(\omega_1)$ along with (17) result in the following expressions of $F_M^{(0,1)}$ in terms of ω–basis:

$$F_M^{(0)} = \left\{\left(\tfrac{s_1}{M}, \tfrac{s_2}{M}\right) \mid s_{0,1,2} \in \mathbb{Z}^{\geq 0}, s_0 + s_1 + 2s_2 = M, s_1 \equiv 1 \mod 2\right\}, \tag{47}$$

$$F_M^{(1)} = \left\{\left(\tfrac{s_1}{M}, \tfrac{s_2}{M}\right) \mid s_0 \in \mathbb{N}, s_{1,2} \in \mathbb{Z}^{\geq 0}, s_0 + s_1 + 2s_2 = M, s_1 \equiv 1 \mod 2\right\}. \tag{48}$$

6 Orthogonality relations of C–functions

The inner product of the complex functions f, g on the grid F_M is defined as

$$\langle f, g\rangle_{F_M} = \sum_{s \in F_M} \varepsilon_{Q^\vee}(s) f(s) \overline{g(s)}. \tag{49}$$

The C–functions defined by the sums (35), are orthogonal with respect to (50),

$$\langle \Phi_\mu, \Phi_\nu \rangle_{F_M} = 16M^2 h_Q\left(\tfrac{\mu}{M}\right)\delta_{\mu\nu}, \quad \mu,\nu \in \Lambda_M. \tag{50}$$

Moreover, it holds that $|\Lambda_M| = |F_M|$, therefore the set of orbit functions $\{\Phi_\lambda \mid \lambda \in \Lambda_M\}$ forms an orthogonal basis of the associated Hilbert space [7]. Similarly, the inner products on the point sets $F_M^{(k)} \subseteq F_M, k = 0, 1$ are given by

$$\langle f, g \rangle_{F_M}^{(k)} = \sum_{s \in F_M^{(k)}} \varepsilon_Q(s) f(s) \overline{g(s)}. \tag{51}$$

The functions $\Phi_{\lambda,\nu}$, labeled by the elements $\lambda, \nu \in \Lambda_M^{(k)}$ satisfy the relation

$$\langle \Phi_\mu, \Phi_\nu \rangle_{F_M}^{(k)} = 8M^2 h_Q\left(\tfrac{\mu}{M}\right)\delta_{\mu\nu}, \quad \mu,\nu \in \Lambda_M^{(k)} \tag{52}$$

for each $k \in \{1, 2\}$. As in the previous case, it holds that $|\Lambda_M^{(k)}| = |F_M^{(k)}|$ for $k = 1, 2$, so the set $\{\Phi_\lambda \mid \lambda \in \Lambda_M^{(k)}\}$ forms an orthogonal basis of each respective Hilbert space. The orthogonality relations (52) and the equal cardinality of the label and point set can be deduced from the general discrete orthogonality theorem for Weyl orbit functions which will be proven in [9].

The Fourier-Weyl transform on the discrete point sets $F_M, F_M^{(0)}$ and $F_M^{(1)}$ are carried out by multiplying the column vector of data by a square matrix. Any arbitrary ordering of the label set Λ_M and point set F_M and the orthogonality relation (50) give rise to the transform matrix $\mathbb{I}_M$, whose entries are defined as

$$[\mathbb{I}_M]_{\lambda,s} = \sqrt{\frac{\varepsilon_{Q^\vee}(s)}{16M^2 h_Q(\frac{\lambda}{M})}}\overline{\Phi_\lambda(s)}, \quad \lambda \in \Lambda_M, s \in F_M. \tag{53}$$

The relation (50) forces the matrix $\mathbb{I}_M$ to be unitary.

For $k = 1, 2$, the matrices $\mathbb{I}_M^{(k)}$, associated with the discrete Fourier-Weyl transform on the point set $F_M^{(k)}$, are determined by

$$\left[\mathbb{I}_M^{(k)}\right]_{\lambda,s} = \sqrt{\frac{\varepsilon_Q(s)}{8M^2 h_Q(\frac{\lambda}{M})}}\overline{\Phi_\lambda(s)}, \quad \lambda \in \Lambda_M^{(k)}, s \in F_M^{(k)} \tag{54}$$

and an ordering of the sets $\Lambda_M^{(k)}$ and $F_M^{(k)}$. Analogously, equation (52) implies that $\mathbb{I}_M^{(0)}$ and $\mathbb{I}_M^{(1)}$ are unitary matrices.

7 Central splitting of the transform matrix

Now fix the orderings of the sets $\Lambda_M^{(0)}$ and $\Lambda_M^{(1)}$, then order the set Λ_M so that the elements of $\Lambda_M^{(0)}$ are situated before the points from $\Lambda_M^{(1)}$, each subset using its respective ordering. Define the unitary splitting matrix $\mathbb{T}_M$ as the block matrix

$$\mathbb{T}_M = \begin{pmatrix} \mathbb{T}_M^{(0)} \\ \mathbb{T}_M^{(1)} \end{pmatrix}, \tag{55}$$

where the elements $\mathbb{T}_M^{(k)}, k = 1, 2$ are defined as

$$\left[\mathbb{T}_M^{(0)}\right]_{s,t} = \frac{1}{\sqrt{2d(s)}}\left(\delta_{s,t} + \delta_{g_2^\vee \cdot s,t}\right), \quad s \in F_M^{(0)}, t \in F_M, \tag{56}$$

$$\left[\mathbb{T}_M^{(1)}\right]_{s,t} = \frac{1}{\sqrt{2d(s)}}\left(\delta_{s,t} + \theta_{\omega_1}(g_2^\vee)\delta_{g_2^\vee \cdot s,t}\right), \quad s \in F_M^{(1)}, t \in F_M. \tag{57}$$

The matrix $\mathbb{T}_M$ relates $\mathbb{I}_M$ with the splitting matrices $\mathbb{I}_M^{(0)}$ and $\mathbb{I}_M^{(1)}$ via the relation

$$\mathbb{I}_M = \left(\mathbb{I}_M^{(0)} \oplus \mathbb{I}_M^{(1)}\right)\mathbb{T}_M \tag{58}$$

where the rows associated with the labels $\lambda \in \Lambda_M$ are ordered as described above. The ordering of the columns of $\mathbb{I}_M$ is determined by the splitting matrix (55).

Choosing $M = 5$, the label set is given by

$$\Lambda_5 = \{(0,2),(2,1),(4,0),(0,1),(2,0),(0,0),(1,1),(3,0),(1,0)\} \tag{59}$$

and it splits into the disjoint union $\Lambda_5 = \Lambda_5^{(0)} \cup \Lambda_5^{(1)}$ whose constituents are

$$\Lambda_5^{(0)} = \{(0,2),(2,1),(4,0),(0,1),(2,0),(0,0)\}, \tag{60}$$

$$\Lambda_5^{(1)} = \{(1,1),(3,0),(1,0)\}. \tag{61}$$

The corresponding point set F_5 is

$$F_5 = \left\{(\tfrac{1}{5},\tfrac{4}{5}),(\tfrac{3}{5},\tfrac{2}{5}),(1,0),(\tfrac{1}{5},\tfrac{3}{5}),(\tfrac{3}{5},\tfrac{1}{5}),(\tfrac{1}{5},\tfrac{2}{5}),(\tfrac{3}{5},0),(\tfrac{1}{5},\tfrac{1}{5}),(\tfrac{1}{5},0)\right\} \tag{62}$$

and the splitting point sets $F_5^{(0)}$ and $F_5^{(1)}$ are

$$F_5^{(0)} = \left\{(\tfrac{1}{5},\tfrac{2}{5}),(\tfrac{3}{5},\tfrac{2}{5}),(1,0),(\tfrac{1}{5},\tfrac{1}{5}),(\tfrac{3}{5},0),(\tfrac{1}{5},0)\right\}, \tag{63}$$

$$F_5^{(1)} = \left\{(\tfrac{1}{5},\tfrac{1}{5}),(\tfrac{3}{5},0),(\tfrac{1}{5},0)\right\}. \tag{64}$$

The transform matrix $\mathbb{I}_5$ is calculated from the definition (53) as

$$\mathbb{I}_5 = \tfrac{1}{10}\begin{pmatrix}
\sqrt{2}(\sqrt{5}-1) & -2\sqrt{3+\sqrt{5}} & 4 & -2 & 3-\sqrt{5} & 3+\sqrt{5} & -2\sqrt{3+\sqrt{5}} & -2 & \sqrt{2}(\sqrt{5}-1) \\
2 & 2 & -4\sqrt{2} & -3\sqrt{2} & 2\sqrt{2} & 2\sqrt{2} & 2 & -3\sqrt{2} & 2 \\
\sqrt{7-3\sqrt{5}} & \sqrt{7+3\sqrt{5}} & 4 & -2 & -2(1+\sqrt{5}) & 2(\sqrt{5}-1) & \sqrt{7+3\sqrt{5}} & -2 & \sqrt{7-3\sqrt{5}} \\
2\sqrt{3+\sqrt{5}} & \sqrt{2}-\sqrt{10} & -4 & 2 & -3-\sqrt{5} & \sqrt{5}-3 & \sqrt{2}-\sqrt{10} & 2 & 2\sqrt{3+\sqrt{5}} \\
\sqrt{7+3\sqrt{5}} & \sqrt{7-3\sqrt{5}} & 4 & -2 & 2(\sqrt{5}-1) & -2(1+\sqrt{5}) & \sqrt{7-3\sqrt{5}} & -2 & \sqrt{7+3\sqrt{5}} \\
2\sqrt{2} & 2\sqrt{2} & 2 & 4 & 4 & 4 & 2\sqrt{2} & 4 & 2\sqrt{2} \\
-2\sqrt{5} & 2\sqrt{5} & 0 & \sqrt{10} & 0 & 0 & -2\sqrt{5} & -\sqrt{10} & 2\sqrt{5} \\
\frac{\sqrt{5}-5}{\sqrt{2}} & -\sqrt{5(3+\sqrt{5})} & 0 & 2\sqrt{5} & 0 & 0 & \sqrt{5(3+\sqrt{5})} & -2\sqrt{5} & \sqrt{15-5\sqrt{5}} \\
-\sqrt{5(3+\sqrt{5})} & \frac{\sqrt{5}-5}{\sqrt{2}} & 0 & -2\sqrt{5} & 0 & 0 & \sqrt{15-5\sqrt{5}} & 2\sqrt{5} & \sqrt{5(3+\sqrt{5})}
\end{pmatrix}.$$

