ISSN 2579-2784 (Print)
ISSN 2538-2788 (Online)

MATHEMATICAL PROBLEMS OF COMPUTER SCIENCE

LXV

Yerevan
2026

Հայաստանի Հանրապետության Գիտությունների ազգային ակադեմիայի
Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ

Институт проблем информатики и автоматизации Национальной академии наук Республики Армения

Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia

Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ

Математические проблемы компьютерных наук

Mathematical Problems of Computer Science

LXV

ՀՐԱՏԱՐԱԿՎԱԾ Է ՀՀ ԳԱԱ ԻՆՖՈՐՄԱՏԻԿԱՅԻ ԵՎ ԱՎՏՈՄԱՏԱՑՄԱՆ ՊՐՈԲԼԵՄՆԵՐԻ ԻՆՍՏԻՏՈՒՏԻ ԿՈՂՄԻՑ
ОПУБЛИКОВАНО ИНСТИТУТОМ ПРОБЛЕМ ИНФОРМАТИКИ И АВТОМАТИЗАЦИИ НАН РА
PUBLISHED BY THE INSTITUTE FOR INFORMATICS AND AUTOMATION PROBLEMS OF NAS RA

ԵՐԵՎԱՆ 2026 YEREVAN

Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ պարբերականը հրատարակվում է տարեկան երկու անգամ ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի (ԻԱՊԻ) կողմից: Այն ընդգրկում է տեսական և կիրառական մաթեմատիկայի, ինֆորմատիկայի և հաշվողական տեխնիկայի ժամանակակից ուղղությունները:

Այն ընդգրկված է Բարձրագույն որակավորման հանձնաժողովի ընդունելի ամսագրերի ցանկում:

Տպագրվում է Խմբագրական խորհրդի 2026թ. մայիսի 29-ի N 26-05/1 նիստի որոշման հիման վրա

ISSN 2579-2784 (Print)
ISSN 2738-2788 (Online)

Математические проблемы компьютерных наук, LXV

Журнал **Математические проблемы компьютерных наук** издается два раза в год Институтом проблем информатики и автоматизации НАН РА. Он охватывает современные направления теоретической и прикладной математики, информатики и вычислительной техники.

Он включен в список допустимых журналов Высшей квалификационной комиссии.

Печатается на основании решения N 26-05/1 заседания
Редакционного совета от 29 мая 2026г.

ISSN 2579-2784 (Print)
ISSN 2738-2788 (Online)

Mathematical Problems of Computer Science, LXV

The periodical **Mathematical Problems of Computer Science** is published twice per year by the Institute for Informatics and Automation Problems of NAS RA. It covers modern directions of theoretical and applied mathematics, informatics and computer science.

It is included in the list of acceptable journals of the Higher Qualification Committee.

Printed on the basis of decision N 26-05/1 of the session of the Editorial Council dated May 29, 2026.

ISSN 2579-2784 (Print)

ISSN 2738-2788 (Online)

CONTENTS

Mathematical Problems of Computer Science 65, 7–15, 2026.
doi:10.51408/1963-0142

UDC 519.178

Deterministic Area Coverage by UAV Swarms via Multi-Eulerian Rotor-Router Walks

Vahagn S. Poghosyan

Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: povahagn@gmail.com

Abstract

The deterministic area coverage problems for UAV swarms based on multi-Eulerian rotor-router walks are investigated. The rotor-router model provides a quasi-random analogue of random walks with strong structural properties. The loop reversibility theorem is used; that is, under mild connectivity assumptions on the underlying discretized environment, a swarm of UAVs executing coordinated multi-Eulerian rotor-router walks achieves complete area coverage, in the sense that each directed edge is traversed exactly once by a single UAV. The results establish provable coverage guarantees that are robust to asynchronous execution. Furthermore, the loop reversibility theorem is generalized to the case of repeated coverage, where the area should be covered a prescribed number M of times, without requiring global synchronization among UAVs at the end of each traversal cycle. A multi-user and multi-agent platform has been developed to support mission planning and coordinated supervision of self-organizing UAV swarms.
Keywords: UAV swarm, Rotor-router model, Eulerian walk, Area surveillance, Self-organizing system.
Article info: Received 3 February 2026; sent for review 12 February 2026; accepted 20 April 2026.
Acknowledgements: This work was supported by the RA Science Committee, in the frames of the research projects 21AG-1B052 and 24DP-1B016.

1. Introduction

Unmanned Aerial Vehicle (UAV) swarms have emerged as an effective technological solution for large-scale area surveillance, environmental monitoring, and distributed sensing tasks. In contrast to single-UAV systems, swarm-based approaches offer greater robustness, scalability, and flexibility, enabling coordinated coverage of complex and dynamic environments without reliance on centralized control. These advantages have motivated extensive research on decentralized coordination, collective motion, and coverage strategies for UAV swarms.

A broad range of swarm intelligence and bio-inspired methods [1] have been proposed for UAV coordination, including particle swarm optimization, ant colony optimization, and related heuristic approaches.

A communication and control model in collaborative UAV swarms is presented in [2]. The authors identify key requirements for autonomous coordination and communication to enable efficient and reliable cooperative task execution.

In [3], the authors introduce a collision-free formation tracking framework for multiple quadrotors under switching directed communication topologies, where UAVs coordinate using only local neighborhood information and a Hooke's-law-inspired collision avoidance mechanism, supported by Lyapunov stability analysis.

The paper [4] introduces a trajectory optimization framework for multiple UAVs performing collaborative assembly tasks, combining centralized planning with distributed coordination to improve efficiency and feasibility of cooperative motion. The path planning process is solved by using a novel central force optimization genetic algorithm.

In [5], the authors present a comprehensive survey of UAV path planning techniques, classifying existing approaches into classical, soft-computing, and hybrid methods, and analyzing their applicability, advantages, and limitations in complex environments.

The paper [6] provides a comprehensive review of swarm intelligence algorithms for multi-UAV collaboration, addressing collision avoidance, task assignment, path planning, and formation reconfiguration. The survey analyzes representative algorithms, discusses their advantages and limitations, and highlights current research trends and open challenges in UAV swarm systems.

Area coverage and persistent surveillance by UAV swarms are fundamental problems in autonomous systems. While randomized strategies are widely used, deterministic algorithms for dynamic environments with provable guarantees remain comparatively underexplored.

This paper addresses this gap by leveraging rotor-router dynamics (multi-Eulerian walks) [7] to guarantee full coverage. In this setting, UAVs follow simple local rules (rotating outgoing edges and moving accordingly) yet collectively produce globally coordinated exploration patterns. Such an approach naturally supports decentralized execution, collision avoidance, and automatic workload distribution, making it well-suited for swarm-based surveillance applications.

A detailed survey of the theoretical results and fundamental properties of the rotor-router model and the related Abelian sandpile (chip-firing) model is presented in [8]. The loop-reversibility theorem, probabilistic properties, and related results are presented in [9, 10, 11].

The generalized loop-reversibility theorem for the simultaneous motion of multiple particles (chips), proved in [12], establishes deterministic guarantees for the complete traversal of the underlying graph. This theorem formed the theoretical basis for the creation of a multi-user and multi-agent platform for self-organizing UAV swarm mission planning and supervision [13, 12, 14, 15]. This platform employs graph-based representations of the operational area, enabling multiple operators to collaboratively construct mission graphs, generate Eulerian traversal structures, and supervise execution. By compiling surveillance tasks into deterministic traversal plans rather than continuous trajectories, such systems ensure consistency between planning, simulation, and real-time execution, even in the presence of asynchronous UAV behavior or dynamic swarm reconfiguration.

In this paper, we develop a deterministic framework for area coverage by UAV swarms based on multi-Eulerian rotor-router walks. Leveraging loop reversibility properties of the rotor-router model, we establish provable guarantees of complete coverage under mild con-

nectivity assumptions on the discretized environment. Furthermore, the loop reversibility theorem is generalized to the case of repeated coverage, where the area is required to be traversed a prescribed number M of times, without the need for global synchronization between UAVs at the end of each traversal cycle. The proposed approach supports asynchronous execution and provides a mathematically rigorous alternative to heuristic and probabilistic coverage strategies for swarm-based surveillance.

2. Model and Preliminaries

Consider a directed graph (digraph) $G = (V, E)$ with a set of vertices $V = V(G)$ and a set of directed edges $E = E(G)$. A spanning subgraph G' of G is a digraph with the set of vertices $V(G') = V(G)$ and a set of edges $E(G') \subseteq E(G)$. A path of length n from vertex $a \in V$ to $b \in V$ is a sequence of distinct vertices $v_1, v_2, \ldots, v_{n+1}$ such that v_i and v_{i+1} are connected by an edge $e_i \in E$, $i = 1, 2, \ldots, n$, $v_1 = a$, $v_{n+1} = b$. The path becomes a cycle if $a = b$. Edges that connect a vertex to itself are called self-loops (cycles of length 1). A cycle of length 2 consists of two adjacent vertices v_1, v_2, which are connected by a pair of oppositely directed edges from v_1 to v_2 and back. We call such cycles dimers by analogy with lattice dimers (dominoes) covering two neighboring vertices. A cycle formed by more than two edges is called a contour.

An Eulerian circuit on a finite digraph is a walk that starts and ends on the same vertex and visits each directed edge exactly once. If such a walk exists, the digraph is called Eulerian. A digraph is strongly connected if, for any two distinct vertices v and w, there are paths from v to w and from w to v. A strongly connected digraph $G = (V, E)$ is Eulerian if and only if, for each vertex $v \in V$, the in-degree and out-degree of v are equal.

The rotor-router model is defined as follows. Consider an arbitrary connected digraph $G = (V, E)$. Denote the number of outgoing edges (out-degree) from the vertex $v \in V$ by d_v. The total number of edges of G is $|E| = \sum_{v \in V} d_v$. Each vertex $v \in V$ is associated with a rotor, which is directed along one of the outgoing edges from v. The rotor directions at the vertex v are specified by an integer variable α_v, which takes values from $0 \leq \alpha_v \leq d_v - 1$ for $d_v \geq 1$.

The set $\rho = \{\alpha_v |\ v \in V,\ 0 \leq \alpha_v \leq d_v - 1\}$ defines the rotor configuration. The state of the system at any time is defined by a rotor configuration together with the position of a chip at a vertex of G. At each time step, the chip arriving at a vertex v first changes the rotor direction from α_v to $(\alpha_v + 1)$, and then moves one step along the new rotor direction from v to the corresponding neighboring vertex. The rotor directions α_v are taken modulo d_v, that is, $\alpha_v \pm d_v \equiv \alpha_v$.

The rotor configuration ρ can be considered as a spanning subgraph of G ($\rho \subset G$) with the set of vertices $V(\rho) = V(G)$ and the set of directed edges $E(\rho) \subset E(G)$ coinciding with rotors. The state of the system at any moment is given by the pair (ρ, v) of the rotor configuration ρ and the position of the chip $v \in V$. A vertex $v \in V$ is called a sink if its out-degree $d_v = 0$. In the absence of sinks, i.e., when each vertex has at least one outgoing edge, the motion of the chip does not stop.

Now let us formulate the loop reversibility theorem for a multi-particle walk [12].

Consider a bidirected contour $C = (v_1, v_2, \ldots, v_n)$ in a digraph $G = (V, E)$, that is, a contour in which the vertices v_i and v_{i+1} are connected by two edges, one in each direction, $1 \leq i \leq n$, where the n-periodicity is assumed, i.e., $v_{i \pm n} \equiv v_i$. Let e_i^+ and e_i^- be directed edges connecting v_i to v_{i+1} and v_i to v_{i-1}, respectively. Here, the n-periodicity is also assumed,

i.e., $e^+_{i\pm n} \equiv e^+_i$ and $e^-_{i\pm n} \equiv e^-_i$. The superscripts at e^-_i and e^+_i are introduced to denote the negative and positive directions at v_i with respect to the contour C, respectively. Given the rotor-router model defined on G, we say that the bidirected contour C obeys the *domino ordering* if for each vertex v_i, $1 \le i \le n$, there exists a rotor direction $\alpha^\star_{v_i}$ such that the rotor $\alpha^\star_{v_i}$ is oriented along the edge e^-_i and $\alpha^\star_{v_i}+1$ is oriented along the edge e^+_i. The directions $\alpha^\star_{v_1}, \ldots, \alpha^\star_{v_n}$ are called negative with respect to C, whereas the directions $\alpha^\star_{v_1}+1, \ldots, \alpha^\star_{v_n}+1$ are called positive, respectively.

Theorem 1. *(See [12]). Given an arbitrary finite Eulerian digraph G, let $C = (v_1, \ldots, v_n)$ be a bidirected contour obeying domino ordering. Let $k < n$ rotor-router particles start their motion at contour vertices $v_1, \ldots, v_n$, so that there is no more than one particle at one vertex. Denote these vertices as $v_{S_1}, \ldots, v_{S_k}$. Assume that the rotors $v_i, (i = 1, 2, .., n)$ at the contour C initially have positive directions. During the multi-particle walk, if a particle arrives at one of the vertices $v_{S_j}, (j = 1, 2, .., k)$ with a negatively directed rotor at that vertex, that particle stops its motion. Then, after some number of steps, regardless of the order in which the particles move, the walk produces a configuration with negative directions $\alpha^*_{v_1}, .., \alpha^*_{v_n}$. Moreover, all vertices $v_{S_1}, .., v_{S_k}$ will be occupied by one particle each. Assuming the initial rotor-router configuration is recurrent, namely, there are no other cycles except C, each rotor internal to C will perform a full rotation at the end of the walk.*

We now generalize this theorem to the case in which the area is required to be covered a prescribed number M of times, without the need for global synchronization between UAVs at the end of each traversal cycle. To establish this property, we construct M copies of the graph G that coincide at their vertex sets. Applying the above theorem to the resulting graph implies that all of its edges are traversed. Consequently, this construction is equivalent to a rotor-router walk on the original graph G, in which the contour C is reversed M times, and each edge is traversed exactly M times.

3. Application for UAV Swarm Coverage

The above result has a direct interpretation in the context of UAV swarm-based area coverage. In the multi-Eulerian rotor-router framework, repeated coverage with multiplicity M may lead to the accumulation of multiple UAVs at certain vertices of the discretized environment graph. For a vertex v_i with a degree d_i, the maximal number of UAVs that may simultaneously accumulate at v_i is bounded by $M \times d_i$, which grows linearly with the required coverage multiplicity M. While such accumulation is unavoidable in deterministic multi-agent traversal, it can be handled in a controlled manner.

To address this issue, we propose maintaining three-dimensional LIFO (last-in-first-out) stacks of UAVs at each vertex. These stacks provide a collision-free mechanism for temporarily storing multiple UAVs at the same spatial location while preserving the logical structure of the rotor-router dynamics. Furthermore, to avoid collisions during traversal along edges, we introduce an edge-priority rule for UAVs moving in intersecting or opposite directions, ensuring conflict-free motion without global coordination.

Since all UAVs are identical and the rotor-router walk possesses the Abelian property, neither the temporary accumulation of UAVs at vertices nor any permutation of UAV identities affects the final traversal state. Consequently, the proposed stacking and priority mechanisms preserve the deterministic coverage guarantees established by the theory, while

enabling practical realization of repeated area coverage by UAV swarms in three-dimensional space.

4. Multi-User and Multi-Agent Platform: Overview and Operating Regimes

To support the practical deployment of the proposed deterministic coverage algorithms, a multi-user and multi-agent platform for mission planning, execution, and supervision of self-organizing UAV swarms is developed. The platform provides a unified environment in which multiple operators can collaboratively define surveillance tasks, construct discrete mission representations, and monitor swarm execution in real time, while the UAVs themselves operate as autonomous agents executing deterministic traversal rules.

The platform is based on a graph-centric representation of the operational area, where the environment is discretized into a digraph $G = (V, E)$. Vertices represent mission-relevant locations or states, and edges encode admissible traversal paths. This representation is shared consistently across all connected users and UAV agents, ensuring that planning, simulation, and execution rely on the same underlying model. Deterministic compilation of multi-Eulerian traversal tasks allows mission specifications to be translated into reproducible swarm behaviors without dependence on probabilistic decision-making.

From the operator's perspective, the platform enables collaborative mission authoring, where multiple users can simultaneously participate in defining, modifying, and supervising the mission graph. From the UAV perspective, each agent executes local rotor-router rules using only local information, while the collective swarm behavior realizes global coverage objectives derived during mission planning.

The platform supports several operating regimes that correspond to different stages of the mission lifecycle:

1. **Mission Preparation Regime.** In this regime, users collaboratively construct and edit the mission graph, assign roles or constraints to vertices and edges, and specify coverage requirements, including repeated coverage with a prescribed multiplicity M. Additionally, all directed edges are ordered using a fixed priority to organize the above-discussed collision-free mechanism. Deterministic compilation procedures are applied to transform the mission graph into traversal structures compatible with rotor-router execution.

2. **Simulation and Validation Regime.** Before deployment, the compiled traversal plans are executed in a simulation environment. This regime allows users to validate coverage completeness, detect potential conflicts, and assess UAV accumulation behavior at vertices (stack sizes).

3. **Execution and Supervision Regime.** During mission execution, UAVs operate as autonomous agents following rotor-router dynamics. The platform provides supervision and visualization of swarm progress, enabling users to observe traversal states, vertex occupancies, and mission completion without intervention in local UAV decision-making.

4. **Adaptive Reconfiguration Regime.** The platform accommodates dynamic changes such as UAV entry or exit, provided that the underlying graph connectivity conditions are preserved. Due to the Abelian property of the rotor-router model, such reconfigurations do not affect the final coverage outcome, allowing mission continuity without global synchronization.

A key advantage of the platform is the separation between multi-user coordination at the planning and supervision level and multi-agent autonomy at the execution level. Users interact with high-level mission abstractions, while UAVs execute simple, local, deterministic

rules. This separation ensures that concurrent user interactions, UAV permutations, or asynchronous execution do not alter the final coverage guarantees, making the platform well-suited for scalable and reproducible UAV swarm surveillance missions.

A demonstration video illustrating the proposed platform is provided in [16].

5. Conclusion

This paper presented a deterministic framework for area coverage by UAV swarms based on multi-Eulerian rotor-router walks. Using loop reversibility properties, we established provable guarantees for complete and repeated coverage that are robust to asynchronous execution and independent of probabilistic assumptions. The theoretical results were further interpreted in the context of UAV swarm deployment, where UAV accumulation at vertices was shown to scale linearly with the coverage multiplicity and handled using collision-free stacking and priority mechanisms.

In addition, the proposed model naturally supports integration with a multi-user and multi-agent platform for UAV swarm mission planning and supervision, enabling deterministic task compilation, coordinated execution, and consistent supervision across multiple operators. The presented approach provides a rigorous alternative to heuristic swarm coordination methods and forms a foundation for scalable and reproducible UAV-based surveillance systems.

References

[1] E. Bonabeau, M. Dorigo, and G. Theraulaz, "Swarm intelligence: From natural to artificial systems", *Oxford University Press*, New York, 1999.

[2] S. Javaid, M.J. Khan, A. Rehman, and A. Al-Fuqaha, "Communication and control in collaborative UAVs: Recent advances and future trends", *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, pp. 5719–5739, 2023.

[3] J. Guo, Y. Liu, Z. Wang, and H. Zhang, "Collision-free formation tracking control for multiple quadrotors under switching directed topologies: Theory and experiment", *Aerospace Science and Technology*, vol. 131, Article 108007, 2022.

[4] Y. Chen, Y. Zhang, and Z. Wang, "Trajectory optimization of multiple quad-rotor UAVs in collaborative assembling task", *Chinese Journal of Aeronautics*, vol. 29, pp. 184–201, 2016.

[5] S. Ghambari, A.H. Gandomi, and A. Dehghan, "UAV path planning techniques: A survey", *RAIRO–Operations Research*, vol. 58, pp. 2951–2989, 2024.

[6] J. Tang, H. Duan, and S. Lao, "Swarm intelligence algorithms for multiple unmanned aerial vehicles collaboration: A comprehensive review", *Artificial Intelligence Review*, vol. 56, no. 4, pp. 4295–4327, 2023.

[7] V.B. Priezzhev, D. Dhar, A. Dhar, and S. Krishnamurthy, "Eulerian walkers as a model of self-organized criticality", *Physical Review Letters*, vol. 77, pp. 5079–5082, 1996.

[8] A.E. Holroyd, L. Levine, K. Mészáros, Y. Peres, J. Propp, and D. Wilson, "Chip-firing and rotor-routing on directed graphs", in *In and Out of Equilibrium 2*, Progress in Probability, vol. 60, Birkhäuser, Basel, pp. 331–364, 2008.

[9] V.S. Poghosyan and V.B. Priezzhev, "Euler tours and unicycles in the rotor-router model", *Journal of Statistical Mechanics: Theory and Experiment*, P06003, 2014.

[10] V.V. Papoyan, V.S. Poghosyan, and V.B. Priezzhev, "Loop reversibility and subdiffusion of the rotor-router walk", *Journal of Physics A: Mathematical and Theoretical*, vol. 48, no. 28, 285203, 2015.

[11] V.V. Papoyan, V.S. Poghosyan, and V.B. Priezzhev, "Spiral structures in the rotor-router walk", *Journal of Statistical Mechanics: Theory and Experiment*, 043207, 2016.

[12] S. Poghosyan, V. Poghosyan, S. Abrahamyan, A. Lazyan, H. Astsatryan, and Y. Alaverdyan, "Cloud-based mathematical models for self-organizing swarms of UAVs: Design and analysis", *Drone Systems & Applications*, vol. 12, pp. 1–12, 2024.

[13] V. Poghosyan, S. Poghosyan, A. Lazyan, A. Atashyan, D. Hayrapetyan, and H. Astsatryan, "Self-organizing multi-user UAV swarm simulation platform", *Programming and Computer Software*, vol. 49, no. 1, pp. S7–S15, 2023.

[14] A. Atashyan, A. Lazyan, D. Hayrapetyan, H. Astsatryan, V. Poghosyan, and S. Poghosyan, "Mission preparation for self-organizing UAV swarms on a multi-user platform", *Programming and Computer Software*, vol. 50, no. 1, pp. S39–S46, 2024.

[15] S. Poghosyan, V. Poghosyan, H. Astsatryan, and Ye. Alaverdyan, "Decentralized privacy provision in dynamically reconfigurable and self-organizing aero-physical cyber systems", *Physics of Particles and Nuclei*, vol. 56, no. 6, pp. 1309–1317, 2025.

