北京雁栖湖
应用数学研究院
Beijing Institute
of Mathematical
Sciences and Applications

Research Progress Report

Stephen S.-T. Yau （丘成栋）
Affiliation: BIMSA

Pure Mathematics Group Members

Stephen Shing-Toung Yau
Professor

BIMSA Group

Chuangqiang Hu
Assistant Professor

Beihui Yuan
Assistant Professor

Zhiwen Liu
Postdoc

Genglong Lin
Postdoc

Tsinghua Group

Huaiqing Zuo
Professor

Shuanghe Fan
Postdoc

Publication List in 2024

Published or accepted Article

[1] M. DiPasquale and **B. Yuan**, *Bivariate Splines on a Triangulation with a Single Totally Interior Edge*, SIAM Journal on Applied Algebra and Geometry 8, no. 3 (2024): 686-712

[2] M. Sirvent, T. Sorokina, N. Villamizar and **B. Yuan**, *Multivariate polynomial splines on generalized oranges*, Journal of Approximation Theory 299 (2024): 106016

[3] K. Ranestad, M. Kapustka, G. Kapustka, H. Schenck, M. Stillman and **B. Yuan**, *Quaternary quartic forms and Gorenstein rings*, accepted by Memoirs of The American Mathematical Society (2024)

[4]**Chuangqiang Hu**, and Xiao-Min Huang, *Drinfeld Module and Weil pairing over Dedekind domain of class number two*, Finite Field and Their Applications (2024)

Publication List in 2024

[5] Bingyi Chen, Naveed Hussain, **Stephen S.-T. Yau** and Huaiqing Zuo, *On variation of complex structures and variation of new Lie algebras arising from singularities*, Proceedings of the 8th International Consortium of Chinese Mathematicians, 2019, Beijing, China, Vol. 1 (2024), 731-743.

[6] Zida Xiao, **Stephen S.-T. Yau**, Qiwei Zhu and Huaiqing Zuo, *On the Nakai Conjecture for some singularities*, to appear, Annali della Scuola Normale Superiore di Pisa, Classe di Scienze, 2024.

[7] Quan Shi, **Stephen S.-T. Yau**, and Huaiqing Zuo, *On T-maps and ideals of antiderivatives of hypersurface singularities*, Izvestiya: Mathematics, Vol. 88, No. 6 (2024), 190-226.

[8] Guorui Ma, Yang Wang, **Stephen S.-T. Yau**, and Huaiqing Zuo, *Hodge moduli algebras and complete invariants of singularities*, Asian Journal of Mathematics, Vol. 28, No. 1 (2024), 001-046.

[9] Guorui Ma, **Stephen S.-T. Yau**, Qiwei Zhu and Huaiqing Zuo, *A characterization and solvability of quasihomogeneous singularities*, Pacific Journal of Mathematics, Vol. 329, No. 1 (2024), 121-146.

Completed Papers in 2024

[1] S. Fan, N. Hussain, **Stephen S.-T. Yau** and H. Zuo, *New invariants of singularities in terms of higher Nash blow-up local algebras*, preprint.

[2] S. Fan, **Stephen S.-T. Yau** and H. Zuo, *Higher order Jacobian matrix theory and invariants of singularities*, preprint.

[3] S. Fan, **Stephen S.-T. Yau** and H. Zuo, *Higher order Hessian matrix theory and its applications in singularity theory and Calabi-Yau manifolds*, preprint.

[4] S. Fan, **Stephen S.-T. Yau** and H. Zuo, *Finite determinacy of formal power series*, preprint.

[5] S. Fan, **Stephen S.-T. Yau** and H. Zuo, *Higher order Jacobian matrix theory and the inverse of an automorphism of a power series ring*, preprint.

[6] **Genglong Lin**, *Finite energy range of Hessian operator*, preprint.

[7] Yinji Li, **Genglong Lin** and Xiangyu Zhou, *Monge Ampere type equations on compact hermitian manifold*, preprint.

Completed Papers in 2024

[8] **Chuangqiang Hu**, **Stephen S.-T. Yau**, Huaiqing Zuo, *Groupoids derived from the simple elliptic singularitiesm*, https://arxiv.org/abs/2410.10124.

[9] **Chuangqiang Hu**, **Stephen S.-T. Yau**, Huaiqing Zuo, *On the k-th Tjurina number of weighted homogeneous singularities*, https://arxiv.org/abs/2409.09384

[10] Zhuo Chen, **Chuangqiang Hu**, Tao Zhang, Xiaopeng Zheng, *Drinfeld Modular Curves Subordinate to Conjugacy Classes of Nilpotent Upper-Triangular Matrices*, https://arxiv.org/abs/2309.00432.