Analogously, it follows from (54) that the transform matrices $\mathbb{I}_5^{(0)}$ and $\mathbb{I}_5^{(1)}$ take the explicit forms

$$\mathbb{I}_5^{(0)} = \tfrac{1}{10}\begin{pmatrix}
3+\sqrt{5} & 3-\sqrt{5} & 4 & -2\sqrt{2} & -2(1+\sqrt{5}) & 2(\sqrt{5}-1) \\
2\sqrt{2} & 2\sqrt{2} & -4\sqrt{2} & -6 & 2\sqrt{2} & 2\sqrt{2} \\
2(\sqrt{5}-1) & -2(1+\sqrt{5}) & 4 & -2\sqrt{2} & 3+\sqrt{5} & 3-\sqrt{5} \\
\sqrt{5}-3 & -3-\sqrt{5} & -4 & 2\sqrt{2} & 2-2\sqrt{5} & 2(1+\sqrt{5}) \\
-2(1+\sqrt{5}) & 2(\sqrt{5}-1) & 4 & -2\sqrt{2} & 3-\sqrt{5} & 3+\sqrt{5} \\
4 & 4 & 2 & 4\sqrt{2} & 4 & 4
\end{pmatrix},$$

$$\mathbb{I}_5^{(1)} = \frac{1}{10}\begin{pmatrix} -2\sqrt{5} & -2\sqrt{10} & 2\sqrt{10} \\ -2\sqrt{10} & 5+\sqrt{5} & 5-\sqrt{5} \\ 2\sqrt{10} & 5-\sqrt{5} & 5+\sqrt{5} \end{pmatrix}.$$

The splitting transform matrices $T_5^{(1,2)}$ are determined by (56) and (57) as

$$\mathbb{T}_5^{(0)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0&0&0&0&0&\sqrt{2}&0&0&0 \\ 0&0&0&0&\sqrt{2}&0&0&0&0 \\ 0&0&\sqrt{2}&0&0&0&0&0&0 \\ 0&0&0&1&0&0&0&1&0 \\ 0&1&0&0&0&0&1&0&0 \\ 1&0&0&0&0&0&0&0&1 \end{pmatrix}, \quad \mathbb{T}_5^{(1)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0&0&0&-1&0&0&0&1&0 \\ 0&-1&0&0&0&0&1&0&0 \\ -1&0&0&0&0&0&0&0&1 \end{pmatrix}.$$

A direct verification shows that (58) is satisfied.

8 Conclusion

The main result (58) connects the transform matrix $\mathbb{I}_M$, which contains data from a large point set F_M, to the matrices $\mathbb{I}_M^{(0)}$ and $\mathbb{I}_M^{(1)}$, defined on a smaller set of points, via the sparse splitting matrix $\mathbb{T}_M$. For large M, this process offers advantages in computational speed, since the set Λ_M is split into two independent subsets, parallelizing the calculation. Moreover, sparsity of $\mathbb{T}_M$ allows for further optimisations.

Other cases of central splitting mechanism have been studied. For the Lie algebra A_1, it is known as the fast split-radix transform, even allowing for recursive splitting of the label sets [10]. Moreover, explicit description for A_2 has been given in the article [6], including proofs. No general recursive central splitting mechanism of any other type than A_1 is known at the moment of writing of this article. Its existence remains an unsolved problem.

This article relies on yet unpublished general theory of central splitting of Weyl orbit functions which relies on the generalized theory of discrete orthogonality, generalizing the results examined in the articles [3, 7, 11, 5] and [8]. Besides their usefulness for defining Fourier-like transforms, Weyl orbit functions found applications in physics, for example in [12] and [13]. Future areas of interest may include determining whether a recursive splitting process can be defined for all types of orbit functions in the most general case, development of general theory of discrete orthogonality for the so-called E–functions [14], which were not examined in this text, study of the corresponding transforms and determining whether a similar mechanism exists for E–functions as well.

References

[1] Richard Kane. *Reflection Groups and Invariant Theory*. Springer Science & Business Media, 2001.

[2] Nicolas Bourbaki. *Lie Groups and Lie Algebras: Chapters 4-6*. Springer Berlin Heidelberg, 2002.

[3] Jiří Hrivnák and Jiří Patera. On discretization of tori of compact simple Lie groups. *Journal of Physics A: Mathematical and Theoretical*, 42(38):385208, 2009.

[4] Anatoliy Klimyk and Jiří Patera. Orbit functions. *SIGMA. Symmetry, Integrability and Geometry: Methods and Applications*, 2:006, 2006.

[5] Jiří Hrivnák and Lenka Motlochová. Dual–root lattice discretization of Weyl orbit functions. *Journal of Fourier Analysis and Applications*, 25(5):2521–2569, 2019.

[6] Jiří Hrivnák, Mariia Myronova, and Jiří Patera. Central splitting of A_2 discrete Fourier–Weyl transforms. *Symmetry*, 12(11):1828, 2020.

[7] Tomasz Czyżycki and Jiří Hrivnák. Generalized discrete orbit function transforms of affine Weyl groups. *Journal of Mathematical Physics*, 55(11), 2014.

[8] Tomasz Czyżycki, Jiří Hrivnák, and Lenka Motlochová. Generalized dual–root lattice transforms of affine Weyl groups. *Symmetry*, 12(6):1018, 2020.

[9] Vojtěch Teska, Jiří Hrivnák, and Lenka Motlochová. Discrete orthogonality of Weyl orbit functions on invariant lattices and corresponding transforms. *(in preparation)*.

[10] Vladimir Britanak, Patrick C Yip, and Kamisetty Ramamohan Rao. *Discrete cosine and sine transforms: general properties, fast algorithms and integer approximations*. Elsevier, 2010.

[11] Jiří Hrivnák and Mark A. Walton. Weight–lattice discretization of Weyl–orbit functions. *Journal of Mathematical Physics*, 57(8), 2016.

[12] Adam Brus, Jiří Hrivnák, and Lenka Motlochová. Quantum particle on dual weight lattice in Weyl alcove. *Symmetry*, 13(8):1338, 2021.

[13] Jiří Hrivnák and Lenka Motlochová. On electron propagation in triangular graphene quantum dots. *Journal of Physics A: Mathematical and Theoretical*, 55(12):125201, 2022.

[14] Jiří Hrivnák and Jiří Patera. E–discretization of tori of exceptional compact simple Lie groups. In *AIP Conference Proceedings*, volume 1307, pages 89–94. American Institute of Physics, 2010.

Hilbert Envelope-Based EEG Classification for Alzheimer's Disease Detection*

Nichita Vatamaniuc
vatamnic@cvut.cz

study programme: Applied Informatics
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jaromír Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstrakt. Alzheimerova choroba (AD) je běžná neurodegenerativní porucha charakterizovaná progresivním kognitivním úpadkem a změněnou elektrickou aktivitou mozku. Tato studie představuje komplexní rámec pro automatickou klasifikaci AD pomocí spektrálních charakteristik získaných z obálek EEG signálů.

Metodologie využívá 19-kanálová EEG měření od 56 pacientů (28 pacientů s AD, 28 kontrol), získaných při vzorkovací frekvenci 200 Hz. Zpracování signálů používá diskrétní Hilbertovu transformaci k extrakci analytických signálů a výpočtu obálek, které zachycují charakteristiky amplitudové modulace. Signály obálek jsou segmentovány na jedno-sekundové epochy, přičemž spektra výkonu jsou počítána pomocí rychlé Fourierovy transformace a průměrována mezi segmenty pro extrakci charakteristik.

Byly vyhodnoceny tři klasifikační přístupy: ν-Support Vector Classification (ν-SVC), k-Nearest Neighbors (KNN) a Ridge Regression (RR). Křížová validace pomocí leave-one-out a 70/30 dělení trénovací a testovací sady zajistila robustní hodnocení výkonu. Klasifikátor ν-SVC dosáhl optimálního výkonu s citlivostí 91 %, zatímco KNN a RR poskytly porovnatelné výsledky (83 %). Konzistence napříč různými klasifikačními paradigmaty potvrzuje robustnost spektrálních charakteristik založených na obálkách pro diskriminaci AD.

