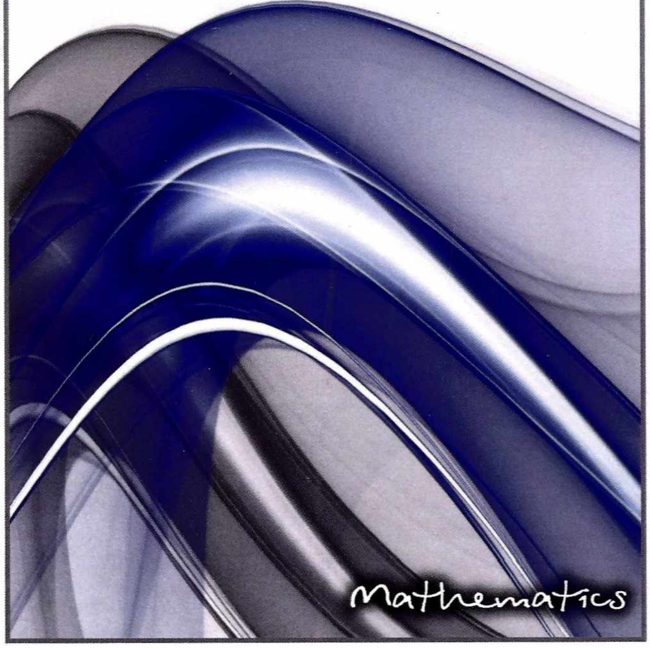
НАУЧНАЯ МЫСЛЬ



GRADIENT-BASED DYNAMICAL SYSTEMS FOR MATRIX COMPUTATIONS AND OPTIMIZATION

Министерство науки и высшего образования Российской Федерации
Сибирский федеральный университет

П. СТАНИМИРОВИЧ
И. ВЕЙ
Л. ЦЗИНЬ
А. СТУПИНА
Л. КАЗАКОВЦЕВ

СИСТЕМЫ ДЛЯ МАТРИЧНЫХ ВЫЧИСЛЕНИЙ И ОПТИМИЗАЦИИ НА ОСНОВЕ ГРАДИЕНТОВ

МОНОГРАФИЯ

Москва
ИНФРА-М
2024

УДК 519.6
ББК 22.192.31
С76

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (грант № 075-15-2022-1121)

Рецензенты:

Тынченко В.С., доктор технических наук, профессор кафедры технологий искусственного интеллекта Московского государственного технического университета имени Н.Э. Баумана (национального исследовательского университета);

Масич И.С., доктор технических наук, профессор кафедры системного анализа Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева

Станимирович П.

С76 Системы для матричных вычислений и оптимизации на основе градиентов : монография / П. Станимирович, И. Вей, Л. Цзинь, А. Ступина, Л. Казаковцев. — Москва : ИНФРА-М, 2024. — 314 с. : цв. ил.

ISBN 978-5-16-020482-6

Рекуррентные нейронные сети (RNN, recurrent neural networks) — один из двух обширных типов искусственных нейронных сетей, позволяющих выходным данным узлов влиять на последующий ввод в те же узлы. RNN непрерывного времени (CTRNN — continuous-time RNN) использует систему обыкновенных дифференциальных уравнений для определения воздействия на поступающие входные данные нейрона. Мы рассматриваем CTRNN, предназначенные для поиска корней уравнений или для минимизации нелинейных функций. Известны два важных класса CTRNN: градиентные нейронные сети (GNN, gradient neural networks) и нейронные сети Чжана (или обнуления — ZNN, Zhang neural networks).

GNN определяется как динамическая эволюция в направлении градиентного спуска нормы Фробениуса матрицы ошибок. Следовательно, существует строгая связь между проектированием динамических систем GNN и методами нелинейной оптимизации.

Цель монографии — обобщить новейшие разработки в теории и вычислительной практике линейной алгебры с использованием алгоритмов, основанных на динамическом системном подходе.

Основные темы — исследование динамических систем GNN, их проектирование и применение при вычислении обычных обратных и обобщенных обратных матриц, решении систем линейных матричных и векторных уравнений.

Ориентирована на аспирантов и исследователей в области математики и инженерии, специализирующихся на числовых методах линейной алгебры, оптимизации, динамических систем, систем управления, обработки сигналов.