[16] "Self-Organized UAV Swarm Demo", Online resource, December 2025. Available: https://www.youtube.com/watch?v=HvQkU4Wotjg

Տարածքի դետերմինացված ծածկումը ԱԹՍ երամի կողմից՝ բազմամասնիկային Էյլերյան rotor-router դեգերմամբ

Վահագն Ս. Պողոսյան

ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ,
Երևան, Հայաստան
e-mail: povahagn@gmail.com

Ամփոփում

Ուսումնասիրվել են տարածքի դետերմինացված ծածկման խնդիրները ԱԹՍ երամների համար՝ հիմնված բազմամասնիկային Էյլերյան rotor-router դեգերումների վրա:

Rotor-router մոդելը հանդիսանում է պատահական դեգերումների քվազի-պատահական անալոգը՝ օժտված հստակ կառուցվածքային հատկություններով:

Կիրառվել է կոնտուրի շրջելիության թեորեմը, ըստ որի՝ դիսկրետացված միջավայրի կապակցվածության որոշակի պայմաններում, համակարգված բազմամասնիկային էյլերյան rotor-router դեգերումներ իրականացնող ԱԹՍ երամը ապահովում է տարածքի ամբողջական ծածկում՝ այն իմաստով, որ յուրաքանչյուր ուղղորդված կողով իրականացվում է ճիշտ մեկ անցում միայն մեկ ԱԹՍ-ի կողմից:

Ստացված արդյունքները սահմանում են ապացուցելի ծածկման երաշխիքներ, որոնք կայուն են ասինխրոն կատարման նկատմամբ:

Կոնտուրի շրջելիության թեորեմը ընդհանրացվել է բազմակի ծածկման դեպքի համար, երբ պահանջվում է տարածքը ծածկել նախապես տրված M անգամ՝ առանց յուրաքանչյուր դեգերման ցիկլի ավարտին ԱԹՍ-ների միջև գլոբալ սինքրոնացման անհրաժեշտության:

Մշակվել է բազմաօգտատեր և բազմաագենտ հարթակ, որն ապահովում է ինքնակազմակերպվող ԱԹՍ երամների առաքելությունների պլանավորումը և համակարգված վերահսկումը:

Բանալի բառեր՝ ԱԹՍ խումբ, rotor-router մոդել, էյլերյան քայլ, տարածքի հսկում, ինքնակազմակերպվող համակարգ:

Детерминированное покрытие области роем БПЛА с использованием многочастичных эйлеровых rotor-router блужданий

Ваагн С. Погосян

Институт проблем информатики и автоматизации НАН РА, Ереван, Армения
e-mail: povahagn@gmail.com

Аннотация

Исследованы задачи детерминированного покрытия области роями беспилотных летательных аппаратов (БПЛА), основанные на многочастичных эйлеровых rotor-router блужданиях.

Модель rotor-router представляет собой квазислучайный аналог случайных блужданий, обладающий выраженными структурными свойствами.

Использована теорема обратимости контуров, согласно которой при определенных условиях связности дискретизированной среды рой БПЛА, выполняющий согласованные многочастичные эйлеровы rotor-router блуждания, обеспечивает полное покрытие области в том смысле, что для каждого ориентированного ребра осуществляется ровно один проход только одним БПЛА.

Полученные результаты устанавливают доказуемые гарантии покрытия, устойчивые к асинхронному выполнению. Кроме того, теорема обратимости контуров обобщается на случай многократного покрытия, когда требуется покрыть область заданное число M раз без необходимости глобальной синхронизации между БПЛА в конце каждого цикла блуждания.

Разработана многопользовательская и многоагентная платформа, обеспечивающая планирование миссий и координированный надзор за самоорганизующимися роями БПЛА.

Ключевые слова: рой БПЛА, модель rotor-router, эйлерово блуждание, наблюдение территории, самоорганизующаяся система.

Mathematical Problems of Computer Science 65, 16–29, 2026.
doi:10.51408/1963-0143

UDC 519.72

Rate-Reliability-Distortion-Equivocation Tradeoffs for Source with Secret Component and Side Information

Jemma S. Santrosyan

Vanadzor State University, Vanadzor, Armenia
e-mail: j.santrosian@gmail.com

Abstract

We consider the lossy source coding problem for one-way sources with correlated outputs and side information available at both the encoder and decoder. In this scenario, one component of the source must be transmitted to the receiver within a prescribed distortion level, while the other component must remain as confidential as possible from the receiver or a potential wiretapper. To characterize this trade-off, we introduce and analyze the rate-reliability-distortion-equivocation (RRDE) function, as well as the corresponding equivocation-reliability-distortion and rate-reliability-distortion functions. The results provide a unified information-theoretic framework that captures the interplay between compression efficiency, reconstruction fidelity, reliability, and secrecy.

Keywords: Rate-reliability-distortion-equivocation function, Lossy source coding, Data compression, Source coding with side information, Rate-distortion theory.
Article info: Received 27 March 2026; sent for review 7 April 2026; accepted 27 April 2026.

1. Introduction

Lossy source coding (also known as lossy compression) is a fundamental concept in information theory that deals with representing a source (e.g., an image, audio signal, or text) using fewer bits than the original, allowing some distortion or loss of information that is acceptable according to a given fidelity criterion [1].

A central concept in the theory of lossy source coding is the rate-distortion (RD) function, which characterizes the minimum coding rate required to satisfy a prescribed average distortion constraint. The RD function captures the fundamental trade-off between compression efficiency and reconstruction fidelity: as the allowable distortion increases, the required rate decreases, whereas stricter fidelity requirements necessitate higher transmission rates. These theoretical foundations originate in Shannon's seminal work [1], and were later developed extensively through the classical contributions of Slepian and Wolf [2], Wyner [3], and Wyner-Ziv [4]. Classical image and audio compression standards such as JPEG and MP3 exemplify the practical importance of the rate-distortion trade-off.

The presence of side information, namely additional data correlated with the source and available at the encoder, the decoder, or both, substantially enriches the source coding problem. By exploiting the statistical dependencies between the source and the side information, one can achieve more efficient compression or improved reconstruction quality. Source coding models with side information have found widespread applications in modern systems, including distributed sensor networks, multimedia communications, cooperative and relay-based wireless networks, privacy-preserving data compression and storage, as well as machine learning applications such as federated learning and distributed inference.

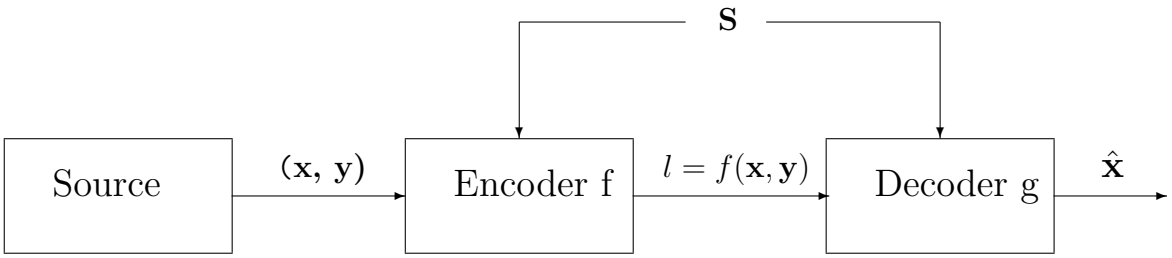

Fig. 1. One-way communication with correlated sources and common side information

Beyond the classical rate-distortion criterion, reliability plays an important role in source coding. Reliability quantifies the rate at which the decoding error probability decays as the block length increases. This leads to the notion of the rate-reliability-distortion (RRD) function, which describes the coding rate as a function of both the distortion level and a prescribed error exponent, or reliability parameter. This function has been investigated for various source models [5, 6].

In many practical scenarios, security is also a critical requirement. Security constraints arise when part of the source information must be protected from unintended receivers. This setting was first formalized by Yamamoto [7], who introduced the rate-distortion-equivocation (RDE) framework, where secrecy is measured via equivocation. In such models, one component of the source is reconstructed subject to distortion constraints, while another component is kept as uncertain as possible at the receiver. Subsequent research has extended this approach to cascade and branching communication systems, as well as to alternative security criteria based on distortion [8, 9, 10].

In our previous works, we studied the RRD and RDE functions. In particular, [11] analyzed the rate-reliability-distortion trade-off for correlated sources, while [12] introduced and investigated a secure source coding model in which secrecy is quantified using equivocation.

Building on these foundations and our previous work, this paper develops a unified framework that simultaneously accounts for rate, distortion, reliability, and secrecy. Specifically, we introduce the rate-reliability-distortion-equivocation (RRDE) function for a source coding model with a secret component and side information available at both the encoder and decoder. We characterize the set of achievable performance tuples and derive the corresponding equivocation-reliability-distortion and rate-reliability-distortion functions as special cases. This framework is particularly relevant for modern applications requiring secure and reliable data transmission, including distributed IoT networks, multimedia streaming with confidentiality constraints, secure communication protocols, and utility-privacy trade-off problems in data analysis [13]. By jointly accounting for side information, reliability, and secrecy, this work bridges classical source coding theory and contemporary challenges in secure communications.

The remainder of the paper is organized as follows. Section 2 introduces the main notations and definitions. The main results are presented in Section 3. The proofs are provided in the next section. Section 5 concludes the paper with a summary and discussion of possible extensions.

2. Notations and Definitions

The Discrete Memoryless Source (DMS) is defined as a sequence $\{(X_i, Y_i, S_i)\}_{i=1}^{\infty}$ of discrete independent identically distributed random variables (i. i. d.) X, Y, and S, taking values in finite sets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{S}$, which are the source messages alphabets, respectively. Let

$$P^* = \{P^*(x, y, s), x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}\}$$

be the generating probability distribution of the source outputs (X, Y, S). The source is memoryless, which means that for N-length vector pairs $\mathbf{x} = (x_1, x_2, ..., x_N) \in \mathcal{X}^N$,$\mathbf{y} = (y_1, y_2, ..., y_N) \in \mathcal{Y}^N$ and $\mathbf{s} = (s_1, s_2, ..., s_N) \in \mathcal{S}^N$

$$P^{*N}(\mathbf{x}, \mathbf{y}, \mathbf{s}) = \prod_{n=1}^{N} P^*(x, y, s).$$

The finite set $\hat{\mathcal{X}}$, different in general from $\mathcal{X}$, is the reproduction alphabet at the receiver. We will use the following distributions:

$$P = \{P(x, y, s) = P_0(s)P_1(x, y|s), x \in \mathcal{X}, y \in \mathcal{Y}, s \in \mathcal{S}\}$$

and the conditional distribution Q

$$Q = \{Q(\hat{x}|x, y, s), x \in \mathcal{X}, y \in \mathcal{Y}, \hat{x} \in \hat{X}, s \in \mathcal{S}\}.$$

A code (f_N, g_N) is defined by a pair of mappings: a coding

$$f_N : \mathcal{X}^N \times \mathcal{Y}^N \times \mathcal{S}^N \to \{1, 2, ..., L(N)\},$$

and decoding

$$g_N : \{1, 2, ..., L(N)\} \times \mathcal{S}^N \to \hat{\mathcal{X}},$$

where $L(N)$ is the code volume. Code rate is

$$R(f_N, g_N) = \frac{1}{N} \log L(N).$$

Throughout this paper, all log-s and exp-s are of base 2.

We consider the distortion measure

$$d : \mathcal{X} \times \hat{\mathcal{X}} \to [0; \infty)$$

between source and reconstruction messages. The distortion measure for N -length sequences is the average of the components' distortions

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N} \sum_{n=1}^{N} d(x, \hat{x}).$$

The task of this system is to ensure restoration of one of the components of source messages, i.e. X, at the receiver within a given distortion level Δ_d and with a small error probability in the case when the state sequence is available to both the encoder and the decoder. At the same time, the other source output Y should be kept as secret as possible from the receiver or wiretapper. This protection level is measured by the **equivocation rate**, defined as

$$R_e = \frac{1}{N} H(\mathbf{Y}|L(N), S),$$

where $H(\mathbf{Y}|L(N), S)$ is the conditional entropy [14]. In other words, the equivocation rate indicates the receiver's uncertainty about $\mathbf{y}$ given l and $\mathbf{s}$.

For the formulation of the result, we remind the following definitions [14].

The **entropy** of RV S is

$$H_{P_0}(S) = -\sum_{s\in\mathcal{S}} P_0(s)\log P_0(s).$$

The **conditional entropy** of RV X, Y relative to the random variable S is

$$H_P(X,Y|S) = -\sum_{x\in\mathcal{X},\, y\in\mathcal{Y},\, s\in\mathcal{S}} P(x,y,s)\log P_1(x,y|s).$$

The **conditional mutual information** of RV X, Y and $\hat{X}$ relative to the random variable S is

$$I_{P,Q}(X,Y;\hat{X}|S)$$

$$= \sum_{x\in\mathcal{X},\, y\in\mathcal{Y},\, \hat{x}\in\hat{\mathcal{X}},s\in\mathcal{S}} P(x,y,s)Q(\hat{x}|x,y,s)\log\frac{Q(\hat{x}|x,y,s)}{PQ(\hat{x}|s)},$$

where

$$PQ(\hat{x}|s) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P_1(x,y|s)Q(\hat{x}|x,y,s).$$

The following property will be used below:

$$I_{P,Q}(X,Y;\hat{X}|S) = H_P(X,Y|S) - H_{P,Q}(X,Y|\hat{X},S).$$

The **Kullback-Leibler divergence** between the distributions P and P^* is defined as follows:

$$D(P||P^*) = \sum_{x\in\mathcal{X},y\in\mathcal{Y},s\in\mathcal{S}} P(x,y,s)\log\frac{P(x,y,s)}{P^*(x,y,s)}.$$

We define the **error probability** of the code (f_N, g_N) as

$$e(f_N, g_N, P^*, \Delta_d) = 1 - \min_{\mathbf{s}\in\mathcal{S}^N} P^{*N}(\mathcal{A}(\mathbf{s})),$$

where $\mathcal{A}(\mathbf{s})$ is the set of satisfactorily transmitted vectors for a given $\mathbf{s}$, which are reconstructed within the distortion constraint $\Delta_d \geq 0$:

$$\mathcal{A}(\mathbf{s}) = \{(\mathbf{x},\mathbf{y}) : g_N(f_N(\mathbf{x},\mathbf{y},\mathbf{s}),\mathbf{s}) = \hat{\mathbf{x}}, d(\mathbf{x},\hat{\mathbf{x}}) \leq \Delta_d\}.$$

Definition 1. The triple (R, Δ_d, Δ_e) is called E-achievable for given $P^*, E > 0, \Delta_d \geq 0, \Delta_e \geq 0$, if for every $\epsilon > 0, \delta > 0$, there exists a code (f_N, g_N) such that

$$\frac{1}{N} \log L(N) \leq R + \epsilon,$$

the error probability is exponentially small

$$e(f_N, g_N, P^*, \Delta_d) \leq \exp\{-N(E - \delta)\}$$

and the equivocation rate

$$R_e \geq \Delta_e - \epsilon.$$

We denote by $\mathcal{R}^*(E)$ the set of all **E-achievable triples**. We will consider the **distortion-equivocation E-achievable region**:

$$\mathcal{R}^*_{\Delta_d, \Delta_e}(E) = \{(\Delta_d, \Delta_e) : (R, \Delta_d, \Delta_e) \in \mathcal{R}^*(E) \text{ for some } R \geq 0\}.$$

Then the **RRDE function** is defined as

$$R^*(E, \Delta_d, \Delta_e) = \min_{(R, \Delta_d, \Delta_e) \in \mathcal{R}^*(E)} R.$$

At last, the **equivocation-reliability-distortion function** (ERD) is:

$$\Gamma^*(E, \Delta_d) = \max_{(\Delta_d, \Delta_e) \in \mathcal{R}^*_{\Delta_d, \Delta_e}(E)} \Delta_e.$$

3. Formulation of Results

Consider the following set of distributions P on $\mathcal{X} \times \mathcal{Y} \times \mathcal{S}$:

$$\alpha(E, P^*) = \{P : D(P||P^*) \leq E\}.$$

Let $\mathcal{Q}(P, \Delta_d, \Delta_e)$ be the set of all conditional PDs $Q_P(\hat{x}|x, y, s) = Q_P$, corresponding to the PD P, for which the following conditions hold:

$$\mathbf{E}d(X, \hat{X}) = \sum_{x,y,s,\hat{x}} P(x, y, s) Q_P(\hat{x}|x, y, s) d(x, \hat{x}) \leq \Delta_d, \tag{1}$$

$$H_{P,Q_P}(Y|\hat{X}, S) \geq \Delta_e.$$

Then

$$\mathcal{Q}(P, \Delta_d) = \bigcup_{H_{P,Q_P}(Y|X,S) \leq \Delta_e \leq H_{P,Q_P}(Y|S)} \mathcal{Q}(P, \Delta_d, \Delta_e).$$

The main result of this paper is presented in the following theorem.

Theorem 1. *For given source distribution P^*, every $E > 0$,*

$$\mathcal{R}^*(E) = \left\{ \begin{array}{l} (R, \Delta_d, \Delta_e) : \Delta_d \geq 0, \Delta_e \geq 0, \\ \\ 0 \leq R_e \leq \min\limits_{P \in \alpha(E, P^*)} \max\limits_{Q_P \in \mathcal{Q}(P, \Delta_d)} H_{P,Q_P}(Y, |\hat{X}, S), \\ \\ R \geq \max\limits_{P \in \alpha(E, P^*)} \min\limits_{Q_P \in \mathcal{Q}(P, \Delta_d, \Delta_e)} I_{P,Q_P}(X, Y; \hat{X}|S) \end{array} \right\}.$$

This result characterizes the fundamental trade-off between rate, distortion, reliability, and secrecy in the presence of side information. The proof is given in the Appendix, based on the method of types [15].

Corollary 2. *The ERD function equals*

$$\Gamma^*(E,\Delta_d)=\min_{P\in\alpha(E,P^*)}\max_{Q_P\in\mathcal{Q}(P,\Delta_d)}H_{P,Q_P}(Y|\hat{X},S).$$

Corollary 3.

$$\mathcal{R}^*_{\Delta_d,\Delta_e}(E)=\left\{\begin{array}{l} R(E,\Delta_d,\Delta_e):\Delta_d\geq 0,\\ \\ 0\leq\Delta_e\leq\Gamma^*(E,\Delta_d)\end{array}\right\}.$$

Corollary 4. *The RRDE function equals*

$$R^*(E,\Delta_d,\Delta_e)=\max_{P\in\alpha(E,P^*)}\min_{Q_P\in\mathcal{Q}(P,\Delta_d,\Delta_e)}I_{P,Q_P}(X,Y;\hat{X}|S).$$

Corollary 5. *The limits of the RRDE and ERD functions when E tends to 0, coincide with the RDR and ED functions stated in [7]:*

$$\lim_{E\to 0}R^*(E,\Delta_d,\Delta_e)=R^*(\Delta_d,\Delta_e)=\min_{Q^*_P\in\mathcal{Q}(P^*,\Delta_d,\Delta_e)}I_{P^*,Q^*_P}(X,Y;\hat{X}|S).$$

$$\lim_{E\to 0}\Gamma^*(E,\Delta_d)=\Gamma^*(\Delta_d)=\max_{Q_{P^*}\in\mathcal{Q}(P^*,\Delta_d)}H_{P^*,Q_{P^*}}(Y|\hat{X},S).$$

Corollary 6. *In the absence of a secret component, the achievable region coincides with [12]*

$$\mathcal{R}^*(E)=\left\{\begin{array}{l}(R,\Delta_d,\Delta_e):\Delta_d\geq 0,\Delta_e\geq 0,\\ \\ 0\leq R_e\leq\min\limits_{P\in\alpha(E,P^*)}\max\limits_{Q_P\in\mathcal{Q}(P,\Delta_d)}H_{P,Q_P}(Y|\hat{X}),\\ \\ R\geq\max\limits_{P\in\alpha(E,P^*)}\min\limits_{Q_P\in\mathcal{Q}(P,\Delta_d,\Delta_e)}I_{P,Q_P}(X,Y;\hat{X})\end{array}\right\}.$$

Corollary 7. *In the absence of side information at the decoder the result reduces to [11]*

$$R(E,\Delta,P^*)=\max_{P\in\alpha(E,P^*)}\min_{Q_P\in\mathcal{Q}(P,\Delta)}I_{P,Q_P}(X;\hat{X}|S).$$

For the proof of Theorem 1, we will use the following modification of the Covering Lemma [16], [5].

Lemma 1. *Let for $\epsilon>0$*

$$J(P,Q)=\exp\{N(I_{P,Q}(X,Y;\hat{X}|S)+\epsilon)\}.$$

Then, for every type P_0 state sequence $\mathbf{s}\in\mathcal{T}^N_{P_0}(S)$, conditional types P_1 and Q, there exists a collection of vectors

$$\{\hat{\mathbf{x}}_j\in\mathcal{T}^N_{P,Q}(\hat{X}|\mathbf{s}),j=1,...,J(P,Q)\},$$

such that the set

$$\{\mathcal{T}^N_{P,Q}(X,Y|\hat{\mathbf{x}}_j,\mathbf{s}),j=1,...,J(P,Q)\},$$

covers $\mathcal{T}^N_P(X,Y|\mathbf{s})$ for N large enough, that is

$$\mathcal{T}^N_P(X,Y|\mathbf{s})\subset\bigcup_{j=1}^{J(P,Q)}\mathcal{T}^N_{P,Q}(X,Y|\hat{\mathbf{x}}_j,\mathbf{s}).$$

4. Proof of Theorem 1

The proof of Theorem 1 consists of two parts: achievability (direct part) and converse (inverse part).

4.1 Achievability

First we shall prove that any triple (R, Δ_d, Δ_e) satisfying the conditions of Theorem 1 is E-achievable or that

$$\mathcal{R}^*(E) \supseteq \left\{ \begin{array}{l} (R, \Delta_d, \Delta_e) : \Delta_d \geq 0, \Delta_e \geq 0, \\ \\ 0 \leq R_e \leq \min\limits_{P \in \alpha(E,P^*)} \max\limits_{Q_P \in \mathcal{Q}(P,\Delta_d)} H_{P,Q_P}(Y|\hat{X}, S), \\ \\ R \geq \max\limits_{P \in \alpha(E,P^*)} \min\limits_{Q_P \in \mathcal{Q}(P,\Delta_d,\Delta_e)} I_{P,Q_P}(X, Y; \hat{X}|S) \end{array} \right\}.$$

Step 1: Type Partitioning

Let us represent the set of all source messages of length N as follows:

$$\mathcal{X}^N \times \mathcal{Y}^N \times \mathcal{S}^N = \bigcup_{P \in \mathcal{P}^N(X \times Y \times S)} \mathcal{T}_P^N(X, Y, S) = \bigcup_{P_0 \in \mathcal{P}^N(S)} \bigcup_{P_1 \in \mathcal{P}^N(X \times Y, P_0)} \mathcal{T}_P^N(X, Y, S),$$

where $\mathcal{P}^N(X \times Y \times S)$ is the set of possible types P of triples $(\mathbf{x}, \mathbf{y}, \mathbf{s}) \in \mathcal{X}^{\mathbf{N}} \times \mathcal{Y}^{\mathbf{N}} \times \mathcal{S}^{\mathbf{N}}$, $\mathcal{P}^{\mathbf{N}}(\mathbf{S})$ is the set of possible types P_0 of vectors $\mathbf{s} \in \mathcal{S}$, $\mathcal{P}^N(X \times Y, P_0)$ is the set of conditional types P_1 of pairs $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^{\mathbf{N}} \times \mathcal{Y}^{\mathbf{N}}$, for given P_0.

Using the properties of types and the definition of the set $\alpha(E, P^*)$ for each $\delta > 0$, the probability of observing sequences with types outside $\alpha(E + \delta, P^*)$ satisfies:

$$\begin{aligned} P^{*N}\left(\bigcup_{P \notin \alpha(E+\delta,P^*)} \mathcal{T}_P^N(X, Y, S) \right) &= \sum_{P \notin \alpha(E+\delta,P^*)} P^{*N}\left(\mathcal{T}_P^N(X, Y, S) \right) \\ &\leq (N+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{S}|} \exp\left\{ -N \min_{P \notin \alpha(E+\delta,P^*)} D(P||P*) \right\} \quad (2) \\ &\leq \exp\left\{ -NE - N\delta + |\mathcal{X}||\mathcal{Y}||\mathcal{S}| \log(N+1) \right\} \\ &\leq \exp\left\{ -N(E + \delta/2) \right\}, \end{aligned}$$

for sufficiently large N.