Budoucí práce se zaměří na současná omezení v efektivitě zpracování, která omezují metodologii na offline analýzu. Nekonečná impulzní odezva tradiční implementace DHT zavádí výpočetní zpoždění, která nejsou kompatibilní s real-time klinickým monitoringem. Směry výzkumu zahrnují vývoj nových návrhů filtrů s konečnou impulzní odezvou (FIR), které kombinují pásmový filtr s modifikovanou Hilbertovou transformací. Imponováním specifických matematických omezení na amplitudové spektrum filtru $F(\omega)$ – včetně sudé symetrie, omezeného rozsahu $[0, 1]$, spojitých druhých derivací a nulových okrajových podmínek při $\omega = 0$ a $\omega = \pi$ – mohou koeficienty filtrů dosáhnout dekadentních rychlostí $O(k^{-3})$ ve srovnání se standardním DHT s $O(k^{-1})$. Tento pokrok umožní přesnou extrakci obálek pomocí filtrů délky menší než 50 koeficientů ve srovnání s více než 200 koeficienty v původních implementacích, což sníží zpoždění při zpracování. Navrhovaná dvoucestná architektura bude počítat in-fázové složky pomocí pásmového filtru, zatímco kvadraturové složky budou podrobeny kombinovanému pásmovému filtru a Hilbertově transformaci, přičemž obálka bude odvozena jako velikost výsledného komplexního signálu. Takové pokroky umožní nasazení překrývajících se filtrů cílených na specifické frekvenční pásma, což usnadní aplikace v reálném čase, včetně detekce epilepsie, klasifikace spánkových fází a mozkových počítačových rozhraní. Přechod od ověřené offline analýzy k efektivnímu

*This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS23/190/OHK4/3T/14.

online zpracování představuje klíčový krok k klinické implementaci automatizovaných systémů neurologického hodnocení, slibujících vylepšené schopnosti pro včasnou intervenci v neurodegenerativních poruchách.

Klíčová slova: Alzheimerova choroba, EEG, frekvenční charakteristiky, klasifikace, automatizovaná diagnostika.

Abstract. Alzheimer's Disease (AD) represents a prevalent neurodegenerative disorder characterized by progressive cognitive decline and altered brain electrical activity. This research presents a comprehensive framework for automatic AD classification using spectral features derived from EEG signal envelopes.

The methodology utilizes 19-channel EEG recordings from 56 patients (28 AD patients, 28 controls) acquired at 200 Hz sampling frequency. Signal processing employs Discrete Hilbert Transform to extract analytic signals and compute envelopes, capturing amplitude modulation characteristics. Envelope signals are segmented into one-second epochs, with power spectra computed via Fast Fourier Transform and averaged across segments for feature extraction.

Three classification approaches were evaluated: ν-Support Vector Classification (ν-SVC), k-Nearest Neighbors (KNN), and Ridge Regression (RR). Cross-validation using leave-one-out and 70/30 train-test splits ensured robust performance assessment. The ν-SVC classifier achieved optimal performance with 91% sensitivity, while KNN and RR produced comparable results (83%). Consistency across diverse classification paradigms validates the robustness of envelope-based spectral features for AD discrimination.

Future work will address current limitations in processing efficiency that restrict the methodology to offline analysis. The infinite impulse response nature of traditional DHT implementation introduces computational delays incompatible with real-time clinical monitoring. Research directions include development of novel finite impulse response (FIR) filter designs combining band-pass filtering with modified Hilbert transformation. By imposing specific mathematical constraints on the filter's amplitude spectrum $F(\omega)$ – including even symmetry, bounded range $[0, 1]$, continuous second derivatives, and zero boundary conditions at $\omega = 0$ and $\omega = \pi$ – filter coefficients can achieve $O(k^{-3})$ decay rates compared to standard DHT's $O(k^{-1})$. This improvement enables accurate envelope extraction using filters of length less than 50 coefficients versus over 200 from original paper implementations, reducing processing delays. The proposed dual-path architecture will compute in-phase components through band-pass filtering while quadrature components undergo combined band-pass and Hilbert transformation, with the envelope derived as the magnitude of the resulting complex signal. Such advances will enable deployment of overlapping filter banks targeting specific frequency bands, facilitating real-time monitoring applications including epilepsy detection, sleep stage classification, and brain-computer interfaces. The transition from validated offline analysis to efficient online processing represents a critical step toward clinical implementation of automated neurological assessment systems, promising enhanced capabilities for early intervention in neurodegenerative disorders.

Keywords: Alzheimer's disease, EEG, frequency features, classification, automated diagnostics.

Full paper: N. Vatamaniuc, J. Kukal, O. Vyšata: Hilbert Envelope-Based EEG Classification for Alzheimer's Disease Detection. Under review in Archives of Control Sciences.

The Best Linear Sub-Model Selection with Entropy Based Regularization*

František Voldřich
voldrichfrantisek@gmail.com

study programme: Mathematical Engineering
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jaromir Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. This paper presents a unified framework for linear sub-model selection based on the principle of entropy maximization. By maximizing entropy under suitable constraints, we derive generalized normal distributions that define both the error model and the prior on regression weights, linking distributional assumptions directly to regularization penalties. The resulting quasi-likelihood formulation integrates likelihood maximization with regularization and allows flexible control of robustness and sparsity through two shape parameters. We develop a convex optimization approach for parameter estimation and propose a heuristic binary optimization algorithm for efficient sub-model selection guided by Bayesian Information Criteria (BIC). The framework is experimentally evaluated on simulated regression tasks with both binary and continuous weights under varying noise conditions. The experiments systematically examine the influence of the distributional parameters and on the accuracy and stability of sub-model identification. The results show that for low and moderate noise levels, the proposed method reliably recovers the true sub-model and achieves information criterion values comparable to those of established regularization techniques such as ridge and lasso regression.

Keywords: linear model, entropy maximization, quasi-likelihood maximization, regularized estimates, information criteria, sub-model selection, binary heuristics

Abstrakt. Tento článek představuje jednotný rámec pro výběr lineárních podmodelů založený na principu maximalizace entropie. Maximalizací entropie za vhodných omezení odvozujeme generalizovaná normální rozdělení, která definují jak model chyb, tak apriorní rozdělení regresních koeficientů, čímž přímo propojují distribuční předpoklady s penalizacemi regularizace. Výsledná formulace kvazi-věrohodnosti integruje maximalizaci věrohodnosti s regularizací a umožňuje flexibilní kontrolu robustnosti a řídkosti prostřednictvím dvou parametrů. Navrhujeme konvexní optimalizační postup pro odhad parametrů a heuristický binární algoritmus pro efektivní výběr podmodelu, řízený podle Bayesovského informačního kritéria (BIC).

Přístup je experimentálně ověřen na simulovaných regresních úlohách s binárními i spojitými váhami při různých úrovních šumu. Experimenty systematicky zkoumají vliv distribučních parametrů na přesnost a stabilitu identifikace podmodelu. Výsledky ukazují, že při nízkých a středních úrovních šumu navržená metoda spolehlivě rekonstruuje skutečný podmodel a dosahuje hodnot informačního kritéria srovnatelných s etablovanými regularizačními metodami, jako jsou ridge a lasso regrese.

*This work has been supported by the grant SGS23/190/OHK4/3T/14 of Czech Technical Universityin Prague.

Klíčová slova: lineární model, maximalizace entropie, maximalizace kvazi-věrohodnosti, regularizované odhady, informační kritéria, výběr submodelu, binární heuristiky

Full paper: F. Voldrich, J. Kukal. *The Best Linear Sub-Model Selection with Entropy Based Regularization*, Submitted for publication to *Statistics and Computing*, September 2025.

Basic Operations in Square and Hexagonal Grids

Jan Vondruška
vondrj35@cvut.cz

study programme: Applied Informatics
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Jaromír Kukal, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstract. This paper investigates fundamental concepts of mathematical morphology applied to binary images represented in square and hexagonal grids. It introduces the basic operations of dilation, erosion, opening, and closing, and explains their properties and interrelations. Special attention is given to skeletonization and object reconstruction, highlighting the role of maximal disks and Lantuéjoul's formulation. Furthermore, the paper compares the characteristics of square and hexagonal grids, emphasizing connectivity, isotropy, and topological preservation. The results show that the hexagonal lattice provides smoother skeletons and more natural representations of objects, making it advantageous for applications in image analysis.

Keywords: Morphology, square grid, hexagonal grid

Abstrakt. Tento článek se zabývá základními koncepty matematické morfologie aplikované na binární obrazy reprezentované ve čtvercové a hexagonální mřížce. Jsou představeny základní operace jako dilatace, eroze, otevření a uzavření, včetně jejich vlastností a vzájemných vztahů. Zvláštní pozornost je věnována skeletonizaci (vytváření koster) a rekonstrukci objektů s důrazem na maximální kruhy a Lantuéjoulovu formulaci. Dále jsou porovnány vlastnosti čtvercové a hexagonální mřížky, zejména z hlediska konektivity, izotropie a zachování topologie. Výsledky ukazují, že hexagonální mřížka poskytuje hladší kostry a přirozenější reprezentaci objektů, což je výhodné pro aplikace v analýze obrazů.