[11] **Zhiwen Liu** and **Stephen S.-T. Yau**, *Variation of Complex Structures and Variation of Lie Algebras III: Yau Algebras Arising from Singularities*, submitted to 'Izvestiya: Mathematics'.

[12] J. Hu, T. Lin, Q. Wu and **B. Yuan**, *The condition for constructing a finite element from a superspline*, arXiv:2407.03680, submitted to Mathematics of Computation, 2024.

[13] A. Mantzaflaris, B. Mourrain, N. Villamizar and **B. Yuan**, *An algebraic framework for geometrically continuous splines*, arXiv:2305.09096, submitted to Mathematics of Computation, 2023, revised in 2024.

Funding Program

Approved:

Chuangqiang Hu and collaborators

Title：Mathematical Theory of isogeny-based Cryptography（椭圆曲线同源密码的数学理论）

Identity: Main Participant

Project Funding Intensity: 1,640,000 Yuan

No. 12441107，From 2025.01 to 2028.12

Research Plan for 2025

Beihui Yuan：

Project 1: Sheaf theory for Finite Element Method (with Ting Lin)

Project 2: Derived symmetries, Fano covers, and normal bundles (with Will Donovan)

Project 3: On Waring rank of hypersurface singularities.

Applying NSFC (青年基金)

Sheaf theory for splines and finite elements

Research Plan for 2025

Chuangqiang Hu：

Project1： Drinfeld Module and Weil pairing over Dedekind domain of class number two, Chuangqiang Hu and Xiuwu Zhu

We derive the new families of Drinfeld modules over non-rational fields.

Project2：Anderson generating function of rank one Drinfeld Module over rational function fields
Hu, Chuangqiang, and Xiao-Min Huang, AND Stephen S.-T. Yau

We construct the Anderson generating functions of the Drinfeld modules and compute the Carlitz period of the associated exponential function.

Applying NSFC (面上基金)

On the caculation of geometric L-functions over function fields

Research Plan for 2025

Zhiwen Liu：

Project 1: Considering several special families of K3 surface singularities and to explore the invariants of their Yau algebras(with Stephen S.-T. Yau)

Project 2: Complex structures and cohomology information on the Yau algebras (with Stephen S.-T. Yau and postdoc-Alejandro) we are exploring which kind of nilpotent Lie algebras can be obtained as Yau algebras, using some tools form complex differential geometry and cohomology theory.

Research Plan for 2025

Genglong Lin：

Project1: Considering fully nonlinear elliptic equations for possibly non-integrable almost complex structure and investigate some almost hermitian geometry

Project 2: consider degenerate Monge Ampere type equation on some HKT manifold by some technique of Lu, Sroka, Dinew

Course Information

Spring Semester:

Computational commutative algebra (Beihui Yuan);
Simplicial homotopy theory and model category (Chuangqiang Hu)

Fall Semester:

Infinite free resolutions (Beihui Yuan);
Introduction to Drinfeld module (Chuangqiang Hu)

Significant Research in 2024

Higher Order Jacobian Matrix Theory,
Higher Order Hessian Matrix Theory,
and Their Applications in Singularity Theory

(joint work with Shuanghe Fan and Huaiqing Zuo)

Overview

In this research, we develop the innovative concepts of "higher order Jacobian matrix theory" and "higher order Hessian matrix theory". In addition, we study their applications in algebraic geometry and singularity theory.

On Higher Order Jacobian Matrix Theory

The higher Nash blow-up is a technique on algebraic varieties that replaces singular points with the limit space of higher order data associated with non-singular points of the variety.

Inspired by this methodology, we introduce the “higher order Jacobian matrix theory”, which generalizes traditional Jacobian matrix theory.

This research explores three significant applications of this theory.

On Higher Order Jacobian Matrix Theory

Firstly, we utilize it to construct new invariants of isolated singularities that generalize the well-known Tjurina and Milnor algebras in algebraic geometry and singularity theory.

These algebras preserve the contact and right invariants respectively.

On Higher Order Jacobian Matrix Theory

Secondly, this theory allows us to compute the inverse of the automorphism of formal power series which can be treated as non-linear extension of the inverse matrix method and the Gauss elimination method respectively.

This approach facilitates the derivation of explicit forms of power series for implicit variables in accordance with the implicit function theorem, enabling the solution of systems of nonlinear equations.

On Higher Order Jacobian Matrix Theory

Finally, this theory provides a novel perspective on the finite determinacy of formal power series over an arbitrary field, while the classical finite determinacy theorem only works for convergent power series over complex number field.

Within the context of formal power series, we introduce a more refined criterion, along with computational techniques for explicit formula of automorphisms and invertible elements. Together with Artin's Approximation Theorem, the proposed method enables explicit calculations related to the finite determinacy theorem.

On Higher Order Hessian Matrix Theory

Projective equivalence is another fundamental problem in algebraic geometry and singularity theory.