Step 2: Code Construction

For each $\Delta_d \geq 0$, fix type $P \in \alpha(E + \delta, P^*)$ and some $Q_P \in \mathcal{Q}(P, \Delta_d, \Delta_e)$. Let for each $\mathbf{s} \in \mathcal{T}_{P_0}^N(S)$

$$C(P, Q_P, j, \mathbf{s}) = \mathcal{T}_{P,Q_P}^N(X, Y|\hat{\mathbf{x}}_j, \mathbf{s}) - \bigcup_{j' < j} \mathcal{T}_{P,Q_P}^N(X, Y|\hat{\mathbf{x}}_{j'}, \mathbf{s}), \quad j = \overline{1, J(P, Q_P)}.$$

Define a code (f_N, g_N) for each $\mathbf{s} \in \mathcal{T}_{P_0}^N(S)$ and vector pairs of conditional type P_1 with encoding as:

$$f_N(\mathbf{x}, \mathbf{y}|\mathbf{s}) = \left\{ \begin{array}{l} j, \ \text{ when } (\mathbf{x}, \mathbf{y}) \in C(P, Q_P, j, \mathbf{s}), \ P \in \alpha(E + \delta, P^*), \\ \\ j_0, \ \text{ when } (\mathbf{x}, \mathbf{y}) \in \mathcal{T}_P^N(X, Y|\mathbf{s}), \ P \notin \alpha(E + \delta, P^*), \end{array} \right.$$

and decoding

$$g_N(j|\mathbf{s}) = \hat{\mathbf{x}}_j, \qquad g_N(j_0|\mathbf{s}) = \hat{\mathbf{x}}_0,$$

where the number j_0 and the reconstruction vector $\hat{\mathbf{x}}_0$ are fixed. Obviously, with such a code, an error occurs only when the number j_0 is sent.

Step 3: Distortion Constraint

According to the definition of the code and the inequality (1), for $P \in \alpha(E+\delta, P^*)$ and $Q_P \in \mathcal{Q}(P, \Delta_d, \Delta_e)$ we have:

$$\begin{aligned} d(\mathbf{x}, \hat{\mathbf{x}}_\mathbf{j}) &= \sum_{x,y,\hat{x},s} P(x,y,s) Q_P(\hat{x}|x,y,s) d(x,\hat{x}) \\ &= \mathbf{E}_{P,Q_P} d(X,\hat{X}) \le \Delta_d, \quad j = \overline{1, J(P,Q_P)}, \end{aligned}$$

which ensures that the distortion constraint is satisfied.

Step 4: Rate Analysis

According to Lemma 1, the number of vectors $\hat{\mathbf{x}}$, each $\mathbf{s}$ for type P and corresponding conditional type $Q_P \in \mathcal{Q}(P, \Delta_d, \Delta_e)$ is:

$$L_{P,Q_P}(N) = \exp\left\{N(I_{P,Q_P}(X,Y;\hat{X}|S) + \epsilon)\right\}.$$

Then, taking into account that the number of types has a polynomial estimate [15]

$$\begin{aligned} L(N) &\le \sum_{P \in \alpha(E+\delta,P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta_d,\Delta_e)} L_{P,Q_P}(N) \\ &\le (N+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{S}|} \max_{P \in \alpha(E+\delta,P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta_d,\Delta_e)} \exp\left\{N(I_{P,Q_P}(X,Y;\hat{X}|S) + \epsilon)\right\}. \end{aligned}$$

Hence, the corresponding limit for the transmission rate is:

$$\begin{aligned} &\frac{1}{N}\log L_{P,Q_P}(N) - \epsilon - \frac{1}{N}|\mathcal{X}||\mathcal{Y}||\mathcal{S}|\log(N+1) \\ &\le \max_{P \in \alpha(E+\delta,P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta_d,\Delta_e)} I_{P,Q_P}(X,Y;\hat{X}|S). \end{aligned} \tag{3}$$

Taking into account the arbitrariness of ϵ and δ and the continuity of the information expression (3), we get:

$$R^*(E,\Delta_d,\Delta_e) \le \max_{P \in \alpha(E,P^*)} \min_{Q_P \in \mathcal{Q}(P,\Delta_d,\Delta_e)} I_{P,Q_P}(X,Y;\hat{X}|S). \tag{4}$$

Step 5: Equivocation Analysis

Using type properties and entropy bounds, for this code the equivocation rate can be evaluated as follows:

$$\begin{aligned} &\frac{1}{N} H(\mathbf{Y}|L(N),S) \\ &\ge \sum_{\mathbf{s}} P_0(\mathbf{s}) \frac{1}{N} \sum_{j=1}^{L(N)} H_{P^*,Q_{P^*}}(Y|(\mathbf{x},\mathbf{y}) \in C(P,Q_P,j,\mathbf{s}))) P^*\{(\mathbf{x},\mathbf{y}) \in C(P,Q_P,j,\mathbf{s})\} \\ &= \sum_{\mathbf{s}} P_0(\mathbf{s}) \frac{1}{N} \sum_{j=1}^{L(N)} \end{aligned} \tag{5}$$

$$\left[-\sum_{\mathbf{y}:\mathbf{x},\mathbf{y}\in C(P,Q_P,j,\mathbf{s})} P^*\{\mathbf{y}|(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})\}\log P^*\{\mathbf{y}|(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})\}\right]$$
$$\times P^*\{(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})\}.$$

For any $\mathbf{y}$ such that $\mathbf{x},\mathbf{y}\in C(P,Q_P,j,\mathbf{s})$ for some $\mathbf{x}$

$$P^*\{\mathbf{y}|(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})\} = \frac{P^*\{(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})|\mathbf{y}\}P^*\{\mathbf{y}\}}{P^*\{(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})\}}$$

$$= \frac{\sum\limits_{(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})} P^*\{\mathbf{x},\mathbf{y}|\mathbf{y},\mathbf{s}\}P^*\{\mathbf{y}\}}{\sum\limits_{(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})} P^*\{\mathbf{x},\mathbf{y}|\mathbf{s}\}} \le \frac{\sum\limits_{\mathbf{x}\in\mathcal{T}^N_{P,Q_P}(X|\mathbf{y},\hat{\mathbf{x}}_j,\mathbf{s})} P^*\{\mathbf{x}|\mathbf{y},\mathbf{s}\}P^*\{\mathbf{y}\}}{\sum\limits_{(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})} P^*\{\mathbf{x},\mathbf{y}|\mathbf{s}\}}. \quad (6)$$

As the probability of the pair $(\mathbf{x},\mathbf{y})$ is constant within the same type, from (6) we obtain that

$$P^*\{\mathbf{y}|(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})\} \le \frac{|\mathcal{T}^N_{P,Q_P}(X|\mathbf{y},\hat{\mathbf{x}}_j,\mathbf{s})|}{|C(P,Q_P,j,\mathbf{s})|}$$

$$\le \frac{\exp[N(H_{P,Q_P}(X|Y,\hat{X},S)]}{(N+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{S}|}\exp[N(H_{P,Q_P}(X,Y|\hat{X},S)]} \le \exp[-N(H_{P,Q_P}(Y|\hat{X},S)-\epsilon)]. \quad (7)$$

Then, from (5), (7) and (2) we obtain that

$$\frac{1}{N}H(\mathbf{Y}|L(N),S)$$

$$\ge \sum_{\mathbf{s}} P_0(\mathbf{s})\frac{1}{N}\sum_{j=1}^{L(N)}\left[N\sum_{\mathbf{y}:(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})} P^*\{\mathbf{y}|(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})\}(H_{P,Q_P}(Y|\hat{X},S)-\epsilon)\right]$$
$$\times P^*\{(\mathbf{x},\mathbf{y})\in C(P,Q_P,j,\mathbf{s})\}$$
$$= \sum_{\mathbf{s}} P_0(\mathbf{s})P^*\{(\mathbf{x},\mathbf{y})\in \bigcup_{j=1}^{L(N)} C(P,Q_P,j,\mathbf{s})\}(H_{P,Q_P}(Y|\hat{X},S)-\epsilon)$$
$$\ge (1-\exp\{-N(E+\delta/2)\})(H_{P,Q_P}(Y|\hat{X},S)-\epsilon).$$

For N large enough, we obtain that

$$R_e \ge H_{P,Q_P}(Y|\hat{X},S)-\epsilon \ge \Delta_e-\epsilon. \quad (8)$$

According to (2), (4) and (8), we state that the triple (R,Δ_d,Δ_e) is E-achievable.

4.2 Converse

Now we show that any E-achievable triple satisfies:

$$\mathcal{R}^*(E)\subseteq\left\{\begin{array}{l}(R,\Delta_d,\Delta_e):\Delta_d\ge 0,\Delta_e\ge 0,\\ \\ 0\le R_e\le \min\limits_{P\in\alpha(E,P^*)}\max\limits_{Q_P\in\mathcal{Q}(P,\Delta_d)} H_{P,Q_P}(Y|\hat{X},S),\\ \\ R\ge \max\limits_{P\in\alpha(E,P^*)}\min\limits_{Q_P\in\mathcal{Q}(P,\Delta_d,\Delta_e)} I_{P,Q_P}(X,Y;\hat{X}|S)\end{array}\right\}.$$

Step 1: Setup

Let $\epsilon > 0$ be fixed. Consider a code (f_N, g_N) for each blocklength N with (R, Δ_d, Δ_e) E-achievable triple. We must show that for some $Q_P \in \mathcal{Q}(P, \Delta_d, \Delta_e)$ the following inequalities hold for N large enough:

$$\frac{1}{N}\log L(N) + \epsilon \geq \max_{P\in\alpha(E,P^*)} I_{P,Q_P}(X,Y;\hat{X}|S), \tag{9}$$

$$\frac{1}{N}H(\mathbf{Y}|L(N),S) - \epsilon \leq \min_{P\in\alpha(E,P^*)} H_{P,Q_P}(Y|\hat{X},S). \tag{10}$$

Step 2: Size of Typical Sets

Let $\mathcal{A}'(\mathbf{s})$ be the complement of the set $\mathcal{A}(\mathbf{s})$. The following statement is true:

$$\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right| = \left|\mathcal{T}_P^N(X,Y|\mathbf{s})\right| - \left|\mathcal{A}'(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right|.$$

For $P \in \alpha(E-\epsilon, P^*)$

$$\begin{aligned}\left|\mathcal{A}'(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right| &= \frac{P^{*N}\left\{\mathcal{A}'(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right\}}{P^{*N}(\mathbf{x},\mathbf{y}|\mathbf{s})}\\ &\leq \exp\left\{N(H_P(X,Y|S)+D(P||P^*))\right\}\exp\left\{-N(E-\epsilon)\right\}\\ &\leq \exp\left\{N(H_P(X,Y|S)-\epsilon)\right\}.\end{aligned}$$

Hence,

$$\begin{aligned}\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right| &\geq (N+1)^{-|\mathcal{X}||\mathcal{Y}||\mathcal{S}|}\exp\left\{NH_P(X,Y|S)\right\} - \exp\left\{N(H_P(X,Y|S)-\epsilon)\right\}\\ &= \exp\left\{N(H_P(X,Y|S)-\epsilon)\right\}\left(\frac{\exp\{N\epsilon\}}{(N+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{S}|}}-1\right) \qquad (11)\\ &\geq \exp\left\{N(H_P(X,Y|S)-\epsilon)\right\}.\end{aligned}$$

For each $\mathbf{s}\in\mathcal{T}_{P_0}^N(S)$ and $\mathbf{x},\mathbf{y}\in\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})$ corresponds a unique vector $\hat{\mathbf{x}}$ such that

$$\hat{\mathbf{x}} = g_N(f_N(\mathbf{x},\mathbf{y}|\mathbf{s}),\mathbf{s}) \ \text{ and } \ \hat{\mathbf{x}}\in\mathcal{T}_{P,Q}^N(\hat{X}|\mathbf{x},\mathbf{y},\mathbf{s}).$$

Step 3: Conditional Type Decomposition

Let us divide the set of all vectors $\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right|$ into subsets by conditional types Q_P. The class having maximum cardinality for given P, we denote by

$$\left(\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right|\right)_{Q_P}.$$

According to the number of conditional types Q, for sufficiently large N, we have:

$$\begin{aligned}\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right| &\leq (N+1)^{|\mathcal{X}||\mathcal{Y}||\mathcal{S}|}\left(\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right|\right)_{Q_P}\\ &\leq \exp\{N\epsilon/2\}\left(\left|\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right|\right)_{Q_P}. \qquad (12)\end{aligned}$$

Let

$$\mathcal{D}(\mathbf{s}) = \left\{\hat{\mathbf{x}} : g_N(f_N(\mathbf{x},\mathbf{y}|\mathbf{s}),\mathbf{s}) = \hat{\mathbf{x}}, \text{for some } (\mathbf{x},\mathbf{y})\in\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\bigcap\mathcal{T}_{P,Q_P}^N(X,Y|\hat{\mathbf{x}},\mathbf{s})\right\}.$$

From the definition of the code $|\mathcal{D}(\mathbf{s})| \leq L(N)$, then

$$\begin{aligned}\left|\left(\mathcal{A}(\mathbf{s})\bigcap\mathcal{T}_P^N(X,Y|\mathbf{s})\right)\right|_{Q_P} &\leq \sum_{\hat{\mathbf{x}}\in\mathcal{D}(\mathbf{s})}\left|\mathcal{T}_{P,Q}^N(X,Y|\hat{\mathbf{x}},\mathbf{s})\right| \\ &\leq L(N)\exp\{NH_{P,Q_P}(X,Y|\hat{X},S)\}. \end{aligned} \tag{13}$$

Step 4: Lower Bound on Rate
From (11-13) follows

$$L(N) \geq \exp\{N(I_{P,Q_P}(X,Y;\hat{X}|S)-\epsilon)\}$$

for each $P \in \alpha(E-\epsilon, P*)$ and some Q_P for which $\mathbf{E}_{P,Q_P}d(X,\hat{X}) \leq \Delta_d$, because $\mathbf{x},\mathbf{y} \in \mathcal{A}(\mathbf{s})$.

Step 5: Equivocation Upper Bound
From the achievability it follows that

$$\Delta_e - \epsilon \leq \frac{1}{N}H(\mathbf{Y}|L(N),S) \leq H_{P,Q_P}(Y|\hat{X},S).$$

Thus, $Q_P \in \mathcal{Q}(P,\Delta_d,\Delta_e)$ and inequalities (9) and (10) are valid. The Achievable triples must satisfy the same constraints as in Theorem 1, completing the proof of Converse part.

Theorem 1 is proved.

5. Conclusion

In this paper, we developed a unified information-theoretic framework for lossy source coding with side information, incorporating reliability and secrecy constraints. We introduced the notion of E-achievable triples and derived a complete characterization of the rate-reliability-distortion-equivocation (RRDE) region.

The obtained results explicitly quantify the interplay between compression efficiency, reconstruction fidelity, decoding reliability, and confidentiality. In particular, we derived closed-form expressions for the RRDE and ERD functions and demonstrated that they naturally generalize classical rate-distortion and rate-distortion-equivocation results.

The proposed framework is well-suited for modern applications involving secure and reliable data transmission, such as distributed sensing systems, privacy-preserving data compression, and communication networks with confidentiality constraints.

Future work may include extensions to multi-user scenarios, continuous alphabets, and alternative secrecy measures such as strong secrecy or semantic security.

References

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion", *IRE National Convention Record*, vol. 7, pp.142–163, 1959.

[2] D. Slepyan and J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Transactions on Information Theory*, vol. 19, no.6, pp. 471 – 480, 1973.

[3] A. D. Wyner, "On source coding with side information at the decoder", *IEEE Transactions on Information Theory*, vol. 21, no. 5, pp. 294 – 300, 1975.

[4] A. D. Wyner and J.Ziv, "The rate distortion function for source coding with side information at the decoder", *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1 – 10, 1976.

[5] E. Haroutunian, M. Haroutunian and A. Harutyunyan, "Reliability Criteria in Information Theory and in Statistical Hypothesis Testing", *Foundations and Trends in Communications and Information Theory*, vol. 4, nos 2-3, pp. 97–263, 2007. doi: 10.1561/0100000008

[6] E. A. Haroutunian, A. N. Harutyunyan and A. R. Ghazaryan, "On rate-reliability-distortion function for a robust descriptions system", *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2690-2697, Nov. 2000, doi: 10.1109/18.887883.

[7] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers", *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 918–923, 1983.

[8] H. Yamamoto, "Source coding theory for cascade and branching communication systems", *IEEE Transactions on Information Theory*, vol. 27, no. 3, pp. 299–308, 1981.

[9] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered", *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 835–842, 1988.

[10] H. Yamamoto, "Rate-distortion theory for the Shannon cipher system", *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 827–835, 1997.

[11] M. E. Haroutunian, J. S. Santrosyan and P. M. Hakobyan "Reliability criteria in source coding problem with secret component", *Mathematical Problems of Computer Science*, vol. 63, pp. 1424, 2025. doi:10.51408/1963-0128.

[12] M. Haroutunian, P. Hakobyan and J. Santrosyan, "Rate-reliability-distortion function for source with side information known to encoder and decoder", *Proceedings International Conference on Computer Science and Information Technologies*, pp. 155–158, 2025. doi: 10.51408/csit2025_38.

[13] L. Sankar, S. R. Rajagopalan and H. V. Poor, "Utility-privacy tradeoffs in databases: an information-theoretic approach", *IEEE Transactions on Information Theory*, vol. 8, no. 6, pp. 838–852, 2013.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition. Wiley, New York, 2006.

[15] I. Csiszár, "Method of types", *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505-2523, 1998.

[16] I. Csiszár and J.Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

Գաղտնի բաղադրիչ և կողմնակի տեղեկատվություն ունեցող աղբյուրների համար արագություն–հուսալիություն–շեղում–անորոշություն փոխզիջումների տեսական ուսումնասիրություն

Ջեմմա Ս. Սանթրոսյան

Վանաձորի պետական համալսարան, Վանաձոր, Հայաստան
e-mail: j.santrosian@gmail.com

Ամփոփում

Մենք դիտարկում ենք կորուստներով աղբյուրի կոդավորման խնդիրը փոխկապակցված ելքերով միակողմանի աղբյուրների համար, որտեղ կողմնակի ինֆորմացիան հասանելի է թե՛ կոդավորիչին, թե՛ ապակոդավորիչին: Այս սցենարում աղբյուրի բաղադրիչներից մեկը պետք է փոխանցվի հասցեատիրոջը՝ շեղման սահմանված մակարդակի շրջանակներում, մինչդեռ մյուս բաղադրիչը պետք է առավելագույնս գաղտնի մնա հասցեատիրոջից կամ հնարավոր գաղտնալսողից: Այս փոխզիջումը բնութագրելու համար մենք ներմուծում և վերլուծում ենք հաղորդման արագություն-հուսալիություն-շեղում-անորոշություն ֆունկցիան, ինչպես նաև համապատասխան անորոշություն-հուսալիություն-շեղում և արագություն-հուսալիություն-շեղում ֆունկցիաները: Արդյունքները տրամադրում են միասնական ինֆորմացիոն-տեսական հիմք, որն արտացոլում է սեղմման արդյունավետության, վերականգնման ճշգրտության, հուսալիության և գաղտնիության փոխազդեցությունը:

Բանալի բառեր՝ արագություն-հուսալիություն-շեղում-անորոշություն ֆունկցիա, կորուստներով աղբյուրի կոդավորում, տվյալների սեղմում, կողմնակի ինֆորմացիայով աղբյուրի կոդավորում, արագություն-շեղում տեսություն:

Исследование компромиссов между скоростью, надежностью, искажением и неопределённостью для источников с секретной компонентой и побочной информацией

Джемма С. Сантросян

Ванадзорский государственный университет, Ванадзор, Армения
e-mail: j.santrosian@gmail.com

Аннотация

В данной работе рассматривается задача кодирования источника с потерями для односторонних источников с коррелированными выходами при наличии побочной информации как на кодере, так и на декодере. В данном сценарии один компонент источника должен быть передан получателю с соблюдением заданного уровня искажения, в то время как другой компонент должен оставаться максимально конфиденциальным для получателя или потенциального перехватчика. Для количественной характеристики данного компромисса вводится и анализируется функция скорость-надежность-искажение-неопределенность, а также соответствующие функции неопределенность-надежность-искажение и скорость-надежность-искажение. Полученные результаты представляют собой единую информационно-теоретическую базу, описывающую взаимосвязь между эффективностью сжатия, точностью восстановления, надежностью и секретностью.

Ключевые слова: функция скорость-надежность-искажение-неопределенность, кодирование источника с потерями, сжатие данных, кодирование источника с побочной информацией, теория скорость-искажение.

Mathematical Problems of Computer Science 65, 30–48, 2026.
doi:10.51408/1963-0144

UDC 004.932:004.62:512.64

On the Theory and Application of Singular Value Decomposition in Image Processing and Data Analysis

Gayane G. Ghazaryan[1] and Artashes L. Ghazaryan[2]

[1]Institute of Physics, Yerevan State University, Yerevan, Armenia
[2] Technical University of Munich, Munich, Germany
e-mail: gayane.ghazaryan@ysu.am, artashes.ghazaryan17@gmail.com

Abstract

Singular Value Decomposition (SVD) is a fundamental factorization method central to image processing, dimensionality reduction, and numerical linear algebra. Its practical strength lies in truncating small singular values, thereby reducing storage while preserving the essential structure. The key challenge is choosing the truncation rank, which determines the trade-off between accuracy, efficiency, and stability.

We introduce an adaptive rank selection method that combines cumulative-energy thresholds with automated elbow detection, yielding a principled alternative to ad hoc rules. The framework is validated on image compression, denoising, and principal component analysis (PCA), with benchmarks against eigenvalue decomposition (EVD) and QR factorization. The results show that the adaptive rule improves fidelity and reproducibility, especially for noisy or ill-conditioned data.

All code, figures, and tables are released as open source, ensuring that the experiments can be reproduced in a clean environment. This unifies theory, automation, and reproducibility, reaffirming that SVD is both mathematically optimal and practically versatile for contemporary data analysis.
Keywords: Singular Value Decomposition, Adaptive rank selection, Cumulative energy, Elbow detection, Image denoising, Image compression, Principal component analysis.
Article info: Received 7 November 2025; sent for review 9 December 2025; accepted 25 May 2026.

1. Introduction

Singular Value Decomposition (SVD) is a cornerstone of numerical linear algebra and a central tool in data science, machine learning, and image processing [1-3]. Its optimality for low-rank approximation, established by the Eckart–Young–Mirsky theorem [1, 4], explains its broad use in image compression [5,6], noise reduction [7-10], and dimensionality reduction via principal component analysis (PCA) [11-13].

Despite its elegance, a persistent practical difficulty is the choice of truncation rank k. This parameter determines the balance between fidelity, compression ratio, and computational cost [14, 15]. Existing strategies rely either on cumulative-energy thresholds, which require task-specific tuning, or elbow rules, which tend to underestimate the rank when singular values decay gradually [16]. Consequently, applications often rely on ad hoc heuristics, which reduce reproducibility.