Klíčová slova: Morfologie, čtvercová mřížka, hexagonální mřížka

1 Introduction

This text explores the theoretical foundations and operations of digital morphology applied to binary images represented in square and hexagonal grids. Mathematical morphology is a powerful tool in image processing that examines the shape and structure of objects within an image using theoretical and lattice-based approaches. The primary focus is on fundamental morphological operations and their properties, the construction of skeletons (medial axes) of objects, and the differences arising from the underlying grid topology.

2 Morphological Operations in Square Grid

2.1 Basic Concepts

2.1.1 Image Representation

An image is viewed as a function $f : M \to \{0, 1\}$, where $M \subset \mathbb{Z}^2$ represents the discrete set of pixel coordinates in a square grid. The pixels are typically arranged so that the top-left corner corresponds to the coordinate $[1, 1]$, with the x -axis increasing to the right and the y -axis increasing downwards.

2.1.2 Object

An object (A) is a set of pixels that has a specific property $p : M \to \{0, 1\}$

$$A = \{a \in M \mid p(a)\},$$

2.1.3 Background

Let $A \subset M$ be an object. Background A^C of the object A is defined as a set of pixels not belonging to the object A.

$$A^C = \{a \in M \mid \neg p(a)\}$$

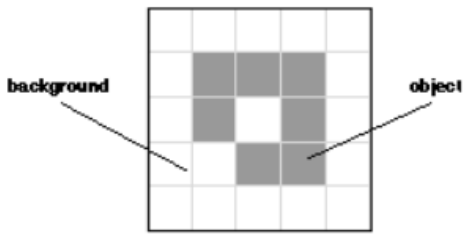


Figure 1: Binary image in which gray pixels represent the object and white pixels represent the background. [1]

2.1.4 Translation

The translation of a set A by vector x is defined as

$$A + x = \{a + x \mid a \in A\}$$

2.1.5 Minkowski addition and substraction

Minkonwski operations for sets $A, B \in M$ are defined as

$$A \oplus B = \bigcup_{b \in B} (A + b), \; A \ominus B = \bigcap_{b \in B} (A - b)$$

2.1.6 Dilation and Erosion

Let A, B be objects in M. Dilation and Erosion are defined by Minkowski operations. In general, dilatation causes the expansion of the object and

$$D(A, B) = A \oplus B$$

Erosion shrinks the object.

$$E(A, B) = A \ominus (-B)$$

, where $-B = \{-b \mid b \in B\}$. While either set A or set B can be thought of as an "image", A is usually considered as the image, and B is called a structuring element.

The two most common structuring elements (given a square grid) are the 4-connected and 8-connected sets, N_4 and N_8.

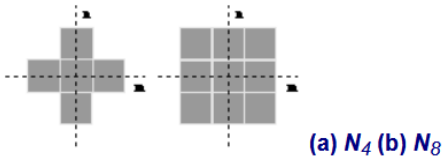


Figure 2: The N_4 and N_8 structuring elements. [1]

2.2 Properties

- Commutativity of dilation: $D(A, B) = D(B, A)$
- Non-commutativity of Erosion
- Associativity of Dilation: $A \oplus (B \oplus C) = (A \oplus B) \oplus C$
- Duality: $D^C(A, B) = E(A^C, -B)$, $E^C(A, B) = D(A^C, -B)$
- Translation Invariance: $A \oplus (B + x) = (A \oplus B) + x$

2.3 Opening and Closing

Dilation and erosion can be combined to build two important higher-order operations - opening (O) and closing (C) [1].

$$O(A, B) = A \circ B = D(E(A, B), B)$$

$$C(A, B) = A \bullet B = E(D(A, B), -B)$$

[1]Opening can be used to separate objects that are connected in the image. This operation smooths the object from the outside and removes its thin edges. The closing function fills small holes in the object and smooths it from the inside. Sequential opening and closing application allows the image to be smoothed and provides a more accurate understanding of its structure.

2.4 Additional Morphological Operators

2.4.1 Top-hat transform

The top hat highlights small elements or details removed by opening.

$$TH(A, B) = A - O(A, B)$$

2.4.2 Bottom-hat transform

The bottom hat emphasizes small holes or dark regions filled by closing.

$$BH(A, B) = C(A, B) - A$$

2.4.3 Minkowski Sausage

It has two defined versions, and these operators extract boundary regions with a thickness controlled by the structuring element.

$$MS^I(A, B) = D(A, B) - E(A, B)$$

$$MS^{II} = MS^I(\delta A, B) = D(\delta A, B) - E(\delta A, B) = D(\delta A, B)$$

where δA is the boundary of A.

These operators extract boundary regions with a thickness controlled by the structuring element.

2.4.4 Hit-or-Miss operation

Let $B = \{B_1, B_2\}$ where B_1, B_2 are disjoint structuring elements.

$$A \otimes B = E(A, B_1) \bigcap E(A^C, B_2)$$

This operation can be used to detect boundary pixels.

3 Skeleton and Object reconstruction

The skeleton may be defined as a line representation of an object.
Line properties:

- One-pixel thick
- Goes through the middle of the object
- Preserves the topology of the object

Sometimes, it is not possible to find the skeleton of an object.

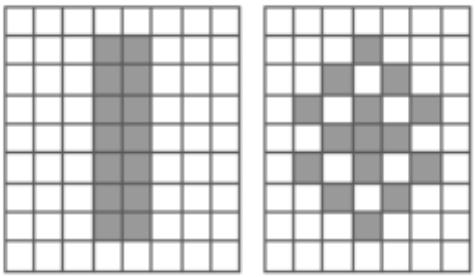


Figure 3: Objects without the skeleton. [1]

3.1 Disk in M

Let M be the base set $\alpha \in M, r \in R$. A disk in M with the center at α and radius r is defined as:

$$K(\alpha, r) = \{\beta \in M \mid \|\beta - \alpha\|_2 \leq r\}$$

where $\|.\|$ is the Euclidean norm.

3.2 Maximal disk representation

Let M be the base set and $K(\alpha, r)$ a disk in M. We say that the disk $K(\alpha, r)$ is maximal in A if $K(\alpha, r) \subseteq A$ and touches the boundary of the object at least at two points.

3.3 Lantuéjoul skeleton

The skeleton subset $S_k(A)$ is defined as:

$$S_k(A) = E^k(A, B) - O(E^k(A, B), B), \ k = 0, 1..., K$$

The skeleton is defined as:

$$S(A) = \bigcup_{k=0}^{K} S_k(A)$$

This formulation for the skeleton does not preserve the topology.

An effect of this skeleton is that the original object can be reconstructed given knowledge of the subsets of the skeleton $S_k(A)$, the structuring element B, and K:

$$A = \bigcup_{k=0}^{K} D(S_k(A), kB)$$

where $kB = (B \oplus B...)$ denotes the k-fold dilatation.

4 Morphology on Hexagonal Grids

In the hexagonal grid, pixels are arranged so that each pixel has six equidistant neighbors sharing a common edge, unlike the square grid, where pixels have 4 or 8 neighbors with varying distances.

The hexagonal grid $M \subset Z^2$ differs from the square one in the arrangement of pixels as regular hexagons, with rows shifted to achieve the equal distance between the centers of the neighboring pixels.

Due to the differing topology, connectivity and neighborhood definitions vary between even and odd rows, requiring separate treatment of structuring elements depending on parity.

4.1 Morphological Operations

Dilation, erosion, opening, and closings are analogously defined as in the square grid but adapted to handle the shifted rows and neighbor patterns of the hexagonal lattice.

Skeletonization and reconstruction methods carry over, but with specific adaptations.

- Two types of Golay alphabets (I and II), crafted for hexagonal topology, are introduced to perform sequential thinning for skeleton extraction.
- The endpoint detection alphabets are expanded to accommodate hexagonal connectivity.

4.2 Comparison of Square and Hexagonal Grids in Morphology

1. Hexagonal grids provide uniform neighbor distances and better connectivity, making definitions of boundaries and skeletons more consistent and natural.
2. Square grids suffer from ambiguous connectivity at pixel corners (diagonal adjacency), complicating boundary and skeleton definitions.
3. Skeletons obtained in hexagonal grids tend to be smoother and topologically preserved more effectively.
4. The hexagonal lattice is generally more suitable for representing natural shapes and avoids the directional bias inherent in square grids.

4.3 Operations

4.3.1 Dilation

$$D(A, B) = A \oplus B = \{z \in M \mid (B)_z \cap = \emptyset\}$$

4.3.2 Erosion

$$E(A, B) = A \ominus B = \{z \in M \mid (B)_z \subseteq A\}$$

4.3.3 Opening

$$O(A, B) = D(E(A, B), B)$$

4.3.4 Closing

$$C(A, B) = E(D(A, B), B)$$

4.3.5 Skeleton decomposition

$$S_k(A) = E^k(A)\ O(E^k(A), B),\ \ A = \bigcup_{k=0}^{K} S_k(A) \oplus B^k$$

5 Conlusion

The theoretical foundations of digital morphology in square and hexagonal grids reveal the pros and cons of hexagonal grids for image representation, particularly in smoothness, uniform connectivity, and natural shape approximation. Morphological operations have the same fundamental properties, but their implementation must be adjusted to the topological differences of the lattice. Skeletonization strategies adapt accordingly, with hexagonal grids offering enhanced preservation of object topology, making them advantageous for applications in biological and medical image analysis.