For projective manifolds defined by equations of degree 2, this can be solved using quadratic form theory and classical Hessian matrix theory.

For degrees greater than 2, inspired by the "higher order Jacobian matrix theory", we develop a new "higher order Hessian matrix theory", which generalizes the classical Hessian matrix theory.

On Higher Order Hessian Matrix Theory

As a consequence, we introduce invariants of projective manifolds under projective transformations and provide numerous examples. The application of these invariants extends beyond the scope of projective transformations.

Notably, one sequence of our invariants bears a striking resemblance to the well-known j-invariant expression.

We use it to solve the longstanding problem of the complete classification of complex structures of K3 surfaces in CP^3.

Summary

The “higher order Jacobian matrix theory” and “higher order Hessian matrix theory” established in this research are foundational and innovative, with promising applications across various branches of mathematics.

Control Group Members

Stephen Shing-Toung Yau
Professor

BIMSA Group

Guoqing Hu
Professor

Yishuai Niu
Associate Professor

Jiayi Kang
Assistant Professor

Xiaopei Jiao
Assistant Professor

Collaborators

Xiuqiong chen
Renmin University
Assistant Professor

Wenhui Dong
Guangxi University
Assistant Professor

PHD Students

Zeju Sun
Tsinghua University

Minli Feng
BIMSA

Publication List in 2024

Published or accepted Article

[1] Z. Sun, and **Stephen S.-T. Yau**, *The Applications of Yau-Yau Algorithm on McKean-Vlasov Filtering Problem*, Journal of Systems Science & Complexity, (2024).

[2] X. Chen, **J. Kang**, and **Stephen S.-T. Yau**, *Continuous discrete optimal transportation particle filter*, accepted by Asian Journal of Mathematics, (2024).

[3] J. Shi, **X. Jiao**, and **Stephen S.-T. Yau**, *DGLG: A Novel Deep Generalized Legendre-Galerkin Approach To Optimal Filtering Problem,* accepted by IEEE Transactions on Automatic Control, (2024)

[4] Y. Tao, **J. Kang**, and **Stephen S.-T. Yau**, *The Stochastic Stability Analysis for Outlier Robustness of Kalman-type Filtering Framework Based on Correntropy-Induced Cost*, accepted by IEEE Transactions on Automatic Control, (2024).

[5] Z. Sun, X. Chen, and **Stephen S.-T. Yau**, *Recurrent Neural Network Spectral Method and its Application in Stable Filtering Problems*, accepted by Automatica, (2024).

[6] **X. Jiao**, and **Stephen S.-T. Yau**, *Weak form Mitter conjecture on nonmaximal rank estimation algebra: state dimension 4 and rank 3*, accepted by Journal of Systems Science & Complexity, (2024)

Publication List in 2024

Published or accepted Article

[7] J. Shi, **X. Jiao**, and **Stephen S.-T. Yau**, *A Novel Logarithmic Transformed Deep Galerkin Approach to Optimal Filtering Problem,* Proceeding of 2024 IEEE 63rd Conference on Decision and Control (CDC), (2024)

[8] Z. Sun, X. Chen, Y. Tao, and **Stephen S.-T. Yau**. *A Uniform Framework of Yau-Yau Algorithm Based on Deep Learning with the Capability of Overcoming the Curse of Dimensionality*, accepted by IEEE Transactions on Automatic Control, (2024).

[9] **X. Jiao**, and **Stephen S.-T. Yau**, *Finite-dimensional estimation algebra on arbitrary state dimension with nonmaximal rank: linear structure of Wong matrix*, International Journal of Control, 97(11), 2669-2676, (2024).

[10] Y. Tao, **J. Kang**, and **Stephen S.-T. Yau**, *Neural Projection Filter: Learning Unknown Dynamics Driven By Noisy Observations*, IEEE Transactions on Neural Networks and Learning Systems, 35(7), 9508-9522, (2024).

[11] X. Chen, **J. Kang**, and **Stephen S.-T. Yau**, *Time-varying Feedback Particle Filter*, Automatica, 167, 111740, 1-8, (2024).

Completed Papers in 2024

[1] **J. Kang**, A. Salmon and **Stephen S.-T. Yau**, *Nonexistence of finite-dimensional estimation algebras on closed smooth manifolds*, submitted to IEEE Transactions on Automatic Control (2024).

[2] **J. Kang**, **X. Jiao**, and **Stephen S.-T. Yau**, *Estimation of the Linear System via Optimal Transportation and Its Application for Missing Data Observations*, submitted to IEEE Transactions on Automatic Control (2024).

[3] **J. Kang**, X. Chen, and **Stephen S.-T. Yau**, *Explicit Convergence analyses of PDE-based filtering algorithms*, submitted to SIAM Journal on Control and Optimaization (2024).