УДК 519.6
ББК 22.19

ISBN 978-5-16-020482-6

Ministry of Science and Higher Education of the Russian Federation
Siberian Federal University

P. STANIMIROVIĆ
Y. WEI
L. JIN
A. STUPINA
L. KAZAKOVTSEV

GRADIENT-BASED DYNAMICAL SYSTEMS FOR MATRIX COMPUTATIONS AND OPTIMIZATION

MONOGRAPH

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2022-1121)

Reviewers:
Tynchenko V.S., Doctor of Technical Science, Professor of Bauman Moscow State Technical University;
Masich I.S., Doctor of Technical Science, Professor of Reshetnev Siberian State University of Science and Technology

Stanimirović P.
Gradient-based dynamical systems for matrix computations and optimization : monograph / P. Stanimirović, Y. Wei, L. Jin, A. Stupina, L. Kazakovtsev. — Moscow : INFRA-M, 2024. — 314 p. : col. ill.

ISBN 978-5-16-020482-6

Recurrent Neural Networks (RNN) is one of the two broad types of artificial neural network that allow the output from some nodes to affect subsequent input to the same nodes. A continuous-time recurrent neural network (CTRNN) uses a system of ordinary differential equations to define effects on incoming inputs on a neuron. We consider CTRNN dedicated to find zeros of equations or to minimize nonlinear functions. Two important classes of CTRNNs are known: Gradient Neural Networks (GNN) and Zhang (or Zeroing) Neural Networks (ZNN).

GNN is defined as the dynamical evolution along gradient-descent direction of the Frobenius norm of the error matrix. Therefore, there is a strict relationship between the design of GNN dynamic systems and nonlinear optimization methods.

The aim of this monograph is to collect the latest developments in the theory and computation in numerical linear algebra by means of algorithms based on dynamical system approach.

Main topics included in this book are investigation of GNN dynamical systems, their design, and applications in computation of the usual matrix inverse and generalized inverses, solving systems of linear matrix and vector equations.

This monograph is aimed to mathematics, engineering graduate students and researchers in the areas of numerical linear algebra, optimization, dynamical systems, control systems, signal processing.

ISBN 978-5-16-020482-6

Contents

Preface

Recurrent Neural Networks (RNN) is one of the two broad types of artificial neural network that allow the output from some nodes to affect subsequent input to the same nodes. A continuous-time recurrent neural network (CTRNN) uses a system of ordinary differential equations to define effects on incoming inputs on a neuron. We consider CTRNN dedicated to find zeros of equations or to minimize nonlinear functions. Two important classes of CTRNNs are known: Gradient Neural Networks (GNN) and Zhang (or Zeroing) Neural Networks (ZNN). GNN is defined as the dynamical evolution along gradient-descent direction of the Frobenius norm of the error matrix. Therefore, there is a strict relationship between the design of GNN dynamic systems and nonlinear optimization methods. The aim of this book is to collect the latest developments in the theory and computation in numerical linear algebra by means of algorithms based on dynamical system approach. Main topics included in this book are investigation of GNN dynamical systems, their design and applications in computation of the usual matrix inverse and generalized inverses, solving systems of linear matrix and vector equations. The dynamical system approach is a powerful tool for solving many kinds of matrix algebra problems because of:
(a) possibility to ensure a response within a predefined time-frame in real-time applications,
(b) its parallel distributed nature,
(c) convenience of hardware implementation,
(d) global convergence without conditions,
(e) in addition, dynamical system approach is applicable to online computation with time-varying matrices.

Our primary goal is the application of dynamical system approach in numerical linear algebra, especially in computation of generalized inverses, and solving system of linear matrix equations. Considered algorithms are aimed to both time-invariant and time-varying, complex and real matrices. Continuous-time computation has attracted a great scientific research. In particular, there is an increasing interest in continuous-time computation, where the states of the dynamical system evolve continuously. Many biological systems, some control systems can be better described in the analog manner. The main motivation is the goal to derive efficient models for the variety of continuous dynamical systems which appear in the real world phenomena.