References

[1] *Morphology-based Operations.*
`https://cmp.felk.cvut.cz/cmp/courses/dzo/resources/course_ip_tudelft/course/fip-Morpholo.html`.
Accessed: September 21, 2025

[2] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*. Pearson Education India, 2004.

[3] Ch. Lantuéjoul, *Sur le modèle de Johnson-Mehl généralisé*, Technical Report, Centre de, 1977.

Autoencoder-Based Anomaly Detection in AMBER Monitoring Plots

Samuel Zahorec
samuel.zahorec@cern.ch

study programme: Applied Informatics
Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

advisor: Josef Nový, Department of Software Engineering
Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague

Abstrakt. or modern high-energy physics experiments, it is crucial to assess the health and efficiency of the wide range of detector setups. This assessment is usually accomplished by sending a fraction of collected data into a monitoring stream and filling the histograms with these monitoring data. Today's experiments employ a myriad of detectors, each often consisting of a large number of subplanes and smaller units. This means that even when encountering medium-sized experiments, such as the AMBER experiment at CERN [4], one may end up with hundreds or even thousands of monitoring plots. While the monitoring data can be used in later stages of the analysis, in order to confirm the good status of the detectors at the time of recording, ideally, the detector health should be monitored online during the actual data acquisition to warn about the critical conditions as soon as possible. The online monitoring then heavily relies on detector experts and, mainly, shifters who are supposed to be present in the control room and notice any kind of abnormal behavior. This approach then leaves room for human error and the inherent inability to track the vast numbers of monitoring histograms. Approaches to detect anomalies using artificial intelligence (notably autoencoder neural networks) then naturally offer a very appealing way to avoid wasting human potential, given that the algorithms are capable of learning from the collected datasets and helping to draw the attention of the shifters to the suspicious anomalies. Going even further, these artificial intelligence anomaly-detection algorithms could potentially spot trends and dependencies often hidden from people who do not have as much experience and knowledge about all the detectors and their subsystems, thus not only accelerating the anomaly recognition but actually improving it in the early stages of the data acquisition.

1 Introduction

Apparatus for Meson and Baryon Research (AMBER) is a medium-sized experiment based at Conseil Européen pour la Recherche Nucleaire (CERN), Europe's largest particle accelerator complex, situated at the border between Switzerland and France. The main objectives of the experiment are to study the structure of the composite particles belonging to the classes of baryons (3 quarks) and mesons (2 quarks), and to determine the radius of the proton for which the measurements done in the past show significant discrepancies (cite). AMBER is, up to now, the largest surface experiment at CERN. It consists of dozens of detector systems, including, for example, silicon fibers (SciFis),

gas-electron multipliers (GEMs), or multi-wire proportional chambers (MWPCs), with the latest addition of the time projection chamber (TPC).

The ultimate analysis of the physics data relies heavily on monitoring the health and correct performance of these detectors. For monitoring purposes, a fraction of acquired data is extracted into the so-called monitoring stream. Dedicated programs then process this monitoring data and produce a set of hundreds of monitoring plots, each containing information about the particular physics aspect of the data collected by the detector (e.g., timing, channel occupancies, or 2D hit maps). These monitoring plots are then displayed in the control room for shifters to detect immediately the potential problems and inform detector experts, or log these findings to be taken into account during the later stages of analysis. This approach is largely susceptible to human error, not to mention the fact that navigating in the hundreds or even thousands of different plots is a challenging task, and only the experts dedicated to one particular detector system are capable of noticing the issues that are of a more subtle nature. Statistically-based methods to determine anomalous conditions offer a solution to cope with this problem, however, they may not always recognize all possible issues, and their performance is determined by the expert knowledge of the detectors and a narrow set of conditions to be considered faulty.

With the latest boom of AI, many techniques of anomaly detection are now being widely studied and employed in various fields of applications, ranging from video surveillance systems [8] to removing anomalies in the data preprocessing steps [10], or identifying system failures in the complex networks of IoT devices [3]. The main objective of this study is to develop a model capable of recognizing anomalous monitoring plots. Rather than studying the basic machine learning algorithms, this work explores possibilities in the field of deep learning and neural networks. The data collected by the AMBER experiment is framed by the so-called runs, which offer the highest level of granularity regarding the data-taking time. Each run corresponds to hundreds of spills (finer time units derived from the operational aspects of the SPS accelerator) and corresponds to roughly one or two hours of data taking (absolute timespan depends on the beam conditions). After the run is finished, shifters are required to label it depending on the quality of the data. If the conditions are mostly plausible, the run is labeled as a "Good Run". That, however, doesn't mean that all the detector monitoring plots are completely correct. Therefore, the training dataset for this study was made of only good runs, but also contained the anomalous plots, which were unlabeled. Inherently, the learning itself is unsupervised, and the choice of a neural network had to reflect this learning mode.

2 Autoencoders

A widely used type of neural network for unsupervised learning is an autoencoder [7]. The goal of this neural network is to learn a compressed representation of the input data in two steps: encoding and decoding. The first part, the encoder, compresses the input data into a latent space representation (i.e., in a reduced dimension). After encoding the input data, the decoder part attempts to reconstruct the data from the encoded representation, with the output being as close as possible to the original input. The key idea is to train the autoencoder to minimize the reconstruction errors. For obvious reasons, autoencoders are particularly useful in the context of anomaly detection. Given that the training dataset

contains mostly normal data, the autoencoder is able to learn its representation. When it attempts to reconstruct input data that deviates significantly from the normal data, it will produce a large reconstruction error, implying that the reconstruction went poorly and the data is, from the point of view of the model, anomalous. The crucial step is, naturally, determining the threshold of reconstruction error from which the input data can be considered anomalous. In an ideal scenario, when the input data contains only normal samples, the threshold may be set reasonably far away from the reconstruction errors distribution. In the case of the monitoring plots, which may contain anomalies already in the training set, the situation becomes more difficult, and the distributions of the reconstruction errors may exhibit long tails. This issue, as well as how the author dealt with the threshold settings, will be discussed in detail in the ?? section.

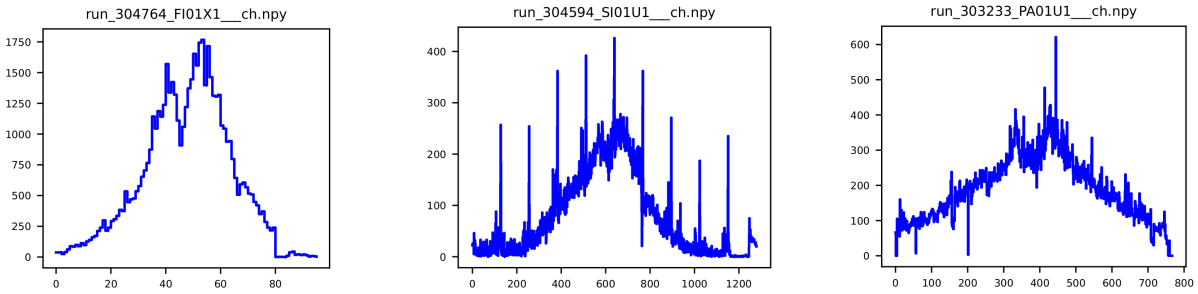
In principle, both the encoder and decoder parts of the autoencoder may contain multiple hidden layers of nodes with activation functions. The number of these layers, as well as the number of nodes in each layer, is highly application-specific and depends on the complexity of the data representations. Implementing more than one layer may bring several advantages. In general, multiple layers allow the network to learn more complex and nonlinear representations. While lower layers often learn to recognize simple patterns, deeper layers specialize in more complicated patterns. It is also very important not to reduce the dimensions too drastically. Multiple layers leave more space for a gradual dimensional reduction, which can help to preserve important information through the network. From the point of view of anomaly detection, employing multiple layers allows for extracting more robust and discriminative features, thus increasing the chances of anomalies scoring higher reconstruction errors.

Nevertheless, basic autoencoders are not particularly well-suited for image reconstruction tasks. In general, when working with datasets consisting of images, convolutional neural networks represent a more effective approach. Consequently, in recent years, researchers have proposed a more advanced subclass of autoencoders, known as convolutional autoencoders [6]. In these architectures, the encoder corresponds to a conventional convolutional neural network composed of convolutional and pooling layers, which progressively reduce the spatial dimensionality of the input data [5]. The decoder, in turn, employs transposed convolutional layers to reconstruct the image and restore its original dimensions. Owing to the characteristics of the data considered in this study, convolutional autoencoders constitute a highly promising modeling approach.

3 Dataset Selection and Hardware

For the purpose of designing and developing the anomaly detection model, a dataset containing 620 monitoring ROOT [1] files from 2024 AMBER data-taking was collected. AMBER tracks all monitoring files and their parameters in a custom database, each file corresponding to a particular data-taking run. As the dataset should contain data taken under very normal conditions, files were filtered according to the type of data (physics, good data) and the general labels for the detector health (no errors on detectors). Since every file contains almost 40 000 histograms, channel occupancy plots (showing number of particles hitting the individual channels of a detector) from three typical and well-known detector subsystems were chosen for performing the tests: `FI01X1` (a silicon fiber detector

plane), SI01U1 (a silicon detector plane), and PA01U1 (a multi-wire proportional chamber plane). Before the model training took place, all these histograms were extracted using a dedicated script and stored in .npz files, each containing a one-dimensional array of hit occupancies. These then served as a convenient input into the designed neural network. Figure 1 shows the three mentioned types of plots under ideal circumstances.