[4] Minli Feng, Xiuqiong Chen, **Stephen S.-T. Yau**, *The Extended Direct Method for Generalized Time-varying Yau Filtering Systems,* submitted to Automatica, (2024).

[5] W. Dong, and **Stephen S.-T. Yau**, *Splitting-up Method for Zakai Equation with Multiplicative α-Stable L\'evy Noise: Convergence Analysis and Applications to Nonlinear Filtering Problems*, submitted to IEEE Transactions on Automatic Control (2024).

[6] W. Cao, S. Liu, C. Liu, Z. He, **Stephen S.-T. Yau**, and S. Li, *Convolutional Bayesian Filtering*, submitted to IEEE Transactions on Automatic Control (2024).

[7] W. Cao, T. Zhang, Z. Sun, C. Liu, **Stephen S.-T. Yau**, and S. Li, *Nonlinear Bayesian Filtering with Natural Gradient Gaussian Approximation,* submitted to IEEE Transaction on Pattern Analysis and Machine Intelligence, 2024.

Funding Program

Approved:

Zeju Sun

Title: On the Algorithms and Properties of McKean-Vlasov Equation with its Nonlinear Filtering Problem (McKean-Vlasov方程及其滤波问题的解法和性质研究)

Identity: Host

Project Funding Intensity: 300,000 Yuan

No. 123B2020, From 2024.05 to 2025.12

Research Plan for 2025

Jiayi Kang:

1. Extend the theoretical framework of the Yau-Yau filtering algorithm, including concepts based on manifolds, generalized noise, and related areas.

2. Explore practical applications of existing algorithms in a wide range of real-world scenarios, such as finance, engineering, and biology. Collaborations are currently underway with several engineering institutes of the Chinese Academy of Sciences.

3.Investigate the intersection of filtering and optimization theories.

The goal is to collaboratively complete 3-4 research papers in the aforementioned areas. In 2025, Jiayi Kang plans to apply as a participant for the 2025 General Program of the National Natural Science Foundation of China (NSFC).

Research Plan for 2025

Xiaopei Jiao:

1. Based on Yau-Yau filer, “direct method” has been proposed to deal with finite dimensional filter with quadratic potential in the recent years. One meaningful attempt is to extend “direct method” to handle a class of infinite dimensional filters such as cubic sensors. Fast solver of time-varying Schrodinger equation will be designed for example based on fundamental solution.

2. With the help of development of scientific machine learning, it is promising to accelerate offline part calculation of Yau-Yau filter. One appealing direction is to combine Physics-Informed Neural Networks (PINNs) to incorporate the information of forward Kolmogorov equation. New fast algorithm will be exploited further compared with traditional filter like extended Kalman filter and particle filter.

In the next coming year, I plan to publish 3-4 papers based on what we have finished. At the same time, I also shall attempt to apply the grant provided by national science foundation and other available sources.

Course Information

Spring Semester:

Deep Learning (Guoqing Hu);
Optimization algorithms in machine learning (Yishuai Niu)