In this work, we propose an adaptive rank selection strategy that integrates cumulative energy and elbow detection within a unified SVD pipeline. Concretely, we adopt a conservative hybrid rule

$$k^\star = \max\big(k_\tau,\, k_{\text{elbow}}\big),$$

which guards against under-selection while preserving interpretability. Defaults such as $\tau \approx 0.995$ are reliable for clean images.

On standard images (e.g., `astronaut`, `cameraman`), the hybrid rule avoids elbow under-selection and maintains a higher structural similarity at comparable ranks (see Section 4, Fig. 2).

Positioning vs. alternatives. Classical fixed-rank tuning lacks reproducibility. Energy thresholds are interpretable but require task-dependent choices of τ. Elbow rules adapt automatically but tend to under-select when the singular values decay slowly. Denoising-specific criteria such as optimal hard thresholding [15] assume particular noise models and are not directly comparable across tasks, such as compression or PCA. In contrast, our hybrid rule is task-agnostic, reproducible, and guards against under-selection, without requiring prior calibration.

Contributions. The main contribution of this work is methodological and reproducibility-oriented.

- **Unified framework.** A reproducible SVD workflow for image compression, denoising, and PCA with standardized inputs and metrics (PSNR, SSIM, energy retention, and runtime).
- **Adaptive rank selection.** A principled hybrid of cumulative energy and elbow criteria, robust to slow spectral decay.
- **Systematic benchmarking.** Comparisons against EVD and QR factorization under identical conditions, clarifying the trade-offs in accuracy and runtime.
- **Open-source reproducibility.** All Python scripts can regenerate every figure and table from a clean environment using fixed random seeds, explicit data sources, and version-pinned dependencies.

Scope and limitations. We focus on standard truncated SVD for moderate-scale images and matrices; randomized or streaming variants are beyond the scope of this study.

2. Mathematical Foundations of Singular Value Decomposition

2.1 Definition and Existence

For any matrix $A \in \mathbb{R}^{m\times n}$ with rank $r \le \min(m,n)$, the **singular value decomposition (SVD)** is

$$A \;=\; U\,\Sigma\,V^\top,$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal, and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal (rectangular) with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ on the main diagonal and zeros elsewhere. The SVD exists for every real (and complex) matrix, regardless of its shape or rank [1,17]. The singular values are unique; the singular vectors are unique up to orthogonal transformations within degenerate singular subspaces (e.g., sign flips or rotations when some σ_i coincide) [18].[1]

2.2 Relation to Eigendecomposition

SVD is tied to the eigendecomposition of symmetric positive semidefinite matrices $A^\top A$ and $AA^\top$ [1,18]:

- Columns of V are eigenvectors of $A^\top A$ with eigenvalues $\lambda_i = \sigma_i^2$;
- Columns of U are eigenvectors of $AA^\top$ with the same nonzero eigenvalues;
- Hence $\sigma_i = \sqrt{\lambda_i(A^\top A)}$.

Consequently, $\{u_i\}$ and $\{v_i\}$ form orthonormal bases for the column and row spaces of A, respectively, and describe the directions of maximal variance/energy in the data representations.

2.3 Optimal Low-rank Approximation

Let $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^\top$ be the rank-k truncation. The Eckart–Young–Mirsky theorem [4,19] states that A_k is the *best* rank-k approximation to A in all unitarily invariant norms, in particular

$$\|A - A_k\|_F \;=\; \min_{\mathrm{rank(B)}\leq \mathrm{k}} \|A - B\|_F \;=\; \Big(\sum_{i=k+1}^{r} \sigma_i^2\Big)^{1/2}, \qquad \|A - A_k\|_2 \;=\; \begin{cases} \sigma_{k+1}, & k < r, \\ 0, & k \geq r. \end{cases}$$

These identities justify the use of truncated SVD for compression, denoising, and dimensionality reduction.

2.4 Compact and Truncated Forms

Two reduced forms are especially useful [1, 20]:

- **Compact (thin) SVD:** $A = U_r \Sigma_r V_r^\top$, with $U_r \in \mathbb{R}^{m\times r}$, $\Sigma_r \in \mathbb{R}^{r\times r}$ (nonzero singular values), $V_r \in \mathbb{R}^{n\times r}$. This is an exact factorization with minimal storage requirement.
- **Truncated SVD:** $A_k = U_k \Sigma_k V_k^\top$ for $k < r$, where $U_k \in \mathbb{R}^{m\times k}$, $\Sigma_k \in \mathbb{R}^{k\times k}$, $V_k \in \mathbb{R}^{n\times k}$. This trades accuracy for efficiency by retaining the dominant components.

Two useful projection identities are

$$A_k \;=\; U_k U_k^\top A \;=\; A V_k V_k^\top,$$

showing that truncation amounts to an orthogonal projection onto the leading left/right singular subspaces.

[1] For complex matrices, orthogonal/unitary and transpose/conjugate-transpose are replaced accordingly.

2.5 Energy, Variance, and Explained Proportion

SVD provides an orthogonally invariant notion of the total energy as follows:

$$\|A\|_F^2 = \sum_{i=1}^{r} \sigma_i^2.$$

For a rank-k truncation A_k, the retained fraction is

$$\eta_k := \frac{\|A_k\|_F^2}{\|A\|_F^2} = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{r} \sigma_i^2}, \qquad k = 1, \ldots, r.$$

This coincides with the explained variance in PCA for mean-centered data [11]. The quantity η_k underlies the cumulative-energy thresholds used later for adaptive rank selection and provides a task-agnostic basis for comparing datasets.

These properties establish SVD as both theoretically optimal and practically versatile. In the next section, we build on them to address the central challenge of choosing the truncation rank k, and propose an adaptive rule that avoids ad hoc heuristics.

3. Adaptive Rank Selection

Choosing the truncation rank k is the central practical challenge in applying SVD, as it sets the trade-off between the reconstruction fidelity, storage, and runtime. We propose a **unified adaptive rank selection** strategy that replaces ad hoc heuristics with principled and reproducible rules. The method combines two complementary criteria: a cumulative-energy threshold, which is standard in PCA practice [11], and automated elbow detection via the Kneedle algorithm [16]. Together, they yield stable choices for both sharply and slowly decaying singular spectra.

3.1 Cumulative-energy Criterion

As introduced in Subsection 2.5, the cumulative-energy ratio η_k measures the fraction of total energy retained by the first k singular values. The threshold rule selects the smallest k_τ such that $\eta_{k_\tau} \geq \tau$, with $\tau \in (0, 1)$. High thresholds (e.g., $\tau \approx 0.995$) preserve fine details in compression, whereas lower values may be acceptable when noise dominates the tail. The sensitivity to τ is summarized in Appendix B.

3.2 Elbow Detection on the Cumulative Curve

To improve robustness, we pair the threshold rule with elbow detection. We apply Kneedle [16] to the *cumulative* spectrum, using the normalized points $\big(k/r,\, \eta_k\big)$ after light smoothing (Savitzky–Golay) to reduce jitter. Kneedle returns an elbow index k_e corresponding to the onset of diminishing returns in cumulative energy. Thus, we obtain two candidates: k_τ from the threshold rule and k_e from the elbow detection. We round the continuous elbow location to the nearest integer $k \in \{1, \ldots, r\}$.

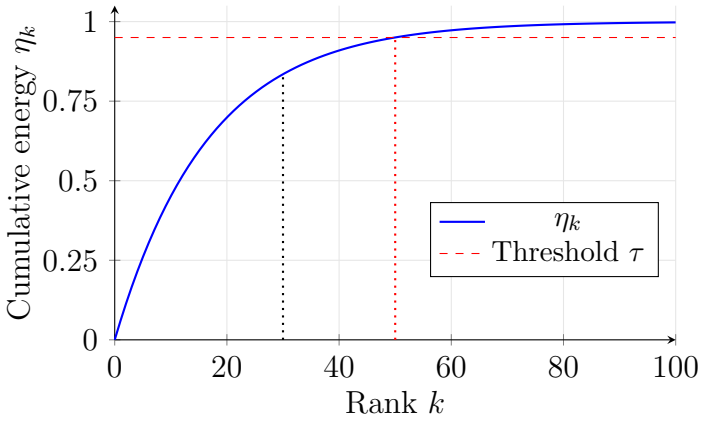

Fig. 1. Cumulative energy η_k with threshold k_τ and elbow k_e. The conservative policy $k^\star = \max(k_\tau, k_e)$ prevents under-selection.

3.3 Unified Policy and Scope

Fig. 1 illustrates the complementarity between the two rules. Our *default, conservative* choice is

$$k^\star \;=\; \max\big(k_\tau,\, k_e\big),$$

which preserves fidelity when the elbow under-selects slowly decaying spectra (Section 2). For practitioners who prefer stronger noise suppression at the cost of detail, we expose an *aggressive* option $k^\dagger = \min(k_\tau, k_e)$. Because the procedure depends only on the singular spectrum $\{\sigma_i\}$, it applies equally to exact and randomized SVD [12].

PCA interpretation. For a mean-centered data matrix $X \in \mathbb{R}^{m\times n}$ with covariance $C = \frac{1}{m-1}X^\top X$, let $s_i = \sigma_i(X)$ denote the singular values of X. Then the eigenvalues of C satisfy $\lambda_i = s_i^2/(m-1)$. The cumulative variance explained by the first k components is

$$\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{i=1}^{r}\lambda_i} = \frac{\sum_{i=1}^{k}s_i^2}{\sum_{i=1}^{r}s_i^2} = \eta_k(X).$$

Therefore, the same policy directly selects the number of principal components.

Algorithm.

1. Compute singular values $\{\sigma_i\}_{i=1}^{r}$ of A and η_k for $k = 1, \ldots, r$.
2. Set $k_\tau \leftarrow \min\{k : \eta_k \geq \tau\}$.
3. Smooth η_k (optional), normalize $k \mapsto k/r$, and compute k_e via Kneedle (increasing/concave) on $(k/r, \eta_k)$; round to an integer.
4. Set default $k^\star \leftarrow \max(k_\tau, k_e)$. (Optional aggressive mode: $k^\dagger \leftarrow \min(k_\tau, k_e)$.)

Output: Truncation rank $k \in \{k^\star, k^\dagger\}$.

Implementation notes. (i) If $r = 0$, return $k = 0$; otherwise clip k to $[1, r]$. (ii) If Kneedle returns no elbow (flat or very steep curves), fall back to k_τ. (iii) Use a short Savitzky–Golay window for smoothing and the "increasing/concave" Kneedle setting, then enforce the monotonicity of η_k. (iv) Default thresholds: $\tau \approx 0.995$ for compression; $\tau \in [0.95, 0.99]$ for denoising. (v) To avoid brittle ties, always choose the smallest k that satisfies each of the rules.

Key insight. By consolidating cumulative-energy thresholds with elbow detection on the cumulative spectrum, the adaptive rule provides a mathematically grounded and reproducible guideline for rank selection. It removes manual tuning, adapts across tasks, and avoids the elbows under-selection failure mode, improving the fidelity in compression and stability in denoising. On spectra with extremely slow decay, the conservative policy may select a larger $k^\star$, trading compression for fidelity; in such cases practitioners may opt for the aggressive mode or impose a budget-constrained cap on k.

4. Applications and Experimental Setup

We evaluate SVD on three canonical low-rank tasks: image compression, image denoising, and principal component analysis (PCA). Each task highlights a different facet of low-rank modeling: storage reduction, noise suppression, and dimensionality reduction. Unless stated otherwise, we use the adaptive selection from Section 3 with the *default, conservative* policy $k^\star = \max(k_\tau, k_e)$; an optional aggressive mode $k^\dagger = \min(k_\tau, k_e)$ is exposed for heavy-noise denoising.

Protocol. All experiments share one codebase, fixed random seeds, and scripts that regenerate figures and tables from a clean environment. The libraries and exact command lines are listed in Appendix B. The metrics are consistent across tasks: PSNR (dB), SSIM, cumulative energy η_k, runtime, and a storage proxy ($mk+nk+k$) corresponding to the parameters of U_k, Σ_k, V_k. SSIM uses an 11×11 Gaussian window (default in `skimage.metrics.ssim`) with a data range $[0, 1]$.

Experimental design principles. Preprocessing, metrics, and defaults are matched across methods (SVD vs. baselines) so differences arise from rank selection or factorization, not setup. When stochasticity is present (e.g., randomized SVD), the seeds are fixed for comparability. We report both numerical fidelity (PSNR, η_k) and perceptual quality (SSIM).

4.1 Image Compression

Given an image matrix A, retaining the top k singular values yields

$$A_k = \sum_{i=1}^{k} \sigma_i \, u_i v_i^\top,$$

preserving the dominant structure while reducing the storage from mn to $mk+nk+k$ parameters. We evaluate on the `astronaut` image (converted to grayscale, 512×512), comparing fixed manual ranks to adaptive choices.

As a representative case, we contrast a fixed $k = 50$ with an adaptive rank $k^\star$ from the τ=0.995 cumulative energy rule (guarded by the elbow via max). Fig. 2 shows the cumulative

energy curve with both choices marked and reconstructions annotated with PSNR/SSIM. In this setting, the elbow typically lies below k_τ, so $k^\star = \max(k_\tau, k_e) = k_\tau$, delivering higher fidelity than small fixed ranks with a modest storage increase. For completeness, we also report a guard against under-selection, $k = \max(k_\tau, k_e)$, across the images and thresholds. The summary statistics are presented in Appendix Table 2.

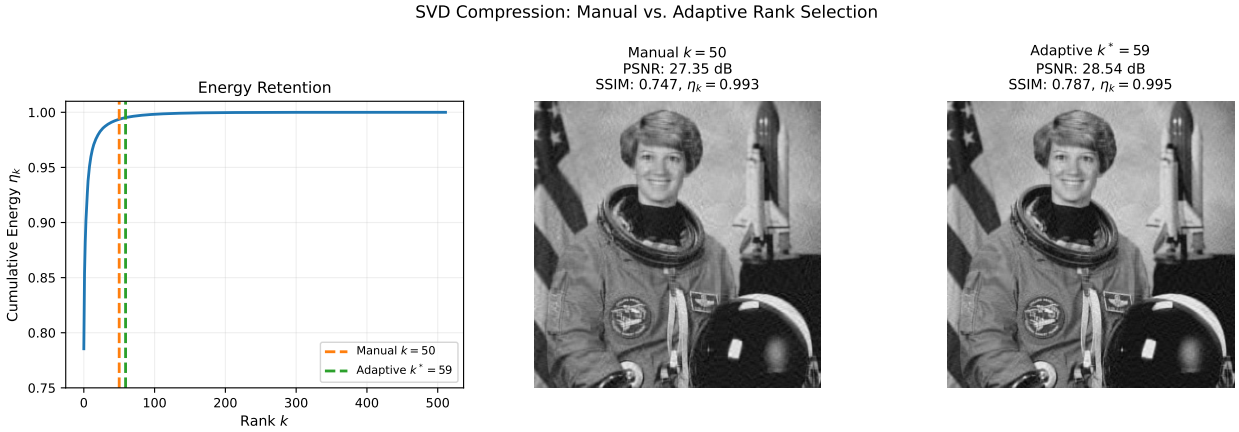

Fig. 2. Grayscale astronaut image: manual $k = 50$ vs. adaptive $k^\star = 59$. Adaptive rank improves fidelity (PSNR +1.2 dB, SSIM +0.04).

4.2 Image Denoising

For denoising, small singular values often capture high-frequency noise; thus, truncation acts as a low-pass spectral filter. We corrupt `astronaut` with additive Gaussian noise ($\sigma = 0.10$) and apply adaptive selection to balance noise removal and detail preservation. Fig. 3 shows the singular-value spectrum of the noisy astronaut image with threshold (k_τ) and elbow (k_e) annotations used for adaptive rank selection.

Benchmark comparisons with fixed ranks are presented in Table 1. The adaptive rule achieves PSNR **on par with or exceeding the best fixed choices**, without per-image tuning, and avoids elbow under-selection on slowly decaying spectra. An example of a full reconstruction is shown in Appendix Fig. 6.

Table 1: Comparison of manual fixed ranks and adaptive rank selection ($k^* = 58$) for denoising the noisy `astronaut` image ($\sigma = 0.10$). Reported are PSNR (dB), SSIM, and retained energy η_k. Adaptive rank achieves quality on par with the best fixed choices while avoiding manual tuning.

Method	**Rank** k	**PSNR (dB)**	**SSIM**	**Energy** η_k
Manual (fixed-k)	40	23.75	0.507	0.967
Manual (fixed-k)	80	23.70	0.467	0.978
Adaptive	58	24.12	0.497	0.973

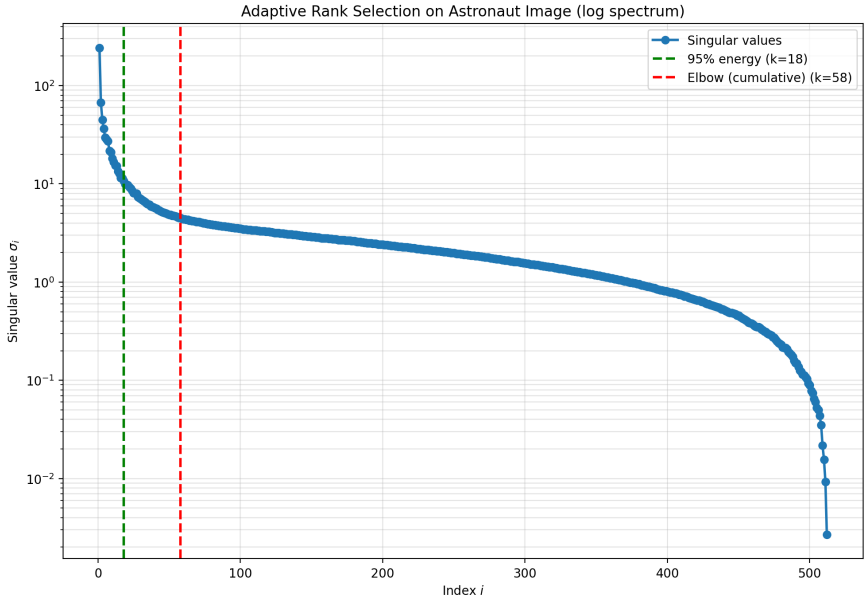

Fig. 3. Noisy astronaut image ($\sigma = 0.10$). Threshold (k_τ) and elbow (k_e) candidates are compared; the default adaptive rule selects $k^\star = \max(k_\tau, k_e)$.

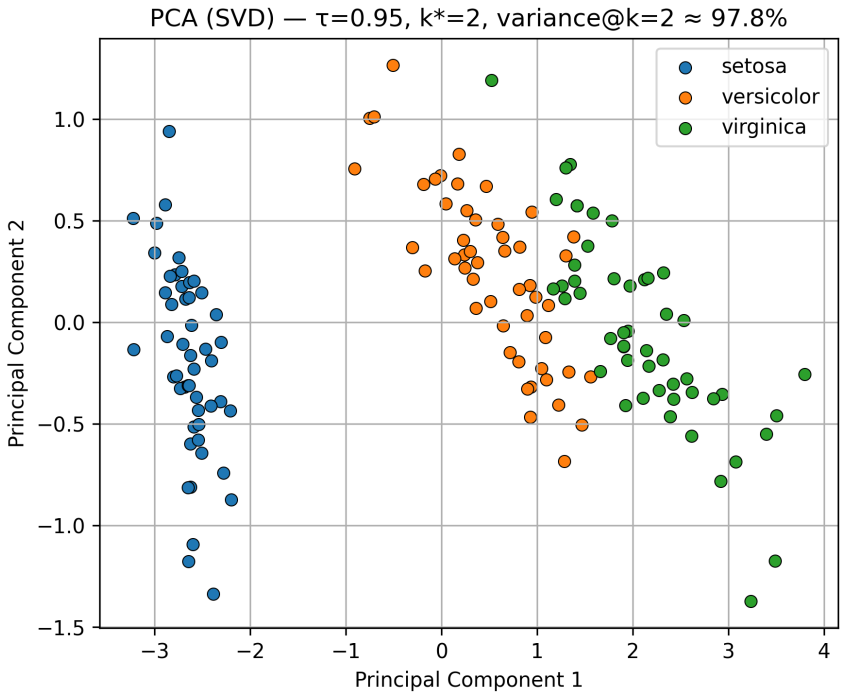

Fig. 4. PCA on the Iris dataset. With $\tau = 0.95$, the adaptive rule selects $k = 2$, retaining 97.8% variance.

4.3 Principal Component Analysis (PCA)

PCA is equivalent to applying SVD to a mean-centered data matrix [11]. For $X \in \mathbb{R}^{m\times n}$ with $X = U\Sigma V^\top$ and covariance $C = \frac{1}{m-1}X^\top X$, the eigenvalues satisfy $\lambda_i = \sigma_i^2/(m-1)$; the cumulative variance explained equals η_k (Section 3).

We illustrate this on the `iris` dataset. Using $\tau = 0.95$, the adaptive rule selects $k = 2$, capturing $\approx 97.8\%$ of the variance. Projecting $Z = XV_k$ yields a two-dimensional embedding that preserves the class structure (Setosa is clearly separated; Versicolor/Virginica partially overlap), as shown in Fig. 4.

The variance-retention curve (Appendix Fig. 7) confirms the stability of the adaptive choice across seeds and preprocessing variants.

Summary. Across all tasks, adaptive rank selection provides a reproducible alternative to manual tuning: (i) In **compression**, it matches or exceeds fixed-k fidelity at modest storage. (ii) In **denoising**, it balances detail and noise suppression while avoiding elbow under-selection; (iii) In **PCA**, component selection is automated to meet variance targets without per-dataset tuning.

5. Comparative Analysis of Matrix Decomposition Methods

To contextualize SVD, we benchmark against eigenvalue decomposition (EVD) and QR factorization on the 512×512 *grayscale* `astronaut` image. We study both the *fixed-rank* and *adaptive-rank* settings, evaluating the reconstruction fidelity (PSNR and SSIM) and runtime.

5.1 Benchmark Setup

All methods use `NumPy/SciPy` and identical hardware; BLAS/LAPACK backends are held fixed (Appendix B). Fixed-rank runs use $k \in \{10, 30, 50, 100\}$; adaptive selection follows Section 3 with the default policy $k^\star = \max(k_\tau, k_e)$ and $\tau = 0.995$ for high-fidelity compression. The elbow-only choices k_e are reported in the appendix. Metrics are averaged over repeated runs (error bars show ±1 std). The runtime is measured on noisy inputs ($\sigma = 0.10$, images normalized to $[0, 1]$); PSNR/SSIM are always computed against a clean reference.

Reproducibility metadata. All figures and tables were generated using a unified Python-based experimental pipeline with fixed random seeds and standardized preprocessing. Although absolute runtimes may vary depending on hardware and numerical backends, the relative trends remained stable across different environments.

Rank-k formulations.