(a) FI01X1 (b) SI01U1 (c) PA01U1

Figure 1: Histograms of channel occupancies from the dataset used for training the model corresponding to silicon fiber detector plane (1a), silicon detector plane (1b), and multi-wire proportional chamber plane (1c). All histograms represent the detectors in their ideal state.

All tests were conducted on a personal Dell Latitude 5420 equipped with an 11th Gen Intel Core i5-1135G7 processor, 16 GB of RAM, and a 512 GB NVMe SSD.

4 Neural Network Design and Architecture

To automatically detect abnormal behavior in the monitoring histograms produced by the experiment's data acquisition system, an unsupervised anomaly detection model based on a one-dimensional convolutional autoencoder has been developed. This model learns a compact latent representation of normal histograms and identifies deviations through elevated reconstruction errors. The approach is conceptually similar to methods used in detector monitoring and calibration control in high-energy physics [9], where autoencoders are applied to learn the manifold of regular detector responses.

4.1 Data Format and Normalization

The network was trained on a collection of one-dimensional histogram data stored in .npz files, each containing a single array of floating-point values. To ensure consistent input dimensionality, all histograms were either padded or truncated to a fixed target length L_t, corresponding to the nearest multiple of 2^n, where n is the number of downsampling layers in the encoder. This guarantees that the encoder–decoder structure can symmetrically halve and then upsample the signal length at each stage without dimensional mismatch.

The model offers four selectable schemes for normalizing the set of histograms $h =$

$(h_1, h_2, \ldots, h_L)$ before training:

$$h_i^{\text{norm}} = \begin{cases} h_i / \max_j h_j & \text{(max normalization)} \\ h_i / \sum_j h_j & \text{(sum normalization)} \\ \log(1+h_i) / \max_j \log(1+h_j) & \text{(log normalization)} \\ h_i & \text{(none)} \end{cases}$$

For the task at hand, max normalization was applied, effectively restricting all values to a $[0, 1]$ range for better numerical stability and convergence during training.

4.2 Model Architecture

The proposed one-dimensional convolutional autoencoder (Conv1D-AE) was implemented in PyTorch. Its architecture consists of two main components: an encoder, which compresses the input signal into a compact latent representation, and a decoder, which reconstructs the original histogram from that representation. The autoencoder architecture is depicted in Figure 2.

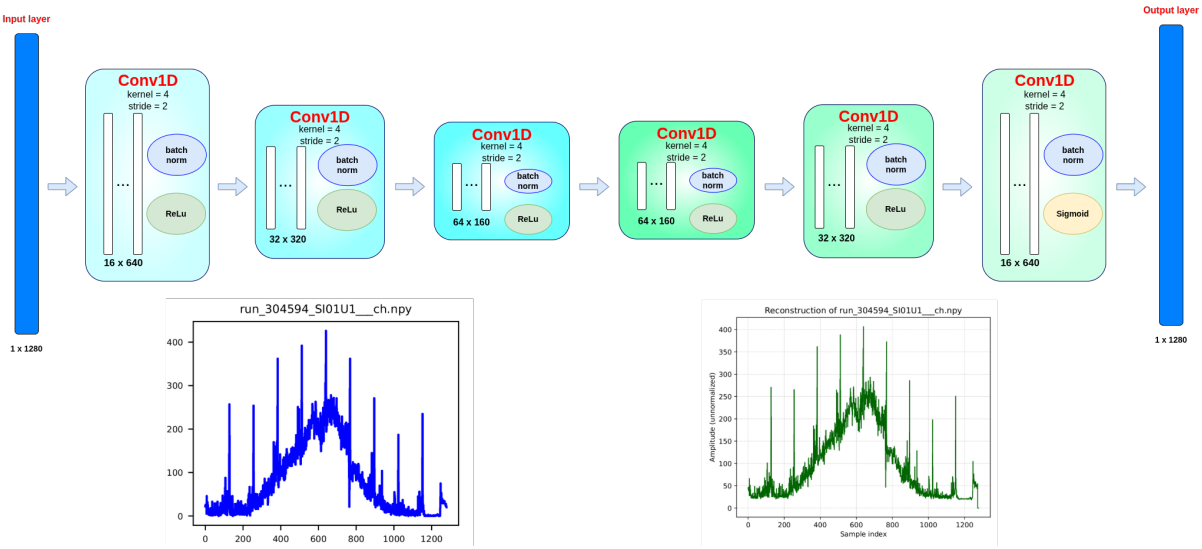


Figure 2: Illustration of the proposed one-dimensional convolutional autoencoder. In this demonstration, both encoder and decoder constitute 3 hidden layers of 16, 32, and 64 channels in the mirror arrangement. Each one-dimensional convolutional layer applies a moving window with a kernel of 4 and a stride 2, effectively doubling the learned detail at each step. With every layer, the number of channels doubles and the scale of the detailed pattern they learn halves. Each layer can optionally apply an intermediate batch normalization. Except for the output, each Conv1D layer applies a ReLu activation function to introduce nonlinearity. To ensure that the reconstructed data array has the same appearance as the input histogram, sigmoid activation is applied in the last convolutional layer. An exemplary channel occupancy plot with 1280 bins from the `SI01U1` dataset was chosen to manifest the network's behavior. The left histogram displays the original input, and the plot on the right shows the version of this histogram reconstructed by the designed autoencoder.

4.2.1 Encoder

The encoder comprises a configurable sequence of one-dimensional convolutional layers with progressively increasing channel counts (for this study, a $[16, 32, 64]$ sequence was used). Each convolutional layer applies a kernel (typically of size 4) with a stride of 2, effectively reducing the input length by half at each stage.

A 1D convolution operates by sliding a small, learnable filter across the input sequence and computing a weighted sum of local values. This allows the network to learn translation-invariant features, such as peaks or localized gradients, that characterize the histogram's shape.

After each convolution, a Rectified Linear Unit (ReLU) activation function is applied. ReLU is defined as

$$\mathrm{ReLU}(x) = \max(0, x)$$

which sets negative values to zero while keeping positive values unchanged. ReLU introduces nonlinearity into the model, allowing it to learn complex, nonlinear relationships in the data. Additionally, ReLU helps to mitigate the vanishing gradient problem that can occur with saturating activations such as sigmoid or tanh.

Optionally, batch normalization layers may follow the convolutions. Batch normalization standardizes the activations of each layer to have zero mean and unit variance within a mini-batch, yielding training acceleration and more stable gradients.

4.2.2 Decoder

The decoder mirrors the encoder's structure, using transposed convolutional layers (ConvTranspose1D) to progressively upsample the latent representation back to the original input length. Each transposed convolution approximately doubles the sequence length (inverse of the encoder's stride) and reduces the number of layer channels in reverse order.

Intermediate decoder layers again use ReLU activations and optional batch normalization to ensure stable learning and nonlinear behavior. The final decoder layer produces a single-channel output corresponding to the reconstructed histogram, followed by a Sigmoid activation function. The Sigmoid function maps each output value to the range $[0, 1]$:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

This is appropriate since normalized input histograms are bounded in the same range. The Sigmoid ensures that the reconstruction remains interpretable as a probability-like distribution.

Finally, minor cropping or zero-padding adjustments are applied to ensure the reconstructed output exactly matches the input length L_t.

4.3 Loss Function

Reconstruction quality is optimized using the Kullback–Leibler (KL) divergence, a measure of dissimilarity between two probability distributions P (the target histogram) and

Q (the reconstructed output):

$$D_{KL}(P||Q) = \sum_i P_i \log\left(\frac{P_i}{Q_i}\right)$$

Both P and Q are first transformed into valid probability distributions by applying the softmax function along the histogram bins. The softmax operation converts an arbitrary real-valued vector into a normalized distribution:

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

This ensures non-negativity and a total sum of one, making the KL divergence mathematically valid and interpretable.

The KL loss encourages the reconstructed histogram to resemble the input not only in amplitude but also in distributional shape, which is particularly meaningful for histogram-like data.

4.4 Optimization and Regularization

Model parameters are optimized using the AdamW optimizer, a decoupled variant of Adam that introduces a more principled treatment of weight decay. The Adam optimizer combines the benefits of two classical stochastic gradient descent improvements:

- Momentum - smooths gradient updates by maintaining an exponentially decaying average of past gradients
- Adaptive learning rates - adjust the update magnitude for each parameter based on the running average of squared gradients

While Adam includes an implicit form of weight decay through its adaptive learning scheme, it tends to overfit when standard L_2 regularization is naively applied. AdamW resolves this by decoupling weight decay from the gradient update, applying it directly to the parameters rather than to the loss-based gradient term. This results in better generalization and more predictable convergence behavior, especially for deep convolutional architectures.

A small weight decay term (10^{-6}) was used to penalize excessively large weights, further reducing overfitting.

4.5 Training Procedure

The dataset was split into training (70%) and validation (30%) subsets. Mini-batches of size 32 were fed into the network for a configurable number of epochs (in this case, 20 or 50). Both training and validation losses were recorded at each epoch to monitor convergence. The model was trained using the AdamW optimizer with a learning rate of 10^{-3}. The training may potentially be performed on an NVIDIA GPU if available. For this article, a personal Dell Laptop was sufficient, and therefore computations defaulted to the CPU.