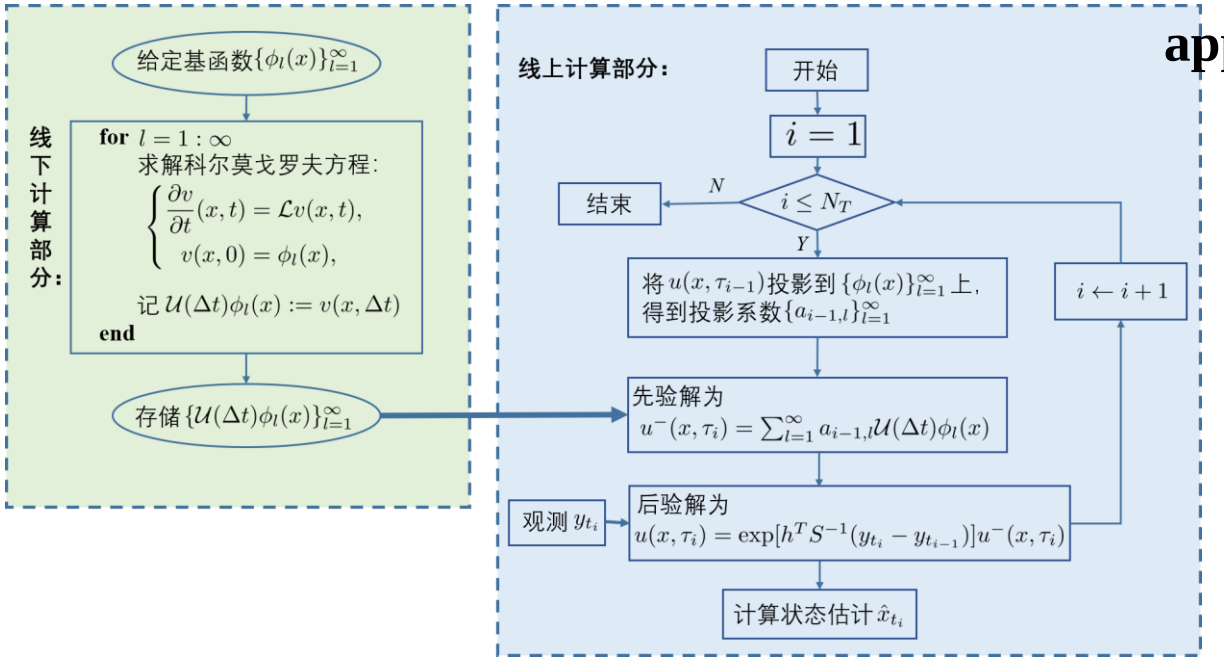
Fall Semester

Deep Learning Theory plus Practical Project Course (Guoqing Hu);
Introduction to Drinfeld module (Yishuai Niu);
The mathematics for control and filtering (Jiayi Kang)

Major Discoveries

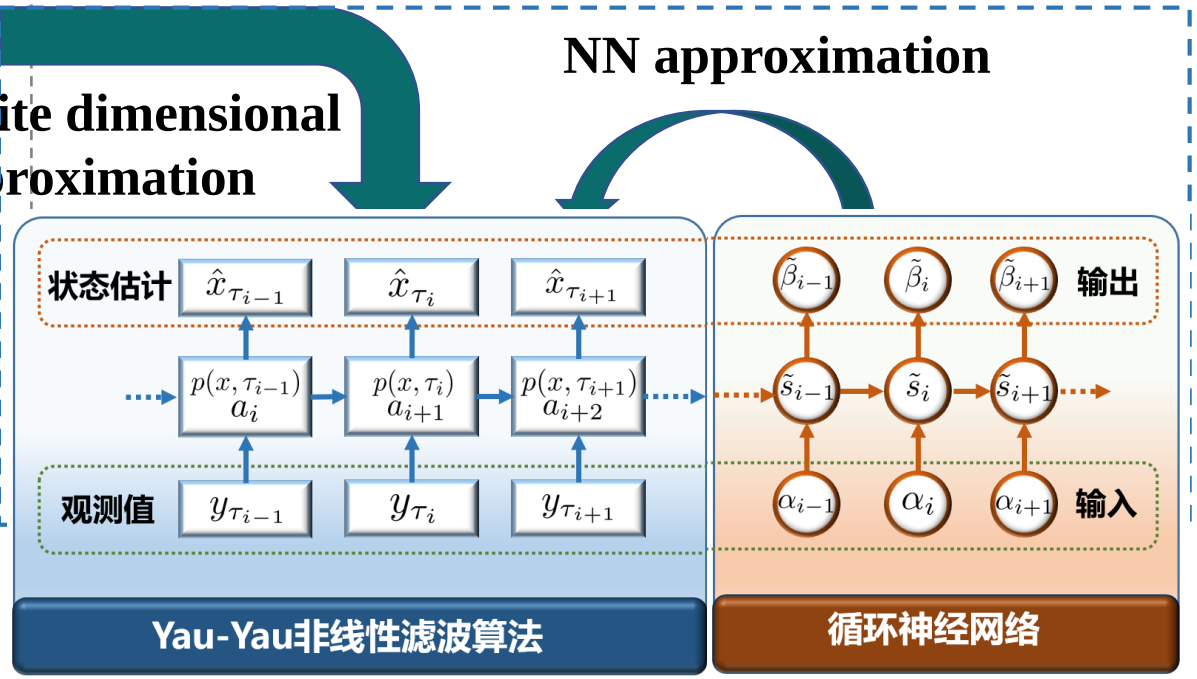
◈ The Yau-Yau nonlinear filtering algorithm addresses core challenges in nonlinear filtering by transforming them into numerical solutions of partial differential equations (PDEs). While traditional methods face the curse of dimensionality, we recently developed a novel solution framework using recurrent neural networks (RNNs).

Yau-Yau algorithm framework



Finite dimensional approximation

NN approximation



Connections

Major Discoveries

Connection between Yau-Yau algorithm and RNN:

The computational graph of the Yau-Yau nonlinear filtering algorithm aligns closely with the structure of RNN.