- **SVD:** $A_k = U_k \Sigma_k V_k^\top$, the optimal rank-$k$ approximation in Frobenius and spectral norms [4].
- **EVD (Gram-based):** Form $A^\top A = V \Lambda V^\top$ with $\lambda_i = \sigma_i^2$. Taking $V_k = [v_1, \ldots, v_k]$ and $\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_{\mathrm{k}})$ yields

$$A_k \;=\; AV_kV_k^\top \;=\; U_k\Sigma_kV_k^\top, \qquad U_k \;=\; AV_k\Sigma_k^{-1}\ (\sigma_i > 0).$$

  This is algebraically equivalent to the truncated SVD, although explicitly forming $A^\top A$ may amplify conditioning errors.
- **QR (projection form):** With column-pivoted thin QR, $A\Pi = QR$. Let Q_k be the first k columns of Q. Then

$$A_k \;\approx\; Q_kQ_k^\top A \;=\; Q_kR_{k,:}\,\Pi^\top,$$

i.e., an orthogonal projection onto span(Q$_k$) (since $Q_k^\top A = R_{k,:}\,\Pi^\top$). This is generally *non-optimal* for a fixed k.

5.2 Benchmark Summary

Fig. 5 plots PSNR versus rank. SVD attains the highest fidelity at each k, consistent with optimality, and EVD reconstructions match SVD to numerical precision (when computed via $AV_kV_k^\top$). See Fig. 8 for the same curve with the adaptive $k^\star$ indicated. QR is faster but yields lower fidelity at the same k. The exact timings depend on the backend and cache effects; see Appendix Table 5.

Complete numerical results: Table 3 (fixed-k), Table 4 (adaptive with $\tau = 0.995$), and Table 5 (runtime).

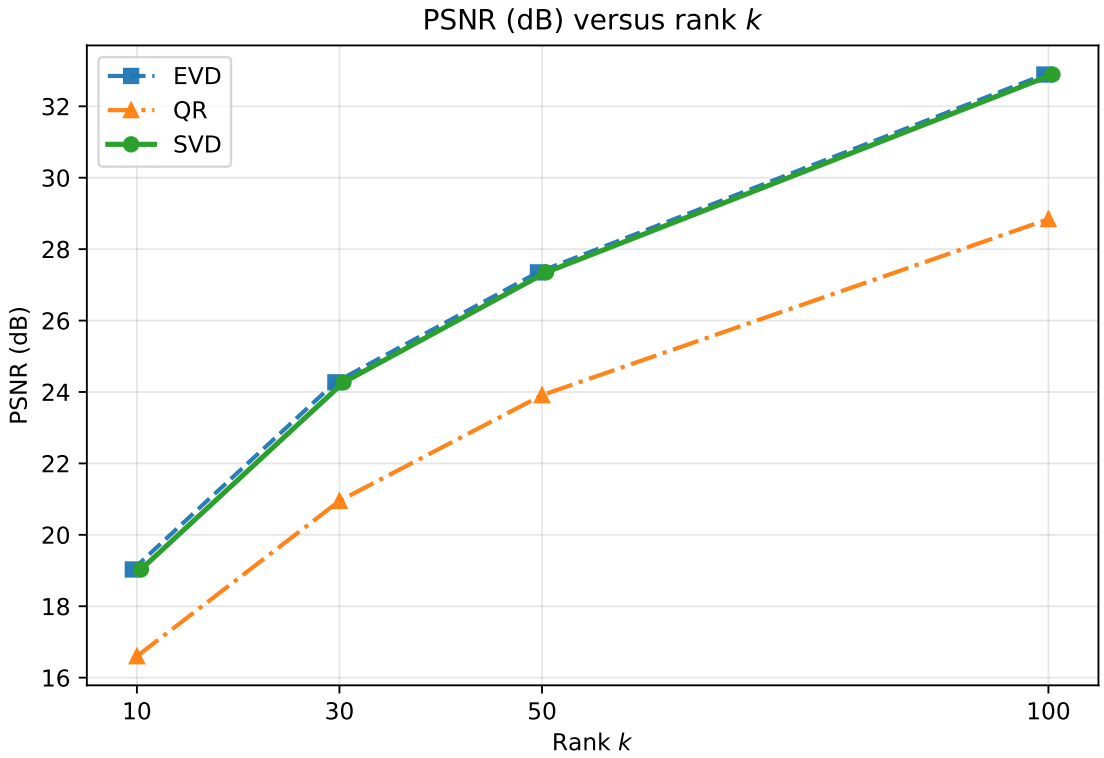

Fig. 5. PSNR vs. rank k for SVD, Gram-based EVD, and pivoted QR. SVD/EVD attain the highest fidelity; QR is faster but less accurate.

5.3 Key observations

- **SVD optimality.** At every tested k, SVD achieved the highest PSNR/SSIM values. For example, at $k = 50$, SVD reached 27.35 dB / 0.747 vs. QRs 23.91 dB / 0.624; at $k = 100$, 32.89 dB / 0.895 vs. 28.84 dB / 0.777 (Table B2).

- **EVD equivalence.** Reconstructions based on $A^\top A$ (via $AV_kV_k^\top$) matched SVD within rounding error, as expected; directly forming $A^\top A$ can, however, increase sensitivity to conditioning.

- **Adaptive rank ($\tau = 0.995$).** The rule selected $k^\star = 59$ in our run, improving fidelity over $k = 50$ by $\approx$1.2 dB PSNR (28.54 vs. 27.35) and $\approx$0.04 SSIM (0.787 vs. 0.747). The QR at $k^\star$ remained lower in quality (24.72 dB / 0.648).
- **Speed–quality trade-off.** QR was $\sim 2\times$ faster than SVD at moderate k (e.g., $k \in \{10, 50, 100\}$), at the cost of noticeable accuracy loss.

5.4 Conclusion

These benchmark experiments confirm the trade-offs. SVD (and Gram-based EVD) achieves the best fidelity at a given rank. QR offers faster runtime at reduced accuracy, and adaptive selection with $\tau = 0.995$ provides a practical balance. These observations motivate the unified framework summarized in the overall conclusion section.

6. Conclusion and Future Work

We presented a unified and reproducible framework for applying SVD to image compression, denoising, and PCA, with systematic comparisons to EVD and QR. Our central contribution is an **adaptive rank selection** strategy that combines cumulative energy thresholds with elbow detection on the cumulative spectrum, eliminating manual tuning while preserving interpretability.

Key findings. (i) SVD consistently achieves the highest reconstruction fidelity, in line with Eckart–Young [4]; (ii) Gram-based EVD reconstructions are numerically indistinguishable from SVD. (iii) QR offers speed advantages at reduced fidelity; (iv) Adaptive selection improves quality over fixed settings while standardizing the decisions across tasks.

Our contribution is to **consolidate** commonly used heuristics into a principled, data-driven, and fully reproducible framework for adaptive rank selection. For clean images, a 99.5% cumulative-energy threshold emerges as a reliable default that balances perceptual fidelity and compression efficiency.

The proposed framework is primarily methodological and does not introduce a new matrix decomposition or new theoretical optimality guarantees. In addition, the elbow-detection component remains partially empirical and may depend on the spectral characteristics of the data. Nevertheless, the approach provides a unified and reproducible strategy for adaptive rank selection across multiple SVD-based applications.

The entire framework, including the code and figure-generation scripts, is openly available.[2]

Implementation updates. All scripts were refactored into a unified modular structure with standardized outputs and harmonized argument parsing. These updates ensure that every figure and table can be reproduced from a clean environment with a single command execution.

Future work. Extensions include color images and video, streaming and distributed settings, and hybrids that pair SVD with learned priors. An especially promising direction is the integration of *randomized and streaming SVD algorithms* [12], enabling scalable low-rank approximation on large datasets while preserving theoretical guarantees. Beyond imaging,

[2]All code and reproducibility materials are publicly accessible at https://github.com/gayaneghazaryan-dot/svd-image-analysis and archived on Zenodo with DOI: https://doi.org/10.5281/zenodo.20418974 .

applications in NLP, recommender systems, and bioinformatics are the natural next steps. More broadly, SVD remains foundational for modern machine learning embeddings and representation learning, suggesting further opportunities for cross-disciplinary impact.

Appendix

A. Compression Quality Metrics

For completeness, we summarize the metrics used in the main text as follows.

- **Compression ratio (CR).** For an $m \times n$ image stored as a dense array, a rank-k SVD uses $mk + nk + k$ parameters (for U_k, Σ_k, V_k) versus mn for the original.[3] We therefore use the proxy
$$\mathrm{CR}_{\mathrm{param}} \;=\; \frac{\mathrm{mn}}{\mathrm{mk} + \mathrm{nk} + \mathrm{k}}.$$
- **Peak signal-to-noise ratio (PSNR).**
$$\mathrm{PSNR} \;=\; 10 \log_{10}\!\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right),$$
where MAX = 1 for our normalized images, and MSE is the mean squared error vs. the clean reference.
- **Structural similarity (SSIM).** A perceptual score in $[0, 1]$ combining luminance, contrast, and structure; higher values indicate better fidelity.
- **Energy retention.**
$$\eta_k \;=\; \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{r} \sigma_i^2} \;=\; \frac{\|A_k\|_F^2}{\|A\|_F^2},$$
coinciding with the cumulative variance explained in PCA for mean-centered data.

These metrics are used consistently in Subsections 4.1 and 4.2.

B. Detailed Benchmarking and Supplementary Figures

B1. Image Compression (Extended Results)

The extended results illustrate how rank selection influences reconstruction quality and storage; see Table 2.

These results support the main-text default: $\tau = 0.995$ balances compression and perceptual fidelity, whereas $\max(k_\tau, k_e)$ guards against elbow under-selection.

B.2 Image Denoising (Extended Results)

To complement the main-text results, we illustrate adaptive rank selection on a noisy `astronaut` image. Fig. 6 shows the cumulative energy curve with the selected $k^\star$, along with the noisy input and the corresponding adaptive reconstruction.

[3]Byte-level ratios depend on datatypes and entropy coding; we report a parameter-count proxy.

Table 2: Comparison of manual and adaptive rank selection on the 512×512 grayscale `astronaut`. Manual fixed-k uses $k = 50$. Adaptive results are grouped by threshold ($\tau = 0.95$ and $\tau = 0.995$), with elbow-only and elbow+guard rules shown separately.

Method	**Rank k**	**PSNR (dB)**	**SSIM**	**Energy η_k**
Manual				
Fixed-k (50)	50	27.35	0.747	0.993
Adaptive ($\tau = 0.95$)				
Energy threshold	9	18.58	0.459	0.951
Adaptive ($\tau = 0.995$)				
Energy threshold	59	28.54	0.787	0.995
Elbow only	35	25.13	0.671	0.989
Elbow + guard ($\max(k_\tau, k_e)$)	59	28.54	0.787	0.995

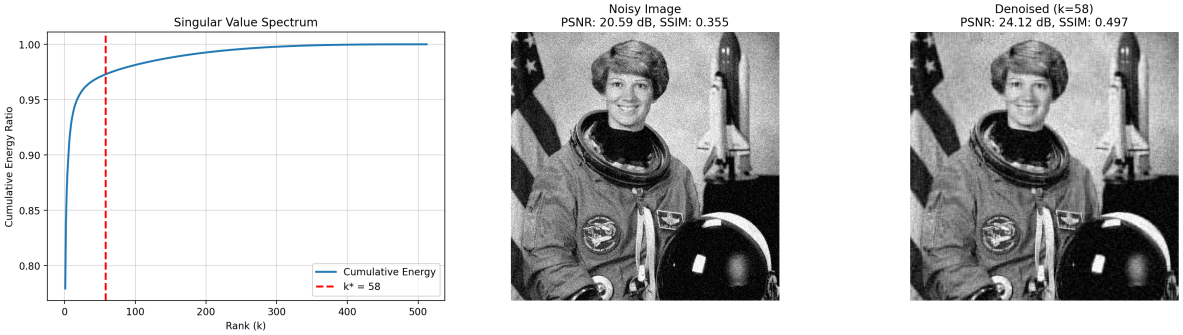

Fig. 6. Noisy astronaut with adaptive rank $k^\star$. Left: cumulative energy curve; middle: noisy input; right: reconstruction with PSNR/SSIM.

B.3 Adaptive Rank Selection in PCA

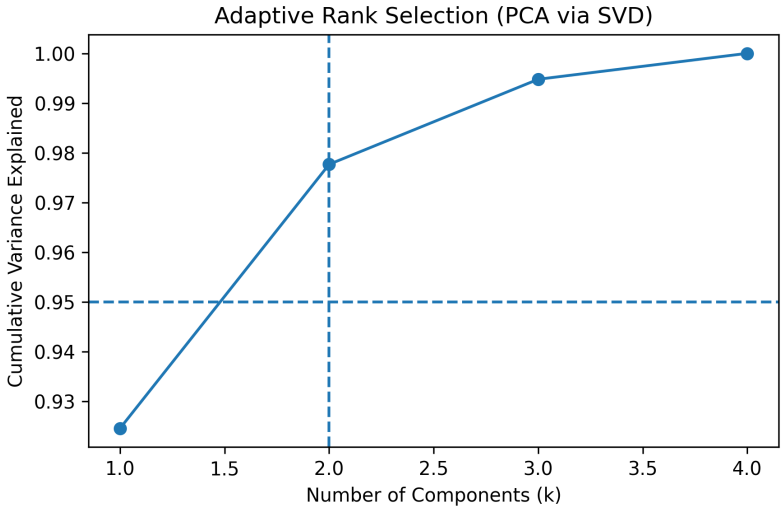

Fig. 7. Cumulative variance explained for the `iris` dataset. With $\tau = 0.95$, the adaptive rule selects $k = 2$, consistent with Section 4.3.

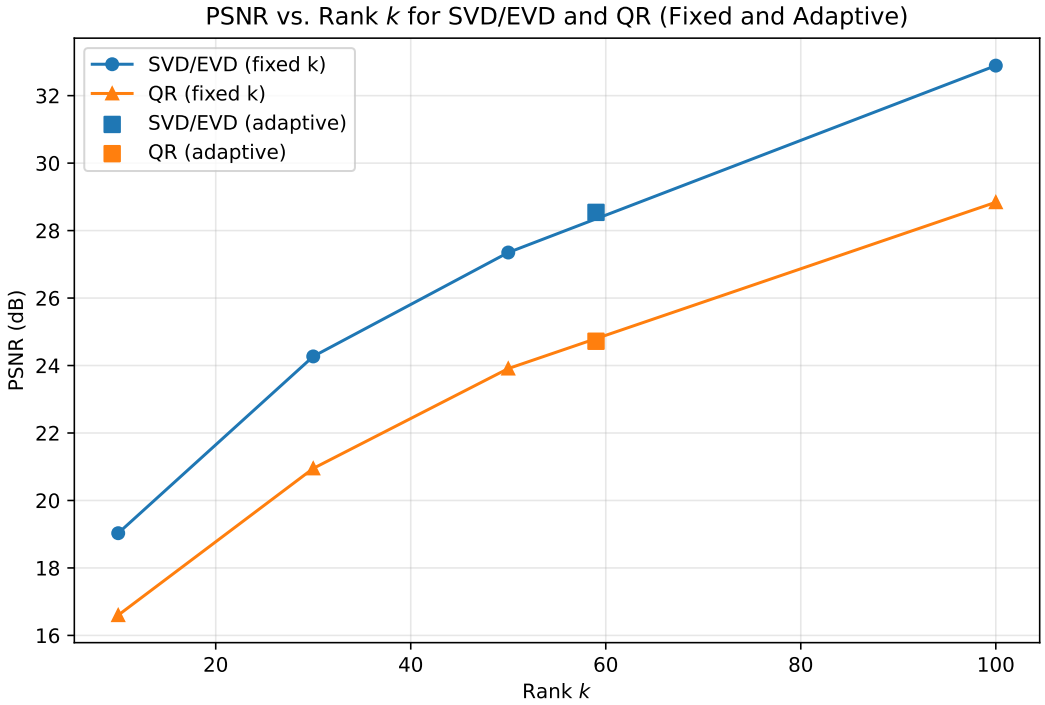

Fig. 8. PSNR vs. rank k for fixed and adaptive policies. Adaptive $k^\star = 59$ improves fidelity by ≈ 1.2 dB over fixed $k = 50$.

B.4 Benchmarking Tables and Figures

Table 3: Fixed-rank reconstructions on the 512×512 grayscale `astronaut` (clean). SVD and Gram-based EVD coincide; QR achieves lower fidelity.

Method	$k = 10$		$k = 30$		$k = 50$		$k = 100$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SVD	19.03	0.473	24.27	0.643	27.35	0.747	32.89	0.895
EVD	19.03	0.473	24.27	0.643	27.35	0.747	32.89	0.895
QR	16.60	0.396	20.95	0.525	23.91	0.624	28.84	0.777

Table 4: Adaptive selection with $\tau = 0.995$ on the grayscale `astronaut`. Δk is relative to fixed $k = 50$.

Method	Adaptive k	PSNR (dB)	SSIM	Δk vs. 50
SVD	59	28.54	0.787	+9
EVD	59	28.54	0.787	+9
QR	59	24.72	0.648	+9

Table 5: Runtime comparison for rank-k approximations on noisy `astronaut` ($\sigma = 0.10$). Values are averages; absolute times depend on hardware/backends, but relative trends are consistent. EVD is excluded from this table as its runtime is dominated by forming $A^\top A$ and is not directly comparable; its reconstruction quality matches SVD to numerical precision (Tables 3,4).

Method	$k = 10$		$k = 50$		$k = 100$		Time (ms)
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
SVD	19.03	0.473	27.35	0.747	32.89	0.895	102–105
QR	16.60	0.396	23.91	0.624	28.84	0.777	49–52

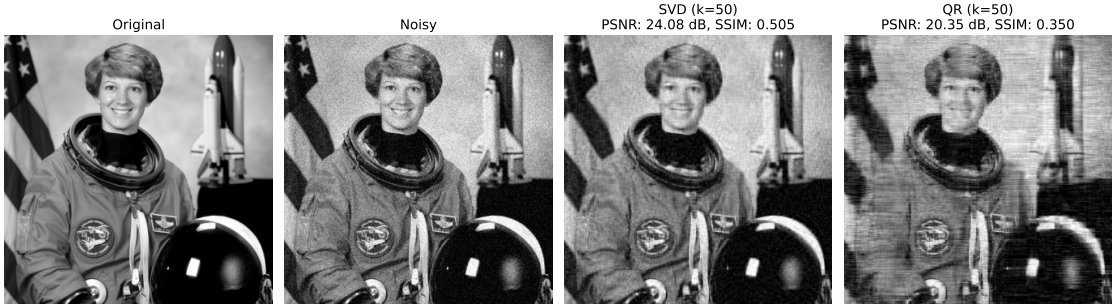

Fig. 9. Visual comparison at $k = 50$. SVD preserves sharper edges; QR exhibits blur and structural loss.

C. Supplementary Materials

To ensure reproducibility, the supplementary materials accompany this submission. They include:

- Python scripts implementing SVD-based compression, denoising, and PCA;
- Example images and configuration files for complete end-to-end replication of the experiments.

Data and Code Availability

All Python scripts, figures, and tables used in this study are publicly available at: https://github.com/gayaneghazaryan-dot/svd-image-analysis.

Each experimentcovering image compression, denoising, benchmarking, and PCAcan be reproduced in full by running one of the four Python scripts provided in the `code/` directory. All figures and tables are generated automatically from these scripts, ensuring full reproducibility without the need for Jupyter notebooks or external data files.

References

[1] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 4th edition edition, 2013.

[2] P.C. Hansen, The truncated svd as a method for regularization", *BIT Numerical Mathematics*, vol. 27, no. 4, pp. 534–553, 1987. doi: 10.1007/BF02579243.

[3] M.E. Wall, A. Rechtsteiner, and L.M. Rocha, "Singular value decomposition and principal component analysis", *A Practical Approach to Microarray Data Analysis*, pp. 91–109, 2003.

[4] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank", *Psychometrika*, vol. 1, no. 3, pp.211–218, 1936. doi: 10.1007/BF02288367.

[5] H. C. Andrews and C. L. Patterson, "Singular value decomposition (svd) image coding", *IEEE Transactions on Communications*, vol. 24, no. 4, pp. 425–432, 1976. doi: 10.1109/TCOM.1976.1093311.

[6] H.R. Swathi and G. Sulochana, "Image compression using svd", *IOP Conference Series: Materials Science and Engineering*, vol. 263, 042082, 2017. doi: 10.1088/1757-899X/263/4/042082.

[7] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, SIAM, 2010.

[8] Q. Guo, C. Zhang, Y. Zhang and H. Liu, "An efficient svd-based method for image denoising", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 868–880, 2016, doi: 10.1109/TCSVT.2015.2416631.

[9] S. Gu, L. Zhang, W. Zuo and X. Feng, "Weighted nuclear norm minimization with application to image denoising", *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2862–2869, 2014, doi: 10.1109/CVPR.2014.366.

[10] E.J. Cands, X. Li, Y. Ma and J. Wright, "Robust principal component analysis?", *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011. doi: 10.1145/1970392.1970395.

[11] I.T. Jolliffe and J. Cadima, *Principal Component Analysis*, Springer, 2nd edition edition, 2016.

[12] N. Halko, P.G. Martinsson and J.A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions", *SIAM Review*, vol. 53, no. 2, pp. 217–288, 2011. doi: 10.1137/090771806.

[13] Z. Lin, "A review on low-rank models in data analysis", *Big Data and Information Analytics*, vol. 1, no. 2-3, pp. 139–161, 2016. doi: 10.3934/bdia.2016001.

[14] P.C. Hansen, *Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion*, SIAM, 1998.

[15] M. Gavish and D.L. Donoho, "The optimal hard threshold for singular values is $4/\sqrt{3}$", *Biometrika*, vol. 101, no.3, pp. 587–602, 2014. doi: 10.1093/biomet/asu056.

[16] V. Satopaa, J. Albrecht, D. Irwin and B. Raghavan, "Finding a "kneedle" in a haystack: Detecting knee points in system behavior", *31st International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pp. 166–171. IEEE, 2011. doi: 10.1109/ICDCSW.2011.20.

[17] L.N. Trefethen and D. Bau, *Numerical Linear Algebra*, SIAM, 1997.

[18] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 2nd edition edition, 2012.

[19] L. Mirsky, "Symmetric gauge functions and unitarily invariant norms", *Quarterly Journal of Mathematics*, vol. 11, no. 1, pp. 50–59, 1960. doi: 10.1093/qmath/11.1.50.

[20] V. Klema and A. Laub, “The singular value decomposition: Its computation and some applications”, *IEEE Transactions on Automatic Control*, vol. 25, no. 2, pp. 164–176, 1980. doi: 10.1109/TAC.1980.1102314.