4.6 Anomaly Detection and Postprocessing

After training, the model was evaluated on the full dataset to compute reconstruction errors for each histogram. As the model was trained using KL divergence, the same function was used for calculating the reconstruction error to ensure consistency and an interpretable measure of anomaly.

A statistical threshold for anomaly detection in the set of reconstruction errors $E = [E_1, \ldots, E_n]$ may be computed using one of four configurable methods:

Median-MAD	$\text{median}(E) + \lambda \times 1.4826 \times \text{MAD}(E)$
Mean Standard Deviation (MSD)	$\mu + \lambda\sigma$
Percentile cutoff	$P_p(E)$
fixed theshold	τ

where μ is the mean of the distribution, σ corresponds to the standard deviation, and

$$\text{MAD}(E) = \text{median}(|E_i - \text{median}(E)|)$$

is the median absolute deviation. The normalization constant 1.4826 comes from the assumption of normally distributed data. For a Gaussian distribution:

$$\text{MAD} \approx 0.6745 \times \sigma \Rightarrow \sigma \approx 1.4826 \times \text{MAD}$$

Therefore, the normalization constant ensures that the MAD-based estimate is comparable to the standard deviation. P_p is the p-th percentile of the error distribution. For Median-MAD and MSD approaches, λ controls the sensitivity of the anomaly detection. As the threshold selection is the most crucial step in labeling the anomalies, it is discussed in greater detail in section 5.2.

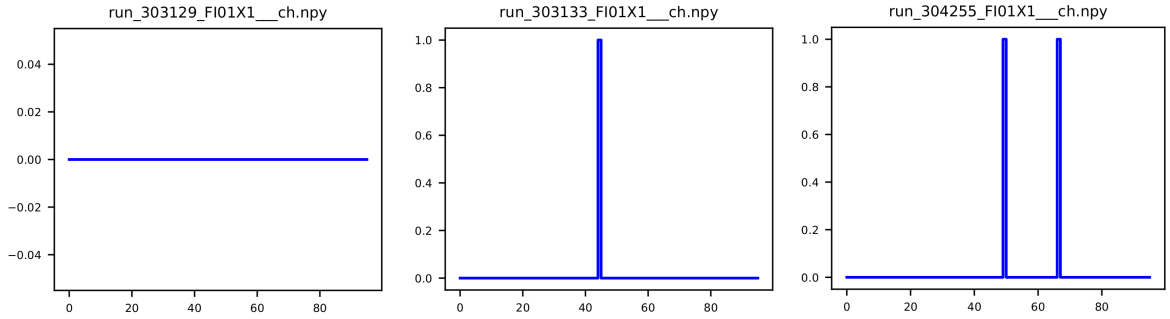
Finally, the model automatically generates a multi-page PDF report, containing:

- Training and validation loss curves
- The distribution of reconstruction errors and detection thresholds
- Visual comparisons of original and reconstructed histograms for the highest-error (most anomalous) samples

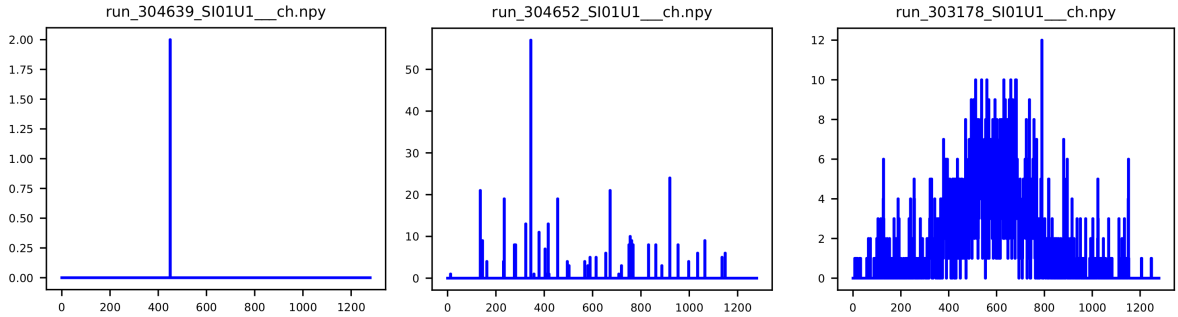
This end-to-end pipeline enables both quantitative and qualitative assessment of the model's anomaly detection performance on the AMBER monitoring data.

5 Performance Evaluation

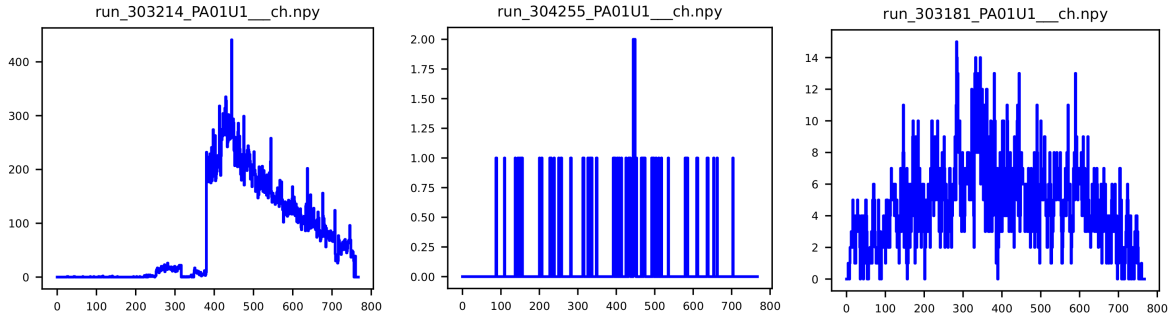
To evaluate the performance of the designed autoencoder, the author of this study made an effort to label at least the most anomalous plots out of all datasets. The examples of what may certainly be considered as anomalous plots are given in Figure 3. The count and exact shape of such plots per dataset were kept for later comparison with the output of the model.



(a) `FI01X1`



(b) `SI01U1`



(c) `PA01U1`

Figure 3: Examples of anomalous plots found in datasets `FI01X1`(3a), `SI01U1`(3b), and `PA01U1`(3c). As can be seen, plots may certainly be considered anomalous when they are empty or contain a very low number of hits. Furthermore, abnormal behavior is often observed in plots that have certain parts missing or are too noisy.

5.1 Model Parameters

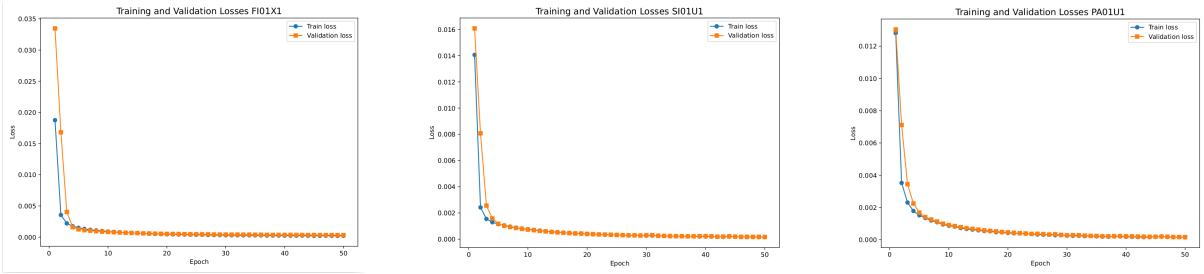
The convolutional autoencoder was trained on the `FI01X1`, `SI01U1`, and `PA01U1` datasets using the configuration summarized in Table 1. The network consisted of three convolutional blocks with channel sizes $[16, 32, 64]$ and a kernel size of 5. Batch normalization was applied after each convolution layer to stabilize training. The input histograms were normalized using max-scaling and padded to the nearest multiple of 2^n based on the number of downsampling steps. The model was trained for 50 epochs using the AdamW optimizer with a learning rate of 1×10^{-3} and a weight decay of 1×10^{-6}. The KL divergence was used as the reconstruction loss.

Parameter	Value
Datasets	`FI01X1`, `SI01U1`, `PA01U1`
Input normalization	`max`
Convolutional channels	[16, 32, 64]
Kernel size	5
Batch normalization	Enabled
Learning rate	1×10^{-3}
Weight decay	1×10^{-6}
Batch size	32
Number of epochs	50
Loss function	Kullback–Leibler divergence (KL loss)
Optimizer	AdamW
Device	CPU

Table 1: Summary of the one-dimensional convolutional autoencoder training parameters.

5.1.1 Training and Validation Losses

To validate the training procedure, the dataset was separated into training (70%) and validation (30%) subsets. Each dataset was further split into mini-batches of size 32. Training then continued in an epoch-wise fashion. In each epoch, the model was trained across all batches of the training data, and the training loss (error of the plot reconstruction by the autoencoder) was computed using the KL divergence. Similarly, but without the actual training, the network was validated on all batches of the validation data, resulting in the validation loss. Both training and validation losses across the epochs were then compared to determine at which point the losses reach their minima and stabilize. One also needed to be aware of the validation loss not starting to increase, as is often the case when overtraining. As can be seen in Figure 4, after 50 epochs, the training and validation losses stabilized, and the model was learning efficiently.