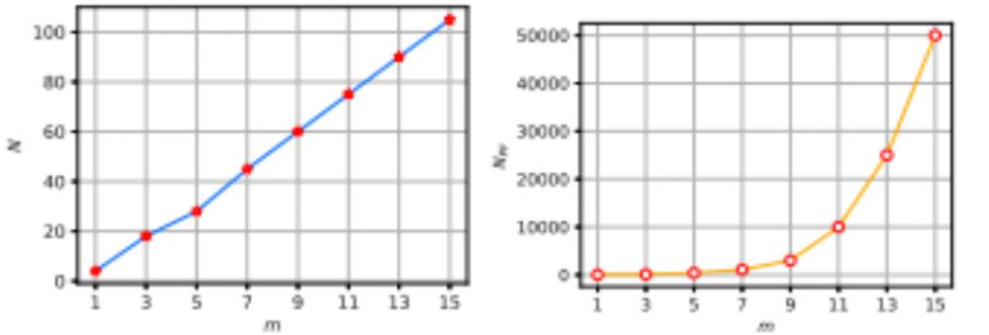
The filtering algorithm based on RNNs achieves high numerical accuracy and effectively solves filtering equations.

Advantages of using RNNs:

The algorithm is highly efficient and accurate, with polynomial parameter growth in high-dimensional filtering problems, effectively overcoming the curse of dimensionality.

A Uniform Framework of Yau-Yau Algorithm Based on Deep Learning with the Capability of Overcoming the Curse of Dimensionality

Xiuqiong Chen, *Member, IEEE*, Zeju Sun, Yangtianze Tao and Stephen S.-T. Yau, *Life Fellow, IEEE*



Bioinformatics Group Members

Stephen Shing-Toung Yau
Professor

BIMSA Group

Guoqing Hu
Professor

Yishuai Niu
Associate Professor

Qi Wu
Associate Professor

Xin Zhao
Associate Professor

Nan Sun
Postdoc

PHD Students

Hongyu Yu

Mengcen Guan

Tao Zhou

Hao wang

Piyu Zhou

Publication List in 2024

Published or accepted Article

[1] **Y.S. Niu**, Hoai An Le Thi, and Dinh Tao Pham, *On Difference-of-SOS and Difference-of-Convex-SOS Decompositions for Polynomials*, SIAM Journal on Optimization, 34(2024), 2, 1852-1878.

[2] H. Zhang, **Y.S. Niu**, *A Boosted-DCA with Power-Sum-DC Decomposition for Linearly Constrained Polynomial Program*, Journal of Optimization Theory and Applications, 201(2024), 720-759.

[3] Hongyu Yu and **Stephen S.-T. Yau**, *On Difference-of-SOS and Difference-of-Convex-SOS Decompositions for PolynoAutomated recognition of chromosome fusion using an alignment-free natural vector method*, Frontiers in Genetics, section Computational Genomics, Vol. 15 (2024), 1364951, 1-10.

[4] Ruohan Ren*, Hongyu Yu*, Jiahao Teng, Sihui Mao, Zixuan Bian and Yangtianze Tao, and **Stephen S.-T. Yau**, *CAPE: a deep learning framework with Chaos-Attention net for Promoter Evolution*, Briengs in Bioinformatics, Vol. 25, No. 5 (2024), 1-12.

Publication List in 2024

Published or accepted Article

[5] Xiaoguang Li*, Tao Zhou*, Xingdong Feng, **Shing-Tung Yau**, and **Stephen S.-T. Yau**, *Exploring geometry of genome space via Grassmann manifolds*, The Innovation, Vol. 5, No. 5 (2024), 100677, 1-8.

[6] Hongyu Yu and **Stephen S.-T. Yau**, *New Virus Variant Detection Based on the Optimal Natural Metric*, Genes, Vol. 15:891 (2024), 1-12.

[7] Harris Song*, **Nan Sun***, Wenping Yu, and **Stephen S.-T. Yau**, *A Novel Natural Graph for Efficient Clustering of Virus Genome Sequences*, Current Bioinformatics, Vol. 19 (2024), 687-703.

[8] Hongyu Yu and **Stephen S.-T. Yau**, *The optimal metric for viral genome space*, Computational and Structural Biotechnology Journal, Vol. 23 (2024), 2083-2096

[9] Mengcen Guan, **Nan Sun** and **Stephen S.-T. Yau**, *Geometric Analysis of SARS-CoV-2 Variants*, Gene, Vol. 909 (2024) 148291, 1-11.

Completed Papers in 2024

[1] **Y.S. Niu**, H.A. Le Thi, D.T. Pham, *An Accelerated DC Programming Approach With Exact Line Search for the Symmetric Eigenvalue Complementarity Problem*.