All dynamical systems have discrete-time analogies, given in the form of appropriate difference equations. Main correlations between the continuous and discrete-time algorithms are considered in this book. Many computational problems (for example, structural analysis, electrical network analysis, weather forecasting) are based on matrix algorithms, such as solving linear systems, eigenvalue problem, singular value decomposition. These algorithms are often given in the discrete form of successive iterations $x(k+1) = F(x(k))$, where F is appropriate function. Such iterations are *discrete-time* (DT systems). On the other hand, the iterations $x(k+1) = F(x(k))$ can be considered as dynamical systems where the *state variable* x depends on the time k that takes discrete integer values. A sequence of points $\{x(k)\}_{k=-\infty}^{\infty}$ satisfying $x(k+1) = F(x(k))$ is called the orbit of F at $x(0)$. Nevertheless, most physical systems are dynamical continuous-time (CT) systems. CT systems are described in terms of ordinary differential equations (ODE) or partial differential equations (PDE) in the form $\dot{x}(t) = g(x(t))$, where $\dot{x}$ means the time derivative of $x(t)$ and g is appropriate time-varying function. In general, continuous-time algorithms and analog VLSI systems could be analyzed by means of time discretization that could destroy the main characteristics of these systems. There are important differences between CT and DT systems. The problem of the existence and uniqueness is an important issue for CT systems, but not for DT systems. Further, CT systems are easier for analysis that DT systems, because the orbits of many CT systems are continuous curves in the state space, while the orbits of DT systems are sequences of points, difficult for analysis. But, in spite of great difference, there is a great influence between the continuous-time and discrete-time computation. Results obtained in the research of dynamical systems can help in developing new iterations; and opposite, knowledge of derived iterations can help in better understanding and even in defining new dynamical evolutions.

Generalized inverses are included an extensive variety of mathematical fields, for example, matrix theory and operator theory. Many types of generalized inverses have various applications such as linear estimation, differential and difference equations, Markov chains, graphics, cryptography, coding theory, robotics, incomplete data recovery, sociology, demography and many other fields. A special case of the Drazin inverse, called Group inverse, has found application in characterizing the sensitivity of the stationary probabilities to perturbations in the underlying transition probabilities. Finally, the group inverse has recently proven to be fundamental in the analysis of Google's PageRank search engine. Generalized inverses play an important role in finding solutions of many stochastic models, in particular Markov chains in discrete or continuous time and Markov renewal processes [92]. The Drazin inverse has been successfully and extensively applied in different fields of science; for example, in finding closed form solutions of singular differential equations with matrix coefficients, in solving difference equations, in Markov chains, multibody system dynamics as well as in finding solutions of various iterative methods. The Moore-Penrose inverse has found a wide range of applications in many areas of science and became a useful in finding least squares solutions of linear systems, in optimization problems, in data analysis, in finding the solution of linear integral equations, etc. Global overview of various applications of generalized inverses can be found in [11]. The intrinsic estimator (IE) has become a widely used tool for the analysis of age-period-cohort (APC) data in sociology, demography, and other fields. It was recently observed in [60] that the IE is a subtype of a larger class of estimators based on the Moore-Penrose generalized inverse (MP estimators). Also, different estimators can lead to radically divergent estimates of the true, unknown APC effects [60]. For more information of he history of generalized inverses the reader is referred to two survey papers [12, 11].

On the other hand, time-varying problems appear more and more frequently in scientific and engineering applications, such as circuit parameters in electronic circuits, aerodynamic coefficients in high-speed aircraft, and mechanical parameters in machinery [318, 86]. Our intention was the attempt to overcome differences between the disciplines of engineering and mathematics, in such a way that the book is interesting for engineers and adequately rigorous for mathematicians.

More precisely, our main interest is computation of main problems in numerical linear algebra by means of dynamical system approach in both time-varying and time-invariant case. Algorithms for their computation are developed and the performance comparison of such algorithms is given. Several applications are presented having in mind that gen- eralized inverses are very powerful tools and are applicable in many branches of mathe- matics, technics and engineering. The most frequent and important is the application in finding solution of many matrix equations and system of linear equations. Besides nu- merical linear algebra, there are a lot of other mathematical and technical disciplines in which generalized inverses play an important role. Some of them are: estimation theory (regression), computing polar decomposition, electrical circuits (networks) theory, automatic control theory, filtering, difference equations, pattern recognition, image restoration.

So far, many classes of generalized inverses have been proposed and investigated. The most popular are the Moore-Penrose inverse and the Drazin inverse.