(a) `FI01X1` (b) `SI01U1` (c) `PA01U1`

Figure 4: Comparisons of training and validation losses through 50 training epochs. For all input datasets `FI01XI` (4a), `SI01UI` (4b), and `PA01UI` (4c), training and validation losses reached their minima and stabilized after 50 epochs.

5.2 Threshold Selection

With all parameters configured, three individual models were trained on the studied datasets. After training, the datasets were once again processed by the model to estimate the resulting reconstruction errors (see Figure 5).

(a) `FI01X1` (b) `SI01U1` (c) `PA01U1`

Figure 5: Reconstruction errors recovered from the models trained individually on datasets `FI01X1` (5a), `SI01U1` (5b), and `PA01U1` (5c). All counts are displayed in a logarithmic scale to improve the visualization of outliers. In all cases, the main peak (ideally containing the normal) plots is followed by a long tail corresponding to the anomalous plots.

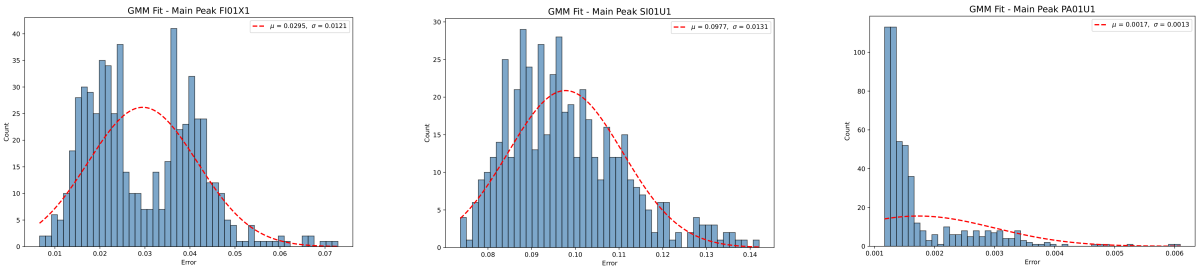
The autoencoder model was originally designed to label outliers according to a threshold computed using one of four schemes defined in 4.6. The model being developed should be robust enough to handle potentially thousands of different types of plots in the future. The fixed threshold option was immediately discarded since it would require establishing this threshold for every dataset manually. The percentile approach is also not well-suited because of the unlabeled nature of the data. One, in principle, doesn't know how many anomalous plots to expect. After careful investigation, the Median-MAD method turned out to be the most robust. Despite its robustness, it still requires careful consideration to define the sensitivity parameter λ (how many standard deviations from the median to consider the reconstruction error an outlier).

As can be seen from the reconstruction error plots in Figure 5, the distributions clearly contain a main peak and the following long tail. With this in mind, a new method, based on fitting Gaussians for determining outliers, was developed. A Gaussian Mixture Model (GMM) from the `scikit-learn` [2] library was applied to the obtained reconstruction errors. The employed GMM had 2 components, basically dividing the distribution into its main peak and tail. All histograms with the reconstruction error outside the main peak were then considered anomalous.

5.3 Anomaly Detection

With two statistical methods to establish the anomaly detection threshold at hand (Median-Mad and GMM fit), the ultimate analysis was performed using both of them to assess their individual performance. The obtained results were evaluated by comparison with the manually prepared labeled set of anomalous plots.

To support the GMM approach, Figure 6 shows the GMM fits of the main peak of reconstruction errors. For simplicity, the fits of the distribution tails were omitted due to very different scales.



(a) `FI01X1` (b) `SI01U1` (c) `PA01U1`

Figure 6: GMM fits of the main peak of reconstruction errors for all studied datasets.

The performance of each of the thresholding methods was evaluated with the well-known metrics: precision, recall, and F1 score. When comparing the manually selected anomalous plots with the ones predicted by the model, one can traditionally compute four categories:

Category	**Autoencoder response**
True Positive (TP)	*Correct anomaly detection*
False Positive (FP)	*False alarm*
False Negative (FN)	*Missed anomaly detection*
True Negative (TN)	*Correct normal plot detection*

From these categories, it is possible to calculate the different metrics as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} = 2 \cdot \frac{\text{Precission} \cdot \text{Recall}}{\text{Precission} + \text{Recall}}$$

Precision illustrates how many of the anomalies the algorithm claimed to find were actually correct, while recall describes the number of true anomalies found by the algorithm. Therefore, high precision means few false alarms, and high recall suggests few misses. One could also think of the precision as the measure of the quality of detections, while the recall is the coverage of detections. The F1 score is a harmonic mean of precision and recall, and provides a balanced measure when trying to trade off between missing anomalies and generating false alarms.

Results of the tests are summarized in Table 2. Two baselines were used for computing the threshold for anomaly detection: GMM fit with two components and Median-MAD with 3σ sensitivity. For Median-MAD, one could experiment with the sensitivity selection.

Dataset	Method	Precision	Recall	F1 Score
FI01X1	GMM (2-comp.)	0.74	1.00	0.85
	MAD (3σ)	0.70	1.00	0.82
	MAD (best λ=14)	1.00	1.00	1.00
SI01U1	GMM (2-comp.)	0.92	0.95	0.93
	MAD (3σ)	0.94	0.93	0.93
	MAD (best λ=3)	0.94	0.93	0.93
PA01U1	GMM (2-comp.)	0.93	1.00	0.97
	MAD (3σ)	0.62	1.00	0.76
	MAD (best λ=21)	0.98	0.98	0.98

Table 2: Summary of anomaly detection performance for three datasets using GMM and Median-MAD methods. For Median-MAD, both fixed 3σ and the best-performing λ (in terms of F1 score) are reported.

Therefore, the table also summarizes the λ values for which the Median-MAD technique performed best on a given dataset.

As shown in Table 2, both the GMM-based and Median–MAD approaches achieve high detection performance across all three datasets. The GMM method, which adapts its threshold through probabilistic modeling, generally provides balanced precision and recall without requiring manual tuning. In contrast, the Median–MAD technique with a fixed 3σ sensitivity exhibits larger variability across datasets, sometimes underperforming when the underlying error distribution deviates from a Gaussian-like shape.

When the Median–MAD sensitivity parameter λ is optimized per dataset, its performance can match or even exceed that of the GMM, as demonstrated for `FI01X1` and `PA01U1`. However, such tuning requires manual calibration and may not scale efficiently when thousands of datasets must be processed autonomously.

Therefore, despite the encouraging results obtained by the best-tuned Median–MAD variant, the GMM-based thresholding appears to be the most robust and practical choice for large-scale deployment. Its natural data-driven adaptation makes it suitable for automated anomaly detection pipelines, where reliability and minimal supervision are required.

6 Conclusion

This work introduced an unsupervised anomaly detection framework based on a one-dimensional convolutional autoencoder for monitoring detector histograms originating from the AMBER 2024 data-taking. The method learns the normal distribution of histogram shapes and detects deviations through reconstruction error analysis. Among the studied thresholding approaches, the Gaussian Mixture Model (GMM) with two components consistently demonstrated robust performance across datasets, providing a self-adaptive threshold without the need for manual tuning. In contrast, the Median–MAD approach was sensitive to the chosen λ value, which limits its applicability when scaling

to a large number of datasets.

The presented results confirm that the GMM-based thresholding achieves high precision and recall with minimal parameterization, making it particularly suitable for large-scale, automated monitoring environments where thousands of datasets may need to be processed continuously. This scalability is a key requirement for future integration into production data quality monitoring systems.

Future work will focus on extending this approach to multi-dimensional histogram data (e.g., 2D or even higher-dimensional distributions), where correlations between axes may carry additional diagnostic information. Another promising direction is to explore whether groups of histograms can be treated jointly as a single dataset to improve robustness and reduce noise in individual anomaly decisions. Finally, incremental or streaming variants of the model could be developed to support real-time anomaly detection and adaptive retraining as new data becomes available.

Overall, the proposed autoencoder–GMM combination provides a strong foundation for building scalable, interpretable, and fully unsupervised anomaly detection systems applicable across diverse detector monitoring tasks.

References

[1] *ROOT reference documentation.* URL: `https://root.cern/doc/v636/`. Accessed: 2025-10-19.

[2] *scikit-learn Documentation.* URL: `https://scikit-learn.org/stable/`. Accessed: 2025-10-23.

[3] A. Chatterjee and B. S. Ahmed. *Iot anomaly detection methods and applications: A survey.* Internet of Things **19** (2022), 100568.

[4] J. Friedrich. *The amber experiment at cern.* EPJ Web of Conferences **303** (09 2024), 06001.

[5] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.* O'Reilly, (2022).

[6] J. Masci, U. Meier, D. Ciresan, and J. Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. 52–59, (06 2011).

[7] A. A. Neloy and M. Turgeon. *A comprehensive study of auto-encoders for anomaly detection: Efficiency and trade-offs.* Machine Learning with Applications **17** (2024).

[8] M. Qasim and E. Verdu. *Video anomaly detection system using deep convolutional and recurrent models.* Results in Engineering **18** (2023), 101026.

[9] A. Radovic, M. Williams, D. Rousseau, et al. *Machine learning at the energy and intensity frontiers of particle physics.* Nature **560** (2018), 41–48.

[10] M. Smith and T. Martinez. Improving classification accuracy by identifying and removing instances that should be misclassified. 2690 – 2697, (09 2011).