[2] **Y.S. Niu**, Y. You, M.F. Benammour, Y.J. Wang, *Parallel DCCUT Algorithms for Mixed-Binary Linear Programs.*

[3] **Guoqing Hu**, Tao Zhou, Piyu Zhou, and **Stephen S.-T. Yau**, *K-mer condition brings asymmetry to the covariance: an efficient improvement of the natural vector method.*

[4] **Qi Wu**, *A mechanic program for phenotypic evolution.*

[5] **Xin Zhao**, *IGHV3-53 Sequence Characteristics and Neutralizing Activity: Predictive Modeling and Broad-Spectrum Antibody Design Against SARS-CoV-2.*

Funding Program

Under research:

Xin Zhao

Title: Genomic Big Data Analysis and Application Research Based on Natural Vector and Convex Analysis Methods（基于自然向量与凸分析方法的基因组大数据分析及应用研究）

Identity: Host

Project Funding Intensity: 300,000 Yuan

No. 12201015，From 2023.01 to 2025.12

Research Plan for 2025

Guoqing Hu：

Project 1: Exploring potential transcription factors and their regulatory mechanisms in humans and mice based on natural vectors and machine learning algorithms
Project 2: Asymmetric Natural Vector Method for Predicting Ambiguous Non-standard Base Codes
Project 3: A Novel Framework for Predicting Genome Sequences and Gene Mutations Based on Natural Vector and Convex Hull Theory
Project 4: Discrimination of human exons and introns and prediction of splice-site based on a novel alignment-free method
Project 5: Fast Algorithm and Application of Asymmetric Covariance Natural Vector
Project 6: A Numerical Method of Gene Sequence Alignment Based on Energy Entropy
Project 7: Exploring the coordination mechanism of tertiary lymphoid structure and immunogenic cell death related genes in lung adenocarcinoma based on machine learning algorithms
Project 8: Screening of MERS-CoV targeting nanobodies
Project 9: Study on the mutation of covid19 gene from alpha to delta
Project 10: Study human chromosome map by finding out the intersection region R of Exon and Intron

Research Plan for 2025

Yishuai Niu：

Project 1: A Novel Framework for Predicting Genome Sequences and Gene Mutations Based on Natural Vector and Convex Hull Theory
Project 2: High-order Moment Portfolio Optimization via an Accelerated Difference-of-Convex Programming Approach and Sums-of-Squares
Project 3: Understand the Effectiveness of Shortcuts through the Lens of DCA
Project 4: On the Convergence Analysis of DCA
Project 5: Hybrid Accelerated DC Algorithms for the Asymmetric Eigenvalue Complementarity Problem

Applying NSFC (面上基金)

Random accelerated multi-block DC programming algorithm and its application

Research Plan for 2025

Qi Wu:

Project 1: Alignment-free methods

a theoretical framework of genetic for the Natural Vector method on genetic diversity analysis

long-term response to selection of complex trait

Project 2: Complex trait evolution

a wave model for evolutionary entities of quantitative trait, fitness, population size, etc.

a “wave interpretation” of unbiased parameter estimation

Research Plan for 2025

Xin Zhao：

Project 1:Research on the association analysis between genotype and phenotype of yeast and using AI technology to analyze large biological dataset.

Project 2: IGHV3-53 Sequence Characteristics and Neutralizing Activity.

Project 3:Single nucleotide polymorphism analyses for Major Depression Disorder.

Applying NSFC (面上基金)

Analysis method and application research of big data based on artificial intelligence and biometrics

Course Information

Spring Semester:

Deep Learning (Guoqing Hu);
Optimization algorithms in machine learning (Yishuai Niu)

Fall Semester:

Deep Learning Theory plus Practical Project Course (Guoqing Hu);
Introduction to Drinfeld module (Yishuai Niu);
Evolutionary genetics and bioinformatics (Qi Wu);
Advanced Theory of Statistics (Xin Zhao)

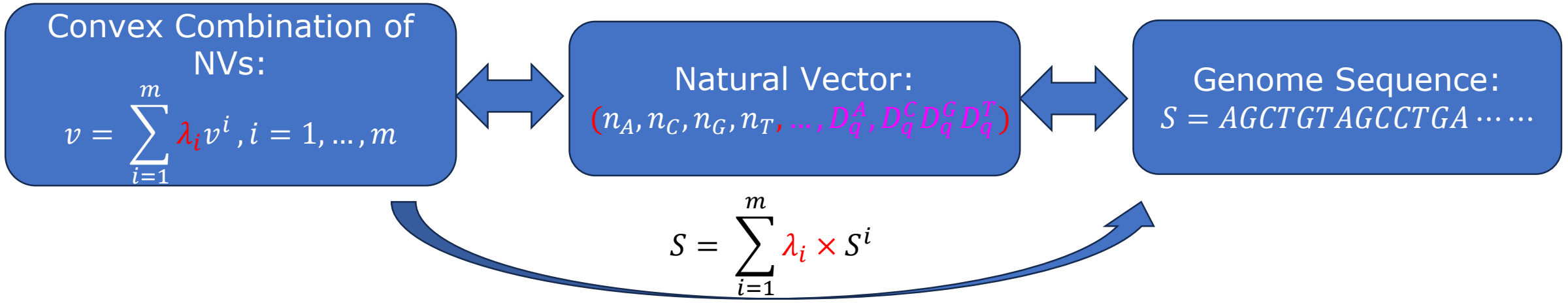
Significant Research in 2024

Prediction of Genome Sequences

Prediction of Genome Sequences

Key Objective: Predict new genome sequences based on convex hull of natural vectors.