Time-varying mathematical problems frequently appear in both scientific research and practical applications. For example, main appearances of time-varying problems are circuit parameters in electronic circuits, aerodynamic coefficients in high-speed aircraft, mechanical parameters in machinery, robot motion planning, chaotic noise rejection for sensors, numerical online problems, time-varying nonlinear optimization problems, etc. In 2001 Zhang et al. developed a special class of recurrent neural networks (RNNs), namely Zhang or zeroing neural networks (ZNNs) for solving efficiently time-varying problems. Zeroing neural networks (ZNN) are able to perfectly track time-varying solutions by exploiting the time derivative of time-varying parameters. As a consequence, many researchers make progresses along this direction by proposing various kinds of ZNN models for solving problems with different features. ZNN dynamical systems are completely new, reliable and highly accurate for solving miscellaneous time-varying matrix, vector or scalar problems. It differs from all other time-varying matrix methods in its use of an error equation and initiated model based on ordinary differential equation (ODE) that assures exponentially fast convergence. The ODE used in the model can be transformed into discrete-time iterations by means of different finite difference equations. This gives a new look to classical iterative methods as well as mutual interaction between continuous-time and discrete-time algorithms.

This book was primarily aimed to researchers in matrix theory, numerical linear algebra, particularly in the theory of generalized inverses and its applications. Researchers in recurrent neural networks and artificial intelligence can find useful material in this book. Also, the presented material should be of interest for readers interested in numerical analysis and symbolic computation. In addition, the book can be very useful for researchers interested in nonlinear programming. In general, dynamical systems models are defined as continuous-time analogies of known nonlinear optimization algorithms, such as the class of gradient-descent algorithms or various modifications of the Newton method for solving nonlinear optimization problems. Both the time-varying and time-invariant environments in defining recurrent neural networks are considered. This book is aimed to mathematics and engineering graduate students and researchers in the areas of numerical linear algebra, optimization, dynamical systems, control systems, image processing. It can also be used as a text or reference for many graduate courses or as a reference for many courses in postgraduate levels in computer science, mathematics or in technical faculties. The reader should be familiar with basic linear algebra, matrix theory, ordinary differential equations, mathematical and functional analysis. Basic knowledge of the algorithm theory, matrix theory and dynamical systems is recommended. Knowledge in *Matlab* programming and in *Simulink* modeling, which is a graphical extension to *Matlab* for modeling and simulation of systems, is desirable. We believe that the book should be of use for many researchers, students in applied mathematics, statistics, engineering, and many other scientific disciplines.

According to MSC classification, the topics considered in this book are classified as follows:

- 15A09 Theory of matrix inversion and generalized inverses
- 15A10 Applications of generalized inverses
- 15A24 Matrix equations and identities
- 65F45 Numerical methods for matrix equations
- 68T07 Artificial neural networks
- 90C30 Nonlinear programming.

This book contains 10 chapters. In what follows, we present a short summary focusing on the key concepts of each chapter.

Chapter 1 is a short introduction into basic topics from the matrix theory, optimization theory and neural networks. Section 1.1 contains fundamental notions like Jordan decomposition, singular value decomposition, idempotent matrices and projectors, the trace function, Kronecker product and vectorization of matrices. Section 1.2 provides definitions and properties of generalized inverses together with the most important applications of generalized inverses to linear systems solving. General information about unconstrained optimization is included in section 1.3. In section 1.4 we refer to stability and convergence properties of these networks according to Lyapunov theory. Section 1.5 is dedicated to a brief introduction to gradient neural networks (GNNs) and zeroing neural networks (ZNNs). Then a survey of some additional ZNN models is presented. Chapter 1 concludes with section 1.7, which describes importance of generalized inverses and dynamical systems.

Chapter 2 presents a robust analysis of GNNs. This chapter describes gradient-based recurrent neural networks used for solving linear matrix equations and to compute different generalized inverses of constant real or complex matrices. The focus of this chapter is in solving matrix equations and in the computation of the matrix inverse and various kinds of generalized inverses. Particularly, the first section 2.1 is a short review of the gradient neural networks, especially those that are focused on the computation of the matrix inverse and the Moore-Penrose generalized inverses and the second section 2.2 is devoted to apply GNN dynamics for solving matrix equation $AXB = D$ and its various particular cases.

The third chapter studies recurrent neural networks arising from gradient neural networks. Obtained dynamical systems do not follow true GNN design. The goal of this section is to investigate modified dynamical systems and their applications in computing the Moore-Penrose and the Drazin inverse. Section 3.1 studies a specific GNN model for computing outer inverses with prescribed range and null space. Subsequent section present GNN models for approximating the Moore-Penrose, the Drazin inverse and various expressions involving particular outer inverses. Convergence properties of described GNN models are studied in details.

Chapter 4 investigates dynamical systems based on full rank factorization $A = PQ$ of the matrix A. This approach assumes usage of the input matrix $A \in \mathbb{C}_r^{m\times n}$ in conjunction with the matrix $G \in \mathbb{C}_s^{n\times m}$, $0 < s \leq r$. Two dynamic state equations and corresponding gradient based RNNs for generating the class of outer inverses are proposed in [226].