Սինգուլյար արժեքների վերլուծության տեսությունն ու կիրառությունները պատկերների մշակման և տվյալների վերլուծության մեջ

Գայանե Գ. Ղազարյան[1] և Արտաշես Լ. Ղազարյան[2]

[1]Ֆիզիկայի ինստուտուտ, Երևանի պետական համալսարան, Երևան, Հայաստան
Մյունխենի տեխնիկական համալսարան, Մյունխեն, Գերմանիա
e-mail: gayane.ghazaryan@ysu.am, artashes.ghazaryan17@gmail.com

Ամփոփում

Սինգուլյար արժեքների վերլուծությունը (Singular Value Decomposition, SVD) մատրիցների վերլուծության հիմնարար մեթոդներից է, որը լայնորեն կիրառվում է պատկերների մշակման, չափողականության նվազեցման և հաշվարկային գծային հանրահաշվի խնդիրներում: Մեթոդի գործնական արդյունավետությունը պայմանավորված է փոքր սինգուլյար արժեքների կտրմամբ, ինչը հնարավորություն է տալիս նվազեցնել տվյալների ներկայացման ծավալը՝ միաժամանակ պահպանելով դրանց հիմնական կառուցվածքային առանձնահատկությունները: SVD-ի կիրառման կարևոր խնդիրներից է կտրման ռանգի ընտրությունը, որը որոշում է հավասարակշռությունը վերականգնման ճշգրտության, հաշվարկային արդյունավետության և արդյունքների կայունության միջև :

Աշխատանքում առաջարկվում է ռանգի ադապտիվ ընտրության մեթոդ, որը համատեղում է կուտակային էներգիայի վրա հիմնված չափանիշը և սինգուլյար արժեքների սպեկտրի «ծնկի» (elbow) կետի ավտոմատ հայտնաբերումը: Առաջարկվող մոտեցումը հիմնավորված այլընտրանք է ռանգի ընտրության լայնորեն կիրառվող մոտեցումներին: Մեթոդն ուսումնասիրվել է պատկերների սեղմման, աղմուկի նվազեցման և գլխավոր բաղադրիչների վերլուծության (PCA) խնդիրներում, ինչպես նաև համեմատվել է սեփական արժեքների դեկոմպոզիցիայի (EVD) և QR գործոնավորման հետ: Թվային փորձարկումների արդյունքները ցույց են տալիս, որ ռանգի ընտրության ադապտիվ կանոնն ապահովում է վերականգնման բարձր որակ և նպաստում է հաշվարկային փորձերի վերարտադրելիությանը և հատկապես աղմկոտ կամ վատ պայմանավորված տվյալների դեպքում:

Աշխատանքում օգտագործված ծրագրային կոդը, նկարները և աղյուսակները հրապարակված են բաց հասանելիությամբ, ինչը հնարավորություն է տալիս ամբողջությամբ վերարտադրել ներկայացված փորձերը: Առաջարկվող մոտեցումը միավորում է տեսական հիմքերը, հաշվարկային գործընթացների ավտոմատացումը և վերարտադրելի հետազոտությունների սկզբունքները՝ վերահաստատելով, որ սինգուլյար արժեքների վերլուծությունը մնում է ինչպես մաթեմատիկորեն հիմնավորված, այնպես էլ գործնականում բազմակողմանի գործիք ժամանակակից տվյալների վերլուծության համար:

Բանալի բառեր՝ սինգուլյար արժեքների վերլուծություն, ռանգի ադապտիվ ընտրություն, կուտակային էներգիա, «ծնկի» մեթոդ (elbow), պատկերների աղմուկի նվազեցում, պատկերների սեղմում, գլխավոր բաղադրիչների վերլուծություն:

О теории и приложениях сингулярного разложения в обработке изображений и анализе данных

Гаянэ Г. Казарян[1] и Арташес Л. Казарян[2]

[2]Мюнхенский технический университет, Мюнхен, Германия
e-mail: gayane.ghazaryan@ysu.am, artashes.ghazaryan17@gmail.com

Аннотация

Сингулярное разложение (Singular Value Decomposition, SVD) является одним из фундаментальных методов матричной факторизации, широко применяемых в обработке изображений, снижении размерности и вычислительной линейной алгебре. Практическая эффективность метода связана с возможностью отсечения малых сингулярных значений, что позволяет уменьшать объём хранимых данных при сохранении наиболее существенной информации. Одной из ключевых задач при использовании SVD является выбор ранга усечения, определяющего компромисс между точностью аппроксимации, вычислительной эффективностью и устойчивостью результатов.

В работе предлагается метод адаптивного выбора ранга, объединяющий энергетический критерий, основанный на накопленной доли энергии, и автоматическое обнаружение точки перегиба спектра сингулярных значений (elbow). Такой подход представляет собой обоснованную альтернативу широко используемым эвристическим правилам выбора ранга.

Предложенный метод исследован на задачах сжатия изображений, подавления шума и анализа главных компонент (PCA), а также сопоставлен с разложением по собственным значениям (EVD) и QR-факторизацией. Результаты численных экспериментов показывают, что адаптивное правило выбора ранга обеспечивает высокое качество восстановления и повышает воспроизводимость вычислительных исследований, особенно в условиях шумных и плохо обусловленных данных.
Исходный код, рисунки и таблицы опубликованы в открытом доступе, что обеспечивает возможность полного воспроизведения всех представленных экспериментов. Предлагаемый подход объединяет теоретические результаты, автоматизацию вычислительных процедур и принципы воспроизводимых исследований, подтверждая, что сингулярное разложение остаётся как математически обоснованным, так и практически универсальным инструментом современного анализа данных.

Ключевые слова: сингулярное разложение, адаптивный выбор ранга, накопленная энергия, метод локтя (elbow), подавление шума изображений, сжатие изображений, анализ главных компонент.

Mathematical Problems of Computer Science 65, 49–62, 2026.
doi:10.51408/1963-0145

UDC 004.8

On Measurable Tournaments for Progressing Generalized Cognizers4

Edward M. Pogossian

Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: epogossi@aua.am

Abstract

Researchers such as Kurzweil and Goertzel predict that AI, due to the progress in LLM is entering a period of exponential growth toward Artificial General Intelligence (AGI). They believe that if such AGI is capable of rewriting its own code, it could evolve into a superhuman AI, possessing the cognitive and computational power of all human civilization

We interpret this requirement for AGI as an unavoidable ability of generalized cognizers to generate expectably reasonable versions of development of themselves, along with trustworthy means of measuring these versions and selecting the most promising of them wrt the utilities of cognizers.

Generalizing our prior assertion that knowledge-based chess strategies can be strongly scaled by local tournaments, we argue an analogous statement for suitable measuring of the progress of generalized cognizers in the frame of an adequate theory of cognizing and address to the ways to strengthen it.
Keywords: AGI, Generalized cognizers, Measuring of progressing, Local tournaments, Absolute ordering.
Article info: Received 23 March 2026; sent for review 9 April 2026; accepted 8 May 2026.

1. Introduction

1.1. Significant progress in machine cognizing, thus in its kernel ability to learn, is challenging to empower and accelerate the learning of generalized cognizers.
1.1.1. Futurologists such as Kurzweil [1] and Goertzel[2] interpret these aims as an invention of Artificial General Intelligence (AGI).

Gertzel is certain that AI research is entering a period of exponential growth, arguing it by referring to Kurzweils prognostications, due the progress of LLM and the ongoing combination of paradigms of AI into a single framework of AGI.

He believes that if such AGI could rewrite its own code, it could evolve into a superhuman AI, possessing the cognitive and computational power of all human civilization.

1.2. We interpret the requirement to AGI to rewrite its own code as an unavoidable ability of mighty generalized cognizers to generate expectably reasonable versions of their development, along with trustworthy means of measuring these versions and selection of the most promising of them wrt the utilities of cognizers.
1.3. In fact, such an interpretation reduces the problem of mighty cognizers to the elaboration of effective evolutionary algorithms exempted from their main prohibiting difficulties in providing suitable fitness and selection constituents [3].
1.3.1. These difficulties were relieved for a particular class of problems approximated by chess-like reproducible games (rg). It was argued that the corresponding knowledge-based rg strategies can be, in fact, strongly scaled by local tournaments wrt ideal tournaments of all possible strategies of those problems [4, 5, 6, 7].
1.3.2. The corresponding theorems have been proven [4] stating that in classes A of algorithms enabling **to enumerate** all possible appearances of algorithms by enumerating the states of regulators determining any of their appearances, it is possible to arrange the sequences of local tournaments between those algorithms similar to those by Elo [8] in chess allowing to converge the sequences of chess players to the best of them wrt absolute tournament among all of them.

In other words, it was proven that knowledge-based rg cognizers can progress by local tournaments to the best ones in corresponding classes scaled wrt total tournaments between all class members.

Generalizing the above assertion can allow us to claim that an analogous statement on suitable measuring also be argued for the progress of generalized cognizers in the frame of an adequate theory of cognizing.
1.3.3. Such a theory we, analogously to the theory of algorithms, ground on generalized constructive models *cogs* of acknowledged Piagets descriptive models *pcogs* of human cognizing *hcogs* and specifying *mentals*, as constructive models of mental structures, state that:

- *constructive generalized models* cogs *of human cognizers* hcogs *can be arguably grounded on Piagets descriptive models* pcogs *of* hcogs;

- *there are premises that generalized* cogs *including the ability of learning* mentals *constructively and adequately model stabilized and developmental aspects of Piagets descriptive models* pcogs *including learning of mental structures (mss)* [9, 10].

1.4. In what follows, we argue that, analogous to chess, suitable measuring of progress of cognizing is also valid for the progress of generalized cognizers.

Consequently, we remind the argument in [9] that combinatorial games can serve as acceptable models of means of cognizing the Universe U; therefore *humans cognizing* pcog *of* U *can be approximated by cognizing combinatorial games, particularly, by rg reproducible subsets of such games and corresponding* rgcogs *models of* cogs *and:*

- *cognizing of combinatorial games, particularly chess, can properly approximate human cognizing, thus approximating the theory of cognizing.*

Then, listing some advances in cognizing of chess, we remind the theorems on locally progressing knowledge–based chess cognizers to the best in target classes and argue a trajectory of their proving to avoid the sophisticated one in [4], concluding with the targets for research.

2. Arguing Cognizing as Combinatorial Games

2.1. The entire Universe and new target **problems appear to humans** *combinatorically* because they need to be analyzed not in the frame of absolutely reliable conceptual frameworks, but with involvement of the entire ad hoc knowledge and means of its enrichment.
2.1.1. Such problems, if lucky, can be completely resolved. For example, in our practice [32], such a complete solution was found for Schroders combinatorial problem on specifying the systems of sets with a min number of subsets, where the proof followed from the proofs of the chain of 36 lemmas (reminding also the proof of Nadareshvilys chess etude required the analysis of chess tree in the about 39 depths [11]).

Unfortunately, combinatorial problem, generally, appear as intractable, why are resolved only fragmentarily for sub-problems, which, as a rule, are too local and do not illuminate the solution of the original problem. Nevertheless, some combinatorial classes are extractable, which satisfy additional requirements, allowing the development of theories and tractable algorithms common to these classes .
2.2. Then, the problems and dynamicity of the entire Universe U, ideally, appear to humans by realities affected by a variety of bundles b of known and possibly unknown doers, as a rule, complemented with doers *hcogs* of cognizing these realities.

These effects in time realities humans represent by classifiers of so far gained mss. Namely, *do-* and *systemic* classifiers of mss altogether comprise communalized, thus, estranged-able from particular members *x@C, attributes*, while the activated attributes, *situations*, operable, and communicably represent the peculiarities and relationships of the observable Universe U reflected, particularly, in sentences and phrases of languages.
2.3. Ideally, situations of U appear to humans h as activated at times t classifiers and relationships, particularly, cause-effect ones, so far gained by h afore t. Consequently, the possibilities of humans to reason on the Universe U and prognosticate decisions at time t are bounded by these ad hoc cognized classifiers and relationships, or **ad hoc cognizable universe (ahcU)**.

Seemingly, this ahcU idealization addresses issues rooted in Platos plate and Vernadskys noosphere comprehensions of the cognized and cognizable, Blavatskys cosmic mind [12], as well as Herbrands model of possible propositional inferences [13] and Wolframs *ruliard* "...the abstract object corresponding to the entangled limit of all possible computational processes: the result of following all possible computational rules in all possible ways [14].
2.3.1. However, in the **real** human universe U, interactions and cognizing of U are apparently processed locally. People, professionals, experts are goal-oriented and work with local sets of relevant attributes, thus, with situations and corresponding local problems, and try to resolve them by so far gained mss. Otherwise, if these attempts fail, people associate them with further cognizing of the problems to enrich mss and become more successful.
2.4. Thus, given situations p, certain utilities, bundles b of doers/actors and their possible doings in corresponding problems P, thus, changes/transformations of p at times t caused by b, the experts, in agree with mss gained before t, can prognosticate possible trajectories/branches of changes of p in time t by local trees T_s of situations allowing them to examine T_s for the most perspective trajectories of decisions in p and do correspondingly to solve the problems P.
2.4.1. Moreover, when experts in situations p of T_s aim to prognosticate strategies of actions for particular doers d and utilities, they for each action a of d consider all expected responses of the rest $b \setminus d$ doers of b in p with corresponding possible changes of p allowing to transform

the trees T_s of situations into games on T_s, which comprise all possible strategies of d in interactions of d with $b \setminus d$, thus, to search for the best strategies for d to solve P.

2.4.2. Such games allow experts to preliminarily process the expected versions of strategies of solutions in the framework of the models of problems, i.e., the framework of gained experience on peculiarities of situations of the problems, compare the strategies, choose the most promising of them and apply, instead of riskier immediate elaboration of strategies induced by widespread reasoning common for any situation. And because IDs of mss are communalized, experts can communicate them, i.e., explain and understand mss of each other, thus, collaborate for more effective solutions.

2.4.2.1. For example, in oligopoly competitions producers, say A, B, C, D ones, influence the market situations by bundles of 4P – price, product, promotion and proliferation, actions. Such competitions can be modeled by trees of situations, which focus on the benefits and strategies, for example, for A, can be transformed into game trees, where for each action of A in p, all possible responses of B, C, D will be branched along with corresponding changes of p [10].

While game trees only approximate oligopoly competitions, in chess, where white and black doers possess predetermined actions of available pieces on the chess board and lists of doings of each piece in each situation, game tree models represent chess game adequately.

2.5. Resuming, we state:

St. 1. *Ideally, humans* hcogs *cognizing of the Universe U can be modeled by cognizing global combinatorial games gG on global gT_s trees of situations, which, in turn, are comprised of real $hcogs_i$ cognizing of local sub-games lg_i $(i = 1, 2, \ldots, n)$ of gG on the corresponding local sub-trees lT_{s_i} of situations of gT_s altogether covering the global gT_s ones.*

2.5.1. Addressing the theory of cognizing statement St.1., lets assume:

Cl. 1.1. *The theory of cognizing local combinatorial sub-games lg_i $(i = 1, 2, \ldots, n)$ of the global gG ones by adequate constructive models of $hcogs_i$ can, in principle, be adequate to constructive models for the entire theory of cognizing.*

2.6. Then, addressing the outputs o_{gc} of gU cognizing, i.e., gU gained mss, lets state that, generally, only the composition of outputs o_{lc_i} of o_{lc_i} cognizing of individual local sub-games lg_i, i.e., lg_i gained mss, can represent o_{gc}, since, generally, o_{lc_i} vary by varying lg_i.

Particularly, o_{gs} can be enriched by mss common for all lg_i. This reasoning is similar, for example, to revealing that the mosaic is glass-made by analysis of its variably colored, but commonly glass–made units.

In contrast with o_{lc_i}, the means of cognizing m_{gU} and m_{lc_i} of gU and lg_i, correspondingly, can be accepted as equal.

Indeed, interpreting cognizers *hcogs* as the means of cognizing, then recalling Piagets argument that in solving problems, *hcogs* throughout the entire life are governed by universal laws, then follows the equality of *hcogs* wrt cognizing of global and local games, which allows to assume:

Cl. 1.2. *Adequate constructive models of $hcogs_i$ for cognizing local combinatorial sub-games lg_i $(i = 1, 2, \ldots, n)$, in principle, can approximate adequate constructive models of* hcogs *for cognizing the entire Universe U.*

2.7. Then, accepting that local games lg_i, while varying in their domains, thus, in outputs o_{lg_i}, assumingly, have equal cognizers, it is worth to assuming:

Cl. 1.3. *The accuracy of approximation of* hcogs *by cognizers $hcogs_i$ of classes of diversified local games lg_i having equal cognizers rises with rising the power of these classes and possibly dense coverage of gG by them.*

2.7.1. Following this implication, lets introduce a class of reproducible local games, argue its consistency with the requirements of corollaries Cl.1.1–3, followed by positives for the theory of cognizing inferred from this consistency.
2.8. Analogously to games with perfect information, the **rg class of** *reproducible* **games** is identified in [4, 15] by the following requirements:

- there are (a) interacting actors (doers, players, competitors, etc.) performing (b) identified types of actions at (c) specified moments of time and (d) specified types of situations,
- there are identified benefits for each of the actors,
- the situations perceived by actors of the games and those after their acting have to be specifiable by if-then rules, allowing for regular experimentation with them .

2.8.1. It appears that rg class among chess-like games a suitable for approximating massive application problems in competition, defense, and dialogue, including intrusion protection, marketing, craft defense, tutoring, testing of programs, and, as a rule, there were no difficulties in their proper, or quasi-adequate rg representation [4].

Studies of knowledge-based solutions for such, explicitly or not, identified combinatorial games rooted in the Theory of Games and Shannons works in chess [16], are presented, particularly, in [4, 17, 18, 19, 20, 21, 22].

Even more, rg class can be properly extended, allowing, for example, the appearance of situations that are not only strongly deterministic, but also have some proximity and likelihood.
2.8.2. Hence, we assume:
Ass.1. *Games of rg class and their extensions quasi-adequately approximate the coverage of local lg sub-games of the global gG games.*
2.8.3. Then, following St.1. and Ass.1., it is reasonable to expect that properties common in rg cognizing, at least, approximate *hcogs* cognizing of the Universe U, allowing corollary Cl. 1.2. to proceed to the following assumption:
Ass. 2. *The theory of* hcogs *cognizing of games of rg class based on common properties of the means and outputs of rg cognizing can approximate adequate models* $hcogs_U$ *of* hcogs *cognizing of* U.
2.9. Consequently, it is worth the addressing the impacts on generalized cognizing by cognizing of reproducible games.

Thus, succeeding Ass. 2., it is reasonable to reveal common properties of means and outputs of rg cognizing, followed by conveying rg common of them to the models approximating cognizing the Universe.

3. Impacting to the Means of Generalized Cognizing by Cognizing Local Games

3.1. To achieve these goals, let's preliminarily acknowledge that generally and from rg experimens[4] –[31] follows that
St. 2. *Rg games differ semantically in the attributes of situations of the trees and the cause-effect rules of transitions between them, Therefore, they generally, require cognizing approaches appropriate to particular games,*

while to state

St.3. *Rg games can be syntactically modeled using game trees and have similar strategy search algorithms induced by the structures of these trees.*

3.2. Thus, it follows that, at least, strategy-search-based properties of cognizing are rg common and studying such rg cognizing can be arranged for each rg representative, **rg-kernel**, followed by dissemination of the results to the entire rg class, then, moreover, to approximate adequate models of hcogs cognizing of U ($hcogs_U$).

3.3. Then, rooting in Zermelo's studies [18] and those, particularly, in [4]-[8], [10, 11], [16, 17], [19]–[31], chess historically is acknowledged as convenient model, i.e., rg-kernel, for revealing common properties of means and outputs of rg cognizing.

Resuming, it follows that

Ass. 3. *Strategy–search–based properties of cognizing are rg common and studying rg cognizing can be arranged for the chess rg-kernel, followed by dissemination of the results to the entire rg class.*

3.4. Correspondingly, let's list some chess **properties common in rg cognizing.**

St. 4. *Chess games can be approximated by finite game trees.*

St. 5. *Chess trees comprise all possible strategies of chess.*

St. 6. *Situations of chess trees are strongly divided into three classes: winning, losing, or drawing.*

St. 7. *Comprehensive outputs of cognizing chess, or comprehensive knowledge of chess (cck) encompass classifiers of values of each chess situation and corresponding cause-effect chains, strategies, for attaining these values, thus, allowing to play in absolutely effective ways wrt all other strategies.*

3.5. Consequently, along with commonality of St. 2–7 let's comprise properties so far gained in cognizing of chess, properly rg extend them, followed by focusing on ongoing and intended chess studies arguably relevant to the theory of $hcogs_U$.

St. 8. *Ideally, there are perfect chess solvers with strategies possessing comprehensive knowledge of chess, cck.*

3.5.1. Consequently,

St.9. *These perfect chess solvers become the winners of ideal absolute tournaments between all chess solvers, ordered wrt points each solver gained in games with all others.*

3.5.2. Extending St. 8. allows us to assume

Ass. 4. *The ordering of chess solvers by points gained in absolute tournaments is in accordance with the decreasing of the quality and quantity of the fractions of cck they possess, thus*

Ass. 5. *The degrees of success of rg solvers are directly proportional to the degrees of quality and quantity of the fractions of comprehensive chess knowledge they possess, and therefore, directly proportional to the progress in cognizing of cck by cognizers cogs of rg solvers.*

3.5.3. Ass. 4–5 are supported by analysis of a wide-ranging repository of chess vocabulary units [10] allowing us to state:

St.10. *Classifiers of the values of chess situations, plans and strategies of chess experts can be interpreted as classifiers of constituents of comprehensive chess knowledge cck and their qualification degrees correspond to the quality and quantity of the fraction of cck they possess.*

3.5.3.1. Another argument supporting Ass. 4–5 follows from Elo's analysis of game results of chess players, allowing him to introduce a measure, Elo's rating, qualification of players

wrt absolute tournaments, while measuring it by means of only minor complexity local tournaments between them [4].

3.5.3.2. In parallel, introducing in [4] a hypothesis analogous to Elo's one, based on proven cck theorems, stating, particularly, that in classes A of algorithms allowing enumeration of all possible appearances of algorithms of A by enumeration of states of variables, regulators, determining any of their appearances, it is possible to arrange the sequences of local tournaments between all possible appearances of algorithms, similar to those used by Elo in chess, converging to the strongest algorithms in A.

3.6. The analysis of the repository of chess vocabulary units also lets us accept the following statement:

St. 11. *The likelihood of expert classifiers wrt the ideal ones, as a rule, is small because precise determination of the values of the realities requires generating and evaluating the corresponding strategies, which are computationally hard problems.*

And, consequently, we accept the following assumption:

Ass.6. *Values of chess classifiers, in principle, cannot be even for all experts, and classifiers of experts along with common constituents, which include an essential part of subjective ones.*

3.7. Applications and experiments with competition, defense and dialogue rg problems rooted in cognizing of chess and summarized in [10], can be stated as follows:

St.12. *Properties by St. 1–11 and Ass. 1–6 revealed in cognizing chess can presumably be extended to the entire rg class by properly replacing chess expressions in them with those of targeted rg representatives and chess values with corresponding utilities.*

3.8. Thus, the corollaries of St. 12. are as follows:

Cl.12.1. *Humans* hchcogs *cognizers of chess can properly approximate* hcogs *rg cognizing, thus also approximating the adequate models of* $hcogs_U$.

Cl.12.2. *The theory of chess cognizing, based on the models* hchcogs, *can approximate the entire constructive theory of cognizing.*

3.9. Let's now recall the theorems on the progress of chess cognizers by local tournaments, as provided in Appendix 1, a trajectory of their less sophisticated proof, followed by conclusions and some actual targets for further research.

4. Progress of Chess Cognizers in Local Tournaments

4.1. Interpreting Elo's hypothesis in [8], the theorems on ordering chess strategies by local tournaments were induced based on the assumption that the proportions of wins, draws, and losses of a given control search strategy f in rg games with arbitrary other strategies, given a fixed initial position P, are directly proportional to the proportions of terminal vertices of the corresponding types in the strategy generated by f from P.

4.1.1. Consequently, the problem of comparing the "strength" of two given strategies, f_i and f_j, is reduced to the problem of establishing the relative positions of these strategies in a sequence obtained by linearly ordering the chess strategies based on the results of an absolute tournament. The latter is understood as a round-robin tournament between all possible chess strategies, where each pair of strategies meets in a match under special conditions that ensure that no a priori advantage arises for either strategy due to the nature of the match organization.