Idea: Using the weight of convex combination for a non-vertex NV to serve as probability of prediction of genome sequence.



The probability of nucleotide for each position in the genome sequence should have a correspondence with the weights $\lambda_i, i = 1, \ldots, m$ in its convex combination.

Prediction of Genome Sequences

Step 1: Identify the maximum length of genome sequences
Identify the maximum length of genome sequences S^i associated with the NV $v^i, i = 1, \ldots, m$, denoted by $maxlen$

Step 2: Create the "prediction matrix" M

- Matrix dimensions: $4 \times maxlen$.
- Rows: A, C, G, T representing nucleotides.
- Columns: Positions $\mathrm{P}_1, \ldots, \mathrm{P}_{maxlen}$.
- $p_{s,j}$: Represents the probability of observing nucleotide $s \in \{A, C, G, T\}$ at position P_j.
- Each column will sum-up to 1.

M	$\mathbf{P_1}$	$\mathbf{P_2}$	$\mathbf{P_3}$	$\mathbf{P_4}$	$\cdots$	$\mathbf{P}_{maxlen}$
A	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$	$\cdots$	$p_{1,maxlen}$
C	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$p_{2,4}$	$\cdots$	$p_{2,maxlen}$
G	$p_{3,1}$	$p_{3,2}$	$p_{3,3}$	$p_{3,4}$	$\cdots$	$p_{3,maxlen}$
T	$p_{4,1}$	$p_{4,2}$	$p_{4,3}$	$p_{4,4}$	$\cdots$	$p_{4,maxlen}$

Prediction of Genome Sequences

Example:

- The first genome sequence $S^1 = AGATCTG\cdots$ gives a matrix T_1, where λ_1 is assigned to position P_j if the corresponding nucleotide is A, C, G or T, and 0 otherwise.

T_1	P_1	P_2	P_3	P_4	$\cdots$	P_{maxlen}
A	λ_1	0	0	0	$\cdots$	
C	0	0	λ_1	0	$\cdots$	
G	0	λ_1	0	0	$\cdots$	
T	0	0	0	λ_1	$\cdots$	

- The prediction matrix M is defined as the sum of $T_i, i = 1, \ldots, m$:

$$M = \sum_{i=1}^{m} T_i \,.$$

Remark: Each column of M could be normalized to sum-up to 1 (unnecessary).

Prediction of Genome Sequences

Step 3: Probability-Based Prediction

France_HDF-IPP11267_2021, 29915

Nucleotide	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_10	P_11	P_12	P_13	P_14	P_15
A	0.965646	0.011596	0.939326	0.035222	0.037049	0.016739	0	0.018376	0.013856	0.056923	0.055097	0.053027	0.027522	0.940383	0.946402
C	0.023192	0.025262	0	0.011163	0.928597	0.043501	0.00226	0.018672	0	0.93355	0.00226	0.930857	0.029401	0	0.042002
G	0.011163	0.930424	0.046194	0.004953	0	0.000434	0.943077	0.009526	0	0	0.008676	0.004953	0	0.018238	0.011596
T	0	0.032718	0.01448	0.948662	0.034354	0.939326	0.054663	0.953426	0.986144	0.009526	0.933967	0.011163	0.943077	0.041379	0
	A	G	A	T	C	T	G	T	T	C	T	C	T	A	A

France_BRE-IPP10815_2021, 29857

A kind of mutation

Nucleotide	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_10	P_3688	P_3689	P_3690	P_8085	P_8086	P_8087
A	0.887591	0.051005	0.901826	0.046322	0.010793	0.031247	0.046541	0.021579	0.041965	0.036654	0.87067	0.036891	0.929358	0.030789	0.023298	0.029332
C	0	0	0.060893	0.021579	0.863809	0.046541	0.014351	0.021816	0	0.916385	0.003933	0.876704	0	0.863809	0.905774	0.00257
G	0.074155	0.938202	0	0	0.083433	0.02111	0.874602	0.023782	0.047904	0.023179	0.036654	0.003933	0.028439	0.055972	0.021816	0
T	0.038254	0.010793	0.037281	0.932099	0.041965	0.901102	0.064505	0.932823	0.910131	0.023782	0.088743	0.082472	0.042202	0.04943	0.049112	0.968097
	A	G	A	T	C	T	G	T	T	C	A	C/T	A	C	C/T	T

Prediction of Genome Sequences

Observation:

- **Genome Recovery:** It is possible to recover the genome sequence from NV.
- **Mutation Detection:** It is possible to identify some mutations within the convex hull of the existing genome family.

Thanks for your attention