The influence of non-linear activations of GNNs for the computation of the Drazin inverse and W-weighted Drazin inverse are given in Chapter 5.

Conditions for existence, representations and computation of matrix generalized inverses are presented in Chapter 6 together with several algorithmic procedures tested on a number of versatile numerical simulations. This chapter presents an interesting combination of an algebraic approach and dynamical system approach in computation of various classes of generalized inverses with prescribed range and/or null space. Namely, the algebraic approach gives some useful representations of generalized inverses, while the dynamical systems are defined using obtained representations in order to solve required matrix equations.

Improvements of the GNN dynamics based on the utilization of gradient descent optimization methods is investigated in Chapter 7. Main goal is to solve the equation $AXB = D$ and apply its particular cases in computing generalized inverses in real time improving the GNN standard model GNN(A, B, D). Our motivation is to improve the GNN(A, B, D) and develop the novel gradient-based GNN (GGNN) design, termed as GGNN(A, B, D) utilizing a novel type of dynamical system. The proposed GGNN model is defined evolving the standard GNN dynamics along the gradient of the standard error matrix.

Chapter 8 investigates two continuous-time neural networks for computing generalized inverses of complex-valued matrices based on constrained quadratic optimization. These neural networks are aimed to the Moore-Penrose inverse and outer inverses in continuous time. These neural networks are generated using the fact that outer inverses and the Moore-Penrose inverse can be derived as the solution of appropriate, matrix valued, convex quadratic programming problems. The first of them is applicable in the pseudoinverse computation and the second one is applicable in construction of outer inverses.

Symbolic computation of generalized inverses based on exact-free and symbolic solving of dynamic state equations is presented in Chapter 9. The proposed algorithms are based on the exact solution of first order systems of differential equations which appear in the dynamic state equation that define corresponding outer inverse. The algorithm is applicable to matrices whose entries are integers, rational numbers as well as rational or polynomial expressions.

The last Chapter 10 investigates improved dynamical systems based on modified of modified and time-varying gain parameter. Hybridizations of GNN and ZNN design are presented in Section 10.2. For that purpose, Section 10.1 describe various ZNN models aimed to solve the scalar, vector and matrix inversion problems. Section 10.3. considers GNN and ZNN with variable gain parameter.

The authors are encouraging the readers for their comments, suggestions and corrections.

Acknowledgments

Yimin Wei was supported by the National Natural Science Foundation of China under grant 12271088 and the Joint Research Project between China and Serbia under the grant 2024-6-7, School of Mathematical Sciences and Key Laboratory of Mathematics for Nonlinear Sciences, Fudan University.

Predrag Stanimirović gratefully acknowledges support from the Research Project, grant no. 451-03-65/2024-03/200124 of the Serbian Ministry of Science and Education. Predrag Stanimirović is supported by the Science Fund of the Republic of Serbia, (No. 7750185, Quantitative Automata Models: Fundamental Problems and Applications - QUAM).

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2022-1121).

Some of the work has been written in conjunction with the writing of research papers by the authors or in collaboration with many colleagues and PhD students. During preparation of this book and, in general, during the many years long research in the area of generalized inverses, the authors have had frequent conversations and consultations with numerous colleagues. We would like to thank especially to the following colleagues.

Marko Petković, University of Niš, Faculty of Sciences and Mathematics, Serbia, for significant help in defining new GNN and ZNN models. Also, his contribution in discretization of continuous-time models as well as in *Simulink* implementation of defined dynamical systems is valued.

Vasilios Katsikis, for help in developing many ZNN dynamical system described in the book.

Miroslav Ćirić, University of Niš, Faculty of Sciences and Mathematics, Serbia, helped in algebraic approach in computation of generalized inverses.

Ivan Živković, University of Niš, Faculty of Sciences and Mathematics, Serbia. Research with him motivated research on simplified dynamical systems.

Shuai Li, Faculty of Information Technology, University of Oulu, Finland, for support in Zhang neural design covering various generalized inverses and resistant with respect to various kinds of noise.

Xue-Zhong Wang, Hexi University, Zhangye, China, for immeasurable contribution in both gradient based and zeroing based dynamical systems.

Haifeng Ma, Harbin Normal University, China, for contribution in development and investigation of both gradient and zeroing dynamical systems.