4.1.2. Then, by analogy with Elo, a hypothesis was put forward on interdependencies of knowledge-based strategies in absolute tournaments, stating that based on the position i of given strategy f_i in an absolute tournament, a constant b_i can be specified such that f_i only

loses games to the left of the neighborhood $(i + b_i, i - b_i)$, wins to the right, and it can both win and lose within the neighborhood. The following theorem was proved:

Theorem 1 (page 152, [4]). *For arbitrary strategies f_i and f_j, if f_j wins the game against f_i and there exists a set of strategies F' such that*

- *the power of F' is greater than b_i*
- *f_j wins and f_i loses games with every strategy in F' in the round-robin, absolute tournament for $F' \cup \{f_i, f_j\}$,*

then f_j is stronger than f_i wrt the absolute tournament between all chess knowledge–based strategies.

A similar statement holds for f_i and f_j games that end in draws.

4.1.3. Thus, the question of testing whether a given strategy f_j strengthens the original f_i reduces to constructing the set F' and estimating the parameter b_i.

Even without estimating b_i, it can be stated: the larger $|F'|$, the more likely it is that f_j is stronger than f_i. This last statement provides the necessary foundation for an experimental study of strategy strengthening.

It should also be noted that, following Ass.3, these results can be extended to the entire rg strategies, too.

4.2. In [4], the solvability of the Shannon chess adaptation problem (SAP) was also proven, interpreted as a requirement to arrange the sequences of local tournaments between chess algorithms of classes A analogous to Elo ones for chess players, which under suitable restrictions on the locality, i.e., on the number $\Delta(t)$ of elementary calculations in local tournaments at time t, will converge to the strongest algorithms in A.

4.2.1. Solvability was proven without requiring a strictly sequential strengthening of hypotheses of A in tournaments and assuming that samples of A could be provided by some algorithms with a finite number l of parameters, *regulators* of the provisions.

The latter allows one to construct a procedure for enumerating all combinations of algorithms of A and sequentially checking them for the presence of algorithms stronger than the current algorithms – a hypothesis, to converge in the limit to the strongest of the algorithms in A with a certain accuracy.

4.2.2. More precisely, let c be the maximum number of elementary calculations when performing arithmetic operations or comparing two numbers, and let c_m and c_g be those required to play one game and transit to algorithms from their numbers, respectively.

4.2.3. Let's also accept that algorithms ω resolve the SAP in classes A with accuracy m if, in the limit, the solutions $\omega(A)$ are weaker than the best algorithms in A at no more than m places wrt absolute tournaments.

Under these conditions, Theorem 2 was proved

Theorem 2 (page 158, [4]). *The SAP adaptation problem in classes A of chess algorithms generated by l regulators is solvable with accuracy $2b$ if, at least,*

$$\Delta(t) \geq 2c_m b^2 + 5cb(1 + \log_l(t + b + 2)) + c_g b$$

4.3. The proof of Theorem 2 relies on the technique of producing sequences of compacted and not–coinciding k–element subsets of the finite n naturals, $k < n$, presented in [32] and [10].

Avoiding the complexity of interpretation of this technique presented in [4], while demonstrating the viability of the theorems, Chapter 6 provides an outline of a parallel trajectory of the basics of the proof for the chess rg kernel.

5. Conclusions and Targets

5.1. Theorem 1 provides an explicit, simple measure of ascending order of a group of game strategies in scales induced by tournaments among all possible strategies in the focused game trees.

In contrast to conservative situation trees of chess, such rg and cogs trees can vary depending on the revealed and dated classifiers of situations and relations between them.

Nevertheless, these trees, as a rule, remain stable for some periods of time, allowing for the development, analysis, and comparison of strategies based on the aforementioned local tournaments.

5.2. Particularly, by extending Theorem 2 and assuming that the class of algorithms A contains the highest generalized cognizers $cogs^*$ the search for $cogs^*$, analogously to chess, can be arranged by enumerating possible states of leverages of algorithms of A along with the technique for extracting all possible bundles b from A of these algorithms, thereby guaranteeing the convergence of the winner cogs of the tournaments to $cogs^*$.

5.2.1. A step to $cogs^*$ could provide the acquirers of the thesauri of communities, as well as their enrichment through discovery, revealing and other leverages for gaining new mss. Particularly, a generative pre-trained transformer such as ChatGPT is worth considering for these aims.

5.3. The evolutionary directed development of diversified versions of cogs, particularly, octaves to $cogs^*$, using measurements obtained in local tournaments, also appears promising.

5.3.1. In chess interpretation, given any starting chcogs C_o with certain functional units, experts preserving the root utilities of C_o can diversify the performance of the units of C_o and form the class $\{C_o\}$ of chcogs.

Applying the package of evolutionary transformers T_e, including parenting and mutations, to $\{C_o\}$, they will get new class $\{C'\}$ of chcogs.

Arranging local tournaments T_r for bundles of $\{C'\}$ experts will find the class of winners $\{C'_w\}$ wrt T_r.

Applying T_r to $\{C'_w\}$, they will get a new class $\{C''\}$ to continue the process recurrently.

Expectedly, such an evolution process will converge to the best chcogs Ch^* in the classes created in this way.

5.3.2. For developing octaves O_{ch}, the experts can arrange the learning and growth of O_{ch} in a variety of environments E_i to get the corresponding class $\{C_1\}$ of chcogs.

Through tournaments T_r in $\{C_{1w}\}$, winners in $\{C_1\}$ can be found to continue the process recurrently, while these learning and evolutionary approaches can be combined.

5.4. Concluding, let's recall the assumption that possibilities of humans to reason about the Universe U and prognosticate decisions at time t are bounded by ad hoc cognized classifiers and relationships, or ad hoc cognizable Universe (ahcU).

Then, it was argued that humans hcogs cognizing of the Universe U can be modeled by cognizing local sub-trees T_{s_i} of situations of the global tree gT_s altogether covering the cognition of gT_s.

Thus, given trees T_{s_i} of situations of problems p_i with target situations s_s, ideally, the perfect strategy S_i^* for p_i from s_s at time t can be calculated for a certain depth h of T_{s_i}.

Despite calculations of S_i^* increases exponentially with the rise of h, and hence the power of S_i^* is bounded by the power of computers, this approach can be viable since in applications p_i the depth h is often restricted, while the power of computers increases tremendously.

6. Appendices: A Trajectory of Proving Theorems 1–3

6.1. Recall that in the SAP *Shannon's chess adaptation problem*, it is required to arrange the sequences of local Elo's type tournaments between chess algorithms of A, which, under suitable restrictions on the locality, i.e., on the number $\Delta(t)$ of elementary calculations in local tournaments at time t, will converge to the strongest algorithms in A.

6.1.1. Let us denote that

chG – chess game tree

h – the depth of game trees

n – branches of the nodes

n^h – number of strategies of chG.

6.1.2. Then,

l – the number of leverages, regulators inducing cognizers of chG

m_i – the range $[a_{i1}, a_{i2}, \ldots, a_{im_i}]$ of states of the x_i leverage, or x_i variable, $i \in [1, l]$ and $m \geq m_i$

A – the set of chess solvers, algorithms (chalgs) l-produced and stabilized in learning chG versions of chess cognizers chcogs

m^l – the number of possible l-produced chalgs, thus, the number of elements of A.

6.1.3. Denote also d and b as naturals such that $d > 2b$ and b is the length of the zone of uncertainty of i-th player in ordering according to the absolute tournaments of all cognizers of A

$\binom{m^l}{d} = \frac{m^l!}{(m^l-d)!\,d!}$ – a combination of m^l things d at a time is the number of local d size tournaments in examining the strength of players.

6.1.4. Each chalgs realizes a strategy in chG, why the number of different chalgs can be, at least, even to n^h in the case if l-produced chalgs of A realize all strategies of chG, i.e., $m^l < n^h$ or $m^l = n^h$.

6.1.5. Consider enumeration E_1 of l-positioned numbers corresponding to the indices i of x_i, $i \in [1, \ldots, m_i]$, which sequentially lists all n possible l-bundles of indices of states of x_i variables, thus, all l-bundles of chalgs of A, starting from the number $1, 1, \ldots, 1$ with 1 in all l positions.

For example, if $m_i = 10$, of l-bundles of indices $b_1, b_2, \ldots, b_L$ in decimal scale of enumeration E_1 will start with the decimal number $10^l + 10^{l-1} + \ldots + 10 + 1$ and, generally, correspond to the decimal number $b_l 10^l + b_{l-1} 10^{l-1} + \ldots + 10 + b_1$.

If the scales are not decimal, i.e., $i \in [1, 2, \ldots, m_i]$, the decimal number of l-bundle $b_1, b_2, \ldots, b_l$ will be equal to $b_l(m_{l-1} \ldots m_2 m_1) + b_{l-1}(m_{l-2} \ldots m_2 m_1) + \ldots + b_2 m_1 + b_1$ and the number n of all l-bundles will be equal to $m_L \ldots m_2 m_1$.

Apparently, chalgs of A get unique numbers, code in E_1, and can be unanimously decoded.

6.2. Let T_A be the ordering of all chalgs of A as the result of the ratings they gain in an absolute tournament for all chalgs of A, where each chalgs competes against all the others in A.

Let d, b also be naturals such that $d > 2b$, where b is the max number of places on the left and right of any i position in T_A that games of i with j in the range of $2b$ are uncertain by the results, i.e., i either wins j or loses j [4].

6.3. For the set E_1 of $1, 2, \ldots, n$ of all chalgs of A, let E_2 be a compacted, lexicographic enumeration of all possible d-bundles of different, not coinciding chalgs of A.

Addressing the definitions in [4] for E_2 lexicographic enumeration, let's provide an example of it for $n = 5$ and $d = 3$: 123, 124, 125, 134, 135, 145, 234, 235, 245, 345.

6.4. Following E_2, let's sequentially take d-bundles of E_2, for each b_t of them provide an exhaustive tournament, i.e., each chalgs competes with all others of b_t, then in the ordering o_{b_t} of strategies of these chalgs to examine, whether o_{b_t} meet the requirement that the winner w_{b_t} of o_{b_t} won each other in b_t, while the min rating player l_{b_t} lost to any of b_t.

According to [4] it follows that w_{b_t} has a higher position in the absolute T_A tournament than l_{b_t}.

Thus, in the search for the best s^* in A, we can exclude l_{b_t} from further examination.

6.5. To examine the viability of w_{b_t} as the best chalgs, let's provide tournaments involving w_{b_t} and, sequentially, each d-bundle of E_2 until the $d+1$-bundle $b'_{t'}$ is found, where the winner $w'_{b'_{t'}}$ with max rating wins all others in $b'_{t'}$, while w_{b_t} loses to all of them.

Thus, $w'_{b'_{t'}}$ in T_A is higher (stronger) then w_{b_t}, so that w_{b_t} can be excluded from further examination.

6.6. Continuing the above series of local tournaments, eventually we'll find $\hat{w}_{\hat{b}_{\hat{t}}}$ that, during exhaustive searching of E_2, will not be defied by chalgs of $\hat{b}_{\hat{t}}$, and thus can be considered as an approximation to s^*.

6.7. Along with looking for new selection criteria guaranteeing s^*, heuristics can be applied to indicate chalgs better than $\hat{b}_{\hat{t}}$.

For example, we can weaken the requirement that $\hat{w}_{\hat{b}_{\hat{t}}}$ is beaten by all chalgs of $\hat{b}_{\hat{t}}$ and trade off different options for the % of such events.

References

[1] R. Kurzweil, *The Singularity Is Closer*, 2024.

[2] B. Goertzel, "Singularity can become real already in 2027," *Beneficial AGI Conference*, 2024 [Online]. Available: `https://hightech.plus/2024/03/07/`

[3] Evolutionary algorithms, Wikipedia, [Online]. Available: `https://en.wikipedia.org/wiki/Evolutionary_algorithm`

[4] E. Pogossian, *Adaptation of Combinatorial Algorithms*, Academy of Sciences of Armenia, 1983.

[5] E. Pogossian, "Management strategy search and assessment programming", *Proceedings of International Conference Computer Science and Information Technologies*, Armenia, 1999.

[6] E. Pogossian, "On assessment of performance of systems by combining on-the-job and expert attributes scales," *International Conference in Computer Sciences and Information Technologies*, pp. 331–335, Yerevan, Armenia, 2015.

[7] E. Pogossian, "Standardizing measurements of performance of human and technical systems," *Proceedings of International Conference 'Current Issues of Psychology of Management'*, pp. 354–362, Yerevan, September 10–11, 2017.

[8] Elo rating system, Wikipedia, [Online]. Available: `https://en.wikipedia.org/wiki/Elo_rating_system`

[9] J. Flavell, *The Developmental Psychology of Jean Piaget*, D. VanNostrand Comp. Inc., Princeton, N.j., 1962.

[10] E. Pogossian, *Constructing Models of Being by Cognizing*, Academy of Sciences of Armenia, Yerevan, Armenia, 2020.

[11] E. Pogossian, V. Vahradyan, and A. Grigoryan, "On competing agents consistent with expert knowledge," *Lecture Notes in Computer Science*, AIS-ADM-07, St. Petersburg, Russia, pp. 229–241, 2007.

[12] H. Blavatsky, *Between Light and Darkness*, Zhizn zamechatelnykh lyudei (in Russian), 2012.

[13] C.-L. Chang and R. C.-T. Lee, *Symbolic Logic and Mechanical Theorem Proving*, Academic Press, San Diego, 1987.

[14] S. Wolfram, "Foundations of biological evolution: more results, more surprises," 2024. `https://writings.stephenwolfram.com/2024/12/foundations-of-biological-evolution-more-results-more-surprises/`

[15] E. Pogossian, "Effectiveness enhancing knowledge-based strategies for SSRGT class of defense problems," in *NATO ASI 2011, Prediction and Recognition of Piracy Efforts Using Collaborative Human-Centric Information Systems*, Salamanca, Spain, pp. 16, 2011.

[16] C. Shannon, "Programming a computer for playing chess," *Philosophical Magazine*, Ser. 7, vol. 41, 1950.

[17] A. Newell, J. C. Shaw, and H. A. Simon, "Report on a general problem-solving program," *Proceedings of the International Conference on Information Processing*, pp. 256–264, 1959.

[18] E. Zermelo, "Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels," *Proceedings of the Fifth International Conference of Mathematicians*, Cambridge University Press, pp. 501–504, 1912.

[19] Ch. Brutyan, I. Zaslavski, and L. Mkrtchyan, "On methods of automated synthesis of positional strategies in games," *Problemi Kibernetiki*, vol. 19, pp. 41–75, 1967.

[20] M. Botvinnik, *About Solving Approximate Problems*, S. Radio, Moscow (in Russian), 1979.

[21] R. Benerji, *Theory of Problem Solving*, Mir, Moscow, 1972.

[22] J. E. Laird, A. Newell, and P. S. Rosenbloom, "Soar: an architecture for general intelligence," *Artificial Intelligence*, vol. 33, no. 1, pp. 1–64, 1987.

[23] J. Laird, *The Soar Cognitive Architecture*, MIT Press, England, 2012.

[24] S. V. Grigoryan and Z. H. Naghashyan, "Adequacy and application of models of cognizing by combinatorial games," *MPCS*, vol. 62, pp. 25–42, Dec. 2024.

[25] S. Grigoryan, "Automating acquisition and explanation of strategy knowledge," *Proceedings of CSIT Conference*, pp. 21–23, 2021.

[26] B. Karapetyan, "A high-level language for chess concepts," *Mathematical Problems of Computer Sciences*, 1986.

[27] E. Pogossian, "Specifying personalized expertise," *CELDA*, pp. 151–159, Barcelona, Spain, 2006.

[28] B. Karapetyan, "Classification of incompleteness of strategies," *Mathematical Aspects for Controlling Conflicts*, pp. 45–48, Yerevan, 1987.

[29] R. Simonyan, "On complexity of antagonistic completely informed finite games," *Mathematical Aspects for Controlling Conflicts*, pp. 74–78.

[30] R. Simonyan, "On complexity of specification Nim type strategies," in *Mathematical Aspects for Controlling Conflicts*, pp. 78–81, Yerevan, 1987.

[31] S. Grigoryan, "Research and development of algorithms and programs of knowledge acquisition and their effective application to resistance problems," PhD Thesis, Yerevan, Armenia, 2016.

[32] E. Pogossian, "Systems of incomparable sets with the smallest number of subsets," PhD Thesis, Computing Center of Academy of Sciences, Moscow, 1970; also in *Math. Problems of Cybernetics and Comp. Science*, vol. XVI, pp. 148–161, Yerevan, 1986 (in Russian).

Ընդհանրացված իմացիչների համար չափելի մրցաշարերի մասին

Էդվարդ Մ. Պողոսյան

ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ
Երևան, Հայաստան
e-mail: epogossi@aua.am

Ամփոփում

Հետազոտողներ Կուրցվեյլը և Գյորտցելը, կանխատեսում են, որ արհեստական բանականությունը (AI) մտնում է արհեստական ընդհանուր բանականության (AGI) էքսպոնենցիալ աճի շրջան։

Նրանք կարծում են, որ եթե նման AGI-ը կարողանա վերաշարադրել իր սեփական կոդը, այն կարող է զարգանալ որպես գերմարդկային արհեստական բանականություն, որը կունենա բոլոր մարդկային քաղաքակրթությունների ճանաչողական և հաշվողական ուժը։

Մենք AGI-ի այս պահանջը մեկնաբանում ենք որպես ընդհանրացված իմացիչների անխուսափելի ունակություն՝ ստեղծելու իրենց փոփոխությունների սպասելիորեն ողջամիտ տարբերակներ՝ այդ տարբերակները չափելու վստահելի միջոցներով և դրանցից ամենախոստումնալիցները ընտրելու ճանաչողների օգտակարության վերաբերյալ։

Ընդհանրացնելով մեր նախորդ պնդումը, որ գիտելիքների վրա հիմնված շախմատի ռազմավարությունները կարող են ուժեղ կերպով մասշտաբավորվել տեղական մրցաշարերի միջոցով, մենք հիմնավորում ենք նմանատիպ պնդում՝ ընդհանրացված իմացիչների առաջընթացի համապատասխան չափման համար՝ իմացիության համարժեք տեսության շրջանակներում, և անդրադառնում ենք դրա զարգացման ուղիներին։

Բանալի բառեր՝ AGI, ընդհանրացված իմացիչներ, առաջընթացի չափում, տեղական մրցաշարեր, բացարձակ կարգավորումներ։

Об измеримых турнирах для прогрессирующих обобщенных познавателей

Эдуард М. Погосян

Институт проблем информатики и автоматизации НАН РА, Ереван, Армения
e-mail: epogossi@aua.am

Аннотация

Исследователи, такие как Курцвейл и Гертцель, предсказывают, что ИИ, благодаря прогрессу LLM, вступает в период экспоненциального роста в сторону искусственного общего интеллекта (AGI).

Они считают, что если бы такой AGI мог переписывать свой собственный код, он мог бы эволюционировать в сверхчеловеческий ИИ, обладающий когнитивной и вычислительной мощью всей человеческой цивилизации.

Мы интерпретируем это требование к AGI как неизбежную способность обобщенных познавателей генерировать ожидаемо разумные версии модификаций самих себя, с надежными средствами измерения этих версий и выбора наиболее перспективных из них, с точки зрения полезности познавателей.

Обобщая наше предыдущее утверждение о том, что основанные на знаниях шахматные стратегии могут быть сильно масштабированы локальными турнирами, мы выдвигаем аналогичное утверждение для подходящего измерения прогресса обобщенных познавателей в рамках адекватной теории познания, и рассматриваем способы ее усиления.

Ключевые слова: AGI, обобщенные познаватели, измерение прогресса, локальные турниры, абсолютный порядок.

Mathematical Problems of Computer Science 65, 63–71, 2026.

doi:10.51408/1963-0146

UDC 004.4

Matrix-Vector Multiplication Performances in Multi-Accelerator Architectures

Edita E. Gichunts

Institute for Informatics and Automation Problems of NAS RA, Yerevan, Armenia
e-mail: editagich@iiap.sci.am

Abstract

This paper presents the performance of symmetric and Hermitian matrix-vector multiplication on two Volta 100 graphics processors in single and double precision. The implementations were performed using the Magma library.

The goal of this work is to present the implementations and performance evaluations of symmetric and Hermitian matrix-vector multiplication on 2 GPUs, and to compare their performance with that of a 1-GPU implementation.

Keywords: Multiple GPUs, MAGMA, Matrix-Vector Multiplication.

Article info: Received 24 Murch 2026; sent for review 9 April 2026; receieved in revised from 30 April 2026; accepted 8 May 2026.

1. Introduction

For high-performance computing, architectural solutions are being developed that facilitate work in this environment. In recent years, general-purpose graphics processors have emerged as a prominent topic in high-performance computing. Calculations on GPUs have been and are being developed very quickly, thanks to the CUDA (Compute Unified Device Architecture) [1] platform developed by Nvidia. CUDA is a software-hardware technology, based on the C programming language, along with its compiler and libraries, which is available to all developers.

The popularity of hybrid GPU-based systems began with the introduction of the NVIDIA CUDA architecture and the extension of the standard programming languages C, C++, and Fortran, which simplified GPU programming, allowing developers to leverage the computational power of modern GPUs. The hybrid architecture combines the advantages of shared- and distributed-memory architectures.

The development of computer architecture was followed by software libraries. Computational efficiency in linear algebra is a significant challenge, making optimized program implementations necessary. Linear algebra libraries are crucial in addressing this issue.

In the mid-1960s, IBM released the Scientific Routine Package [2], a set of FORTRAN routines. In 1974, Harwood published EISPACK [3], a FORTRAN routine package designed to

compute the eigenvalues and eigenvectors of matrices. Additionally, BLAS (Basic Linear Algebra Subroutines), the first product of the ACMSIGNUM joint project, was developed during the 1973-1977 period [4]. The LINPACK library was introduced in 1979 as a collection of subroutines specifically designed for the supercomputers of the 1970s and 1980s, particularly those employing vector processors, primarily for solving linear equations and linear least-squares problems. LINPACK (HPL) [6, 7] also enables evaluation of the most powerful supercomputers, as ranked by the TOP500 [8]. The initial version of BLAS (BLAS Level 1) implemented scalar-vector and vector-vector operations. In 1988, BLAS2 (BLAS Level 2) was developed as an extension of BLAS1 to exploit vector processors' capabilities [9, 10]. BLAS2 enables matrix-vector operations.

In 1990, another extension was added to BLAS3 [11, 12]. This extension accounted for advancements in computer memory by implementing matrix-matrix operations.

LAPACK [13], released in 1992, replaced LINPACK and EISPACK, providing better performance. LAPACK focuses on solving systems of linear equations, linear least squares problems, eigenvalue problems, and uniqueness problems. To perform these operations, it also carries out related calculations such as matrix analyses (LU, QR, LDLT, Cholesky, etc.).

For GPUs, NVIDIA offers CuBLAS [14], an implementation of BLAS in NVIDIA CUDA that encompasses all three levels.

The MAGMA [15] project aims to develop a linear algebra library, similar to LAPACK, for heterogeneous/hybrid architectures, starting with modern Multicore + GPU systems. MAGMA also includes MAGMA BLAS, which complements the CUBLAS subroutines.

For linear algebra problems in multi-GPU architectures, the cuBlasXt library [16] and MAGMA library subroutines are used, designed for multi-GPU systems. It should be noted that the cuBlasXt library includes only the BLAS3 level, i.e., matrix-matrix operations. Therefore, in this work, we use the MAGMA library, which also includes several subroutines for multi-GPU systems.

This paper outlines the algorithmic steps for implementing symmetric and Hermitian matrix-vector multiplication on two graphics processors. It also provides performance evaluations for matrices and vectors with dimensions of up to 40,000. Additionally, the paper presents estimates of the performance differences when using a single graphics processor.

The obtained results expand the possibilities of applying GPU computing technologies to high-performance computing tasks and can be used in the development of new parallel algorithms and engineering computations.

The results of this work can be applied:

- In high-performance computing (HPC) systems;
- In the development of software for scientific simulation;
- In artificial intelligence and machine learning tasks;
- In engineering and physical computations;
- In big data processing;
- In supercomputing systems and computational complexes.

2. Steps to Implement a Matrix-Vector Multiplication in Multiple-GPU Architecture

Linear algebra problems are crucial in high-performance computing, with matrix-vector multiplication being one of the most widely used operations. High productivity in these computing systems is often achieved through the use of linear algebra libraries. It is important to note that the

matrix-vector multiplication is classified as a Level 2 subroutine in the BLAS library, which is not included in cuBlasXt. Consequently, in multi-accelerator architectures, the magmablas_xmv_mgpu() subroutines from the MAGMA magmablas library are utilized for performing the matrix-vector multiplication. Additionally, when referring to symmetric or Hermitian matrices, specific abbreviations are indicated instead of 'x.'
The objective of this work is to implement this multiplication across multiple GPUs. When these subroutines are called, the matrix-vector multiplication is performed in parallel across all GPUs simultaneously. The model used in this work describes how the matrix and vector are distributed across each GPU.
Since the result of the matrix-vector multiplication is a vector, for example, the first element is obtained from the multiplication of the first row of the matrix and the vector, the second element is obtained from the multiplication of the second row of the matrix and the vector, and so forth, then the parallelization model used in this problem is as follows: the matrix is divided into as many parts as there are GPUs, and each GPU receives the divided part of the matrix A along with the entire multiplication vector. In this model, each graphics processing unit (GPU) generates a vector element as many times as there are rows in the partitioned matrix.
Below are the steps for implementing a matrix-vector multiplication in a hybrid system with multiple accelerators, indicating the main subroutines involved.

Here, we outline the algorithmic steps for implementing single-precision Hermitian matrix-vector multiplication.

1. We include the following necessary header files:

```
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <math.h>
#include <cuda.h>
#include <cuda_runtime_api.h>
#include <magma.h>
#include <magma_v2.h>
#include "magma_lapack.h"
#include "flops.h"
#include "magma_cbulge.h"
#include "magma_threadsetting.h"
#include "magma_operators.h"
#include <magma_internal.h>
#include "magma_timer.h"
```

2. The MAGMA library is initialized:

```
magma_init().
```

3. Memory is allocated on the CPU for the matrix and vectors. Additionally, memory is allocated for the final result vector transferred from the GPU to the CPU, where hwork serves as an extra workspace on the CPU allocated with the lhwork dimension:

```
magma_cmalloc_cpu(&A,  matsize );
magma_cmalloc_pinned(&X, vecsize );
magma_cmalloc_cpu(&Y,  vecsize );
magma_cmalloc_cpu(&Ymagma, vecsize );
magma_cmalloc_pinned(&hwork, lhwork ),
```

where hwork serves as an extra workspace on the CPU allocated with the lhwork dimension.

4. On the GPU, memory is allocated for the transferred matrix, vector, and result vector:

```
magma_setdevice(opts.device );
```

magma_cmalloc(&dA, matsize);
magma_cmalloc(&dX, vecsize);
magma_cmalloc(&dY, vecsize).

On GPUs, local memory is allocated to the partitioned sections of the matrix, moving cyclically from one GPU to another. In each instance, the magma_setdevice(dev) function is called first, followed by the memory allocation functions:

magma_cmalloc(&d_lA[dev], ldda*n_local[dev]);

magma_cmalloc(&dwork[dev], ldwork), where dwork is an additional workspace on the GPU with dimension ldwork.

Note that n_local = ((n / nb) / ngpu + 1) * nb.

5. Since the result of the matrix-vector multiplication is a vector, the first element of which, for example, is obtained from the product of the first row of the matrix and the vector, the parallelization algorithm is as follows: the matrix is divided into as many parts as there are GPUs, and each GPU is sent the divided part of the matrix A. The transfer of the matrix A is performed using the following function, which sends the matrix A from the CPU memory to the GPU dA, which is distributed cyclically across multiple GPUs into 1D row blocks.

magma_csetmatrix_1D_col_bcyclic(Noffset, Noffset, A, lda, d_lA, ldda, opts.ngpu, nb), where

Noffset = N + offset;
offset = min(n,nb).

6. We fix the initial time using the gpu_time = magma_sync_wtime(0) function.

7. The subroutine

magmablas_chemv_mgpu(opts.uplo, N, alpha, d_lA, ldda, offset, X + offset, incx,

beta, Ymagma + offset, incx, hwork, lhwork, dwork, ldwork, opts.ngpu, nb, queues) is called.

Note that in the program, we have previously introduced all the required parameters to be included in the subroutine.

This subroutine calculates the operation y = alpha*A*x + beta*y. In the case of alpha=1 and beta=0 values, we have the matrix-vector multiplication y = A*x.

8. Using the gpu_time = magma_sync_wtime(0) - gpu_time difference, we get the calculation execution time.

9. After completing the calculations, the results obtained from the GPUs are transferred to the CPU memory using the following function:

magma_cgetvector(Noffset, dY, incx, Ymagma, incx):

10. At the end of the program, the allocated memory on the CPU is cleared:

magma_free_cpu(A);
magma_free_cpu(Y);
magma_free_pinned(X);
magma_free_cpu(Ymagma);
magma_free_pinned(hwork).

11. We clear the allocated memory on GPUs by moving from one GPU to another in a loop, first calling magma_setdevice(dev) function, then magma_free(d_lA[dev]) and magma_free(dwork[dev]) functions.

We also clear the memory allocated for the moved vectors:

magma_free(dA);
magma_free(dX);
magma_free(dY).

12. We finalize MAGMA processing using the magma_finalize() function.

3. Experimental Results

The research was conducted on two NVIDIA Tesla V100-PCIE graphics processors using the Magma 2.6.0 library. To install the MAGMA 2.6.0 library, the BLAS, LaPack, cLaPack, ATLAS libraries, as well as the following static (.a), dynamic (.so) libraries were loaded: libgfortran.a, libf77blas.a, libcblas.a, libf2c.a, libm.a, libstdc++.a, libpthread.a, libdl.a, libcublas.so, libcudart.so, libcusparse.so, libcudadevrt.a. The gcc, g++, nvcc, and gfortran compilers were used to compile the MAGMA library.

It should also be noted that during the research, when applied to 2 GPUs, both symmetric and Hermitian matrix-vector multiplication were accessed with matrices and vectors of dimensions up to n=40000. And when applied to 1 GPU, in the case of symmetric matrices, n=30000 dimensions were accessed for single precision, and n=20000 dimensions were accessed for double precision. In the case of Hermitian matrices, n=20000 dimensions were accessed for both single and double precision.

Let us present the results from the experiments in the form of graphs and tables. The tables display the productivity values for the specified input matrix dimension.

Figures 1 and 2 show the performance graphs of the symmetric and Hermitian matrix-vector multiplication on two GPUs in single and double precision, respectively.

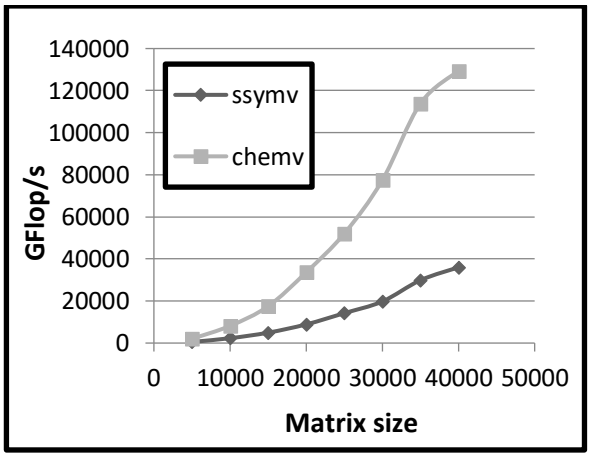

Fig. 1. Single presicion

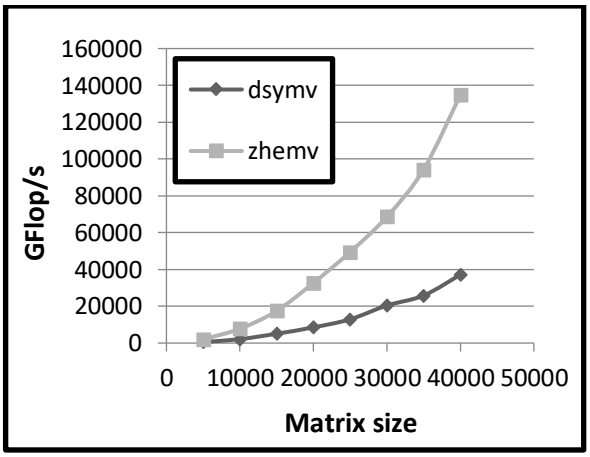

Fig. 2. Double presicion

x	ssymv	chemv
5000	531,03	1979,04
10000	2304,79	8165,32
15000	4941,26	17478,01
20000	8877,28	33641,05
25000	14208,92	52041,63
30000	19764,39	77437,18
35000	29873,08	113866,87
40000	35984,20	129371,34

x	dsymv	zhemv
5000	556,38	1979,04
10000	2107,90	7768,39
15000	5060,49	17478,01
20000	8516,78	32658,83
25000	12914,01	49464,10
30000	20405,40	68637,50
35000	25819,95	94279,68
40000	37180,36	134897,25

Figures 3 and 4 show graphs of the performance difference for symmetric matrix-vector multiplication when using one and two graphics processors in single and double precision, respectively.

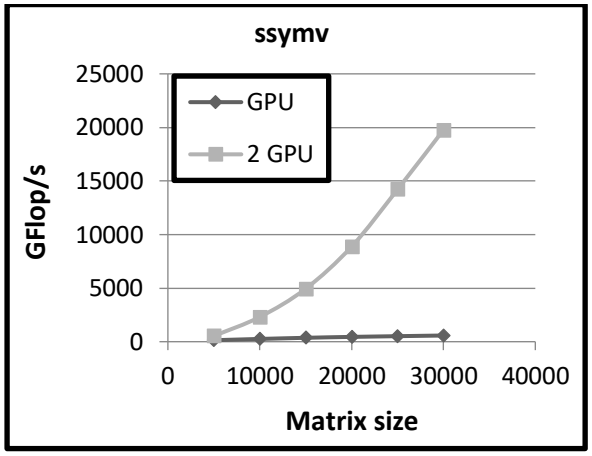

Fig. 3. Symmetrical Single precision

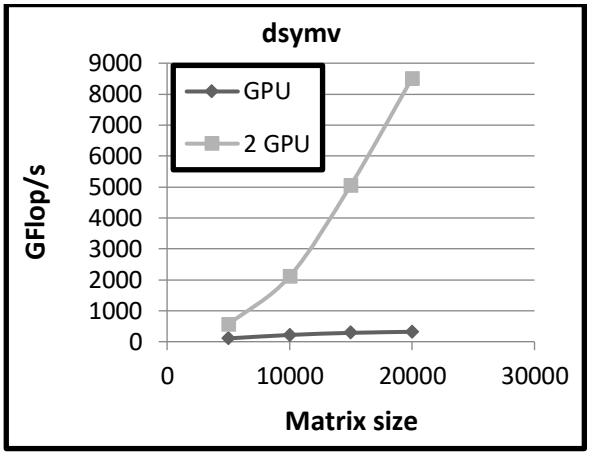

Fig. 4. Symmetrical Double precision

	ssymv	
x	GPU(GFlop/s)	2 GPU
5000	136,29	531,03
10000	256,40	2304,79
15000	362,92	4941,26
20000	448,97	8877,28
25000	514,63	14208,92
30000	577,70	19764,39

	dsymv	
x	GPU(GFlop/s)	2 GPU
5000	105,94	556,83
10000	217,40	2107,90
15000	285,73	5060,49
20000	319,52	8516,78

Figures 5 and 6 show the graphs of the difference in Hermitian matrix-vector multiplication performance when using one and two graphics processors in single and double precision, respectively.

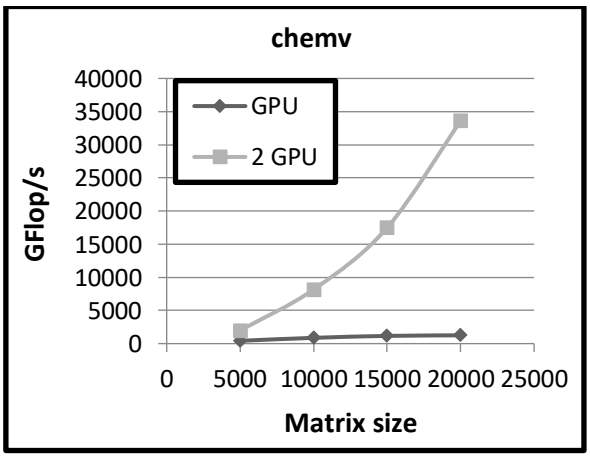

Fig. 5. Hermitian Single precision

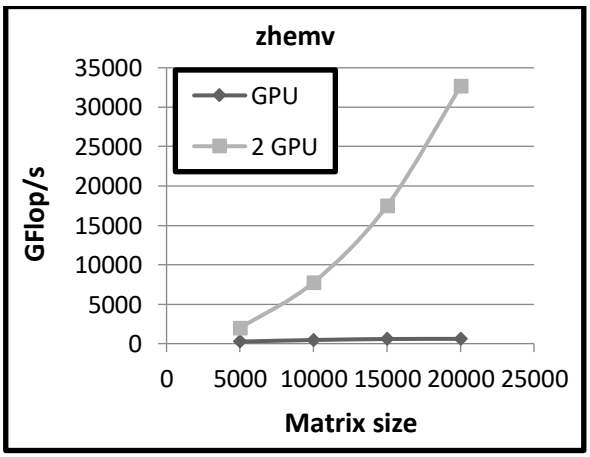

Fig. 6. Hermitian Double precision

	chemv	
x	GPU(GFlop/s)	2 GPU
5000	422,94	1979,04
10000	872,35	8165,32
15000	1163,58	17478,01
20000	1253,53	33641,05

	zhemv	
x	GPU(GFlop/s)	2 GPU
5000	258,51	1979,04
10000	459,84	7768,39
15000	584,09	17478,01
20000	815,45	32658,83

4. Conclusion

As a result of the experiments, we have obtained the following results:

- On two GPUs, in both single and double precision, the performance of the Hermitian matrix-vector multiplication is 3.5 times higher than the performance of the symmetric matrix-vector multiplication.
- In single precision, the performance of the symmetric matrix-vector multiplication of up to 30,000 dimensions on two GPUs is at least 10 times and up to 30 times higher than the performance on a single GPU.
- In double precision, the performance of a symmetric matrix-vector multiplication of up to 20,000 dimensions on 2 GPUs is at least 10 times and at most 25 times higher than that of 1 GPU.
- In single precision, the performance of a Hermitian matrix-vector multiplication of up to 20,000 dimensions on 2 GPUs is at least 10 times and at most 25 times higher than that of 1 GPU.
- In double precision, the performance of a Hermitian matrix-vector multiplication of up to 20,000 dimensions on 2 GPUs is at least 10 times and at most 40 times higher than that of 1 GPU.

References

[1] NVIDIA, “NVIDIA CUDA Parallel Computing Platform”, [Online]. Available:http://www.nvidia.com/object/cuda_home_new.html, NVIDIA, 2013.

[2] International Business Machines Corporation. System/360 Scientific Subroutine Package (360A-CM-03X) Version II, Programmer’s Manual. IBM Technical Publications Department, White Plains, NY, 1967.

[3] B. S. Garbow, “EISPACK-a package of matrix eigensystem routines”, *Computer Physics Communications*, vol. 7, no. 4, pp. 179–184, 1974.

[4] C. L. Lawson, R. J. Hanson, D. R. Kincaid, and F. T. Krogh, “Basic Linear Algebra Subprograms for Fortran Usage”, *ACM Trans. Math. Softw.*, vol. 5, no. 3, pp. 308–323, 1979.

[5] J. Dongarra, C. B. Moler, J. R. Bunch, and G. W. Stewart, LINPACK Users’ Guide, vol. 8. SIAM, 1979.

[6] J. Dongarra and P. Luszczek, “Linpack benchmark”, *Encyclopedia of Parallel Computing*, pp. 1033–1036, 2011.

[7] J. Dongarra, P. Luszczek and A. Petitet, The LINPACK benchmark: Past, present, and future. Concurrency and Computation: Practice and Experience. Concurrency and Computation: Practice and Experience, 15:2003, 2003.
[8] TOP500 Supercomputer Site. http://www.top500.org.
[9] J. Dongarra, J. Du Croz, S. Hammarling and R. J. Hanson, "Algorithm 656: an extended set of basic linear algebra subprograms: model implementation and test programs", *ACM Transactions on Mathematical Software* (TOMS), vol.14, no. 1, pp. 18–32, 1988.
[10] J. Dongarra, J. Du Croz, S. Hammarling and R. J. Hanson, "An extended set of fortran basic linear algebra subprograms", *ACM Trans. Math. Softw.*, vol.14, no. 1, pp.1–17, 1988.
[11] J. Dongarra, J. Cruz, S. Hammerling and I. S. Duff, "Algorithm 679: A set of level 3 basic linear algebra subprograms: model implementation and test programs", ACM *Trans. Math. Softw.*, vol. 16, no. 1, pp. 18–28, 1990.
[12] J. Dongarra, J. Du Croz, S. Hammarling and I. S. Duff, "A set of level 3 basic linear algebra subprograms", *ACM Trans. Math. Softw.*, vol. 16, no.1, pp. 1–17, 1990.
[13] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney and D. Sorensen, *LAPACK Users' Guide. Society for Industrial and Applied Mathematics*, Philadelphia, PA, third edition, 1999.
[14] CUDA Nvidia. Cublas library. NVIDIA Corporation, Santa Clara, California, 15, 2008.
[15] "MAGMA Matrix Algebra on GPU and Multicore Architectures", [Online]. Available: http://icl.cs.utk.edu/magma/, 2014.
[16] [Online]. Available: https://docs.nvidia.com/cuda/cublas/index.html#using-the-cublasxt-api.

Մատրից-վեկտոր արտադրյալի արտադրողականությունները բազմաքանակ արագացուցիչների ճարտարապետությունում

Է. Գիչունց

ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ
Երևան, Հայաստան
e-mail: editagich@iiap.sci.am

Ամփոփում

Այս աշխատանքում ներկայացված են սիմետրիկ և Հերմիտյան մատրից-վեկտոր արտադրյալի արտադրողականությունները երկու Volta 100 գրաֆիկական պրոցեսորների վրա՝ մեկական և երկուական ճշգրտություններում: Իրականացումները կատարվել են ՝ կիրառելով Magma գրադարանը:
Աշխատանքի նպատակն է ներկայացնել սիմետրիկ և Հերմիտյան մատրից-վեկտոր արտադրյալի իրականացումները և արտադրողականությունների գնահատականները 2

GPU-ների վրա, և ցույց տալ նաև արտադրողականության տարբերությունը 1 GPU-ի կիրառման նկատմամբ:
Բանալի բառեր՝ բազմակի GPU-ներ, MAGMA, մատրից-վեկտոր բազմապատկում:

Производительность матрично-векторных произведений в мультиускорительных архитектурах

Э. Гичунц
Институт проблем информатики и автоматизации НАН РА, Ереван, Армения
e-mail: editagich@iiap.sci.am

Аннотация

В данной статье представлены результаты оценки производительности симметричных и эрмитовых матрично-векторных произведений на двух графических процессорах Volta 100 в одинарной и бинарной точности. Реализации выполнены с использованием библиотеки Magma.

Цель данной работы — представить реализацию и оценку производительности симметричного и эрмитова матрично-векторного умножения на двух графических процессорах, а также показать разницу в производительности по сравнению с реализацией на одном графическом процессоре.
Ключевые слова: множественные GPU, MAGMA, матрично-векторное умножение

Կանոններ հեղինակների համար

ՀՀ ԳԱԱ ԻԱՊԻ “Կոմպյուտերային գիտության մաթեմատիկական խնդիրներ” պարբերականը տպագրվում է 1963 թվականից: Պարբերականում հրատարակվում են նշված ոլորտին առնչվող գիտական հոդվածներ, որոնք պարունակում են նոր՝ չհրատարակված արդյունքներ:

Հոդվածները ներկայացվում են անգլերեն՝ ձևավորված համապատասխան “ոճով” (style): Հոդվածի ձևավորման պահանջներին ավելի մանրամասն կարելի է ծանոթանալ պարբերականի կայքէջում՝ http://mpcs.sci.am/:

Rules for authors

The periodical “Mathematical Problems of Computer Science” of IIAP NAS RA has been published since 1963. Scientific articles related to the noted fields with novel and previously unpublished results are published in the periodical.

Papers should be submitted in English and prepared in the appropriate style. For more information, please visit the periodical's website at http://mpcs.sci.am/.

Правила для авторов

Журнал «Математические проблемы компьютерных наук» ИПИА НАН РА издается с 1963 года. В журнале публикуются научные статьи в указанной области, содержащие новые и ранее не опубликованные результаты.

Статьи представляются на английском языке и оформляются в соответствующем стиле. Дополнительную информацию можно получить на веб-сайте журнала: http://mpcs.sci.am/.

The electronic version of the periodical “Mathematical Problems of Computer Science” and rules for authors are available at

http://mpcs.sci.am/

Phone: (+37460) 62-35-51
Fax: (+37410) 28-20-50
E-mail: mpcs@sci.am
Website: http://mpcs.sci.am/

Ստորագրված է տպագրության՝ 29.05.2026
Թուղթը՝ օֆսեթ:
Հրատարակված է ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտի կողմից
Ծավալը՝ 73 էջ: Տպաքանակը՝ 100
ՀՀ ԳԱԱ ԻԱՊԻ Համակարգչային պոլիգրաֆիայի լաբորատորիա
Երևան, Պ. Սևակի 1
Հեռ. +(374 60) 623553
Գինը՝ անվճար

Подписано в печать 29.05.2026
Офсетная бумага.
Опубликовано Институтом проблем информатики и автоматизации НАН РА
Объём: 73 страниц. Тираж: 100
Лаборатория компьютерной полиграфии ИПИА НАН РА.
Ереван, П. Севака 1
Тел.: +(374 60) 623553
Цена: бесплатно

Signed in print 29.05.2026
Offset paper
Published by the Institute for Informatics and Automation Problems of NAS RA
Volume: 73 pages
Circulation: 100
Computer Printing Lab of IIAP NAS RA
Yerevan, 1, P. Sevak str.
Phone: +(374 60) 623553
Free of charge