成果一.

刘晓东研究员与合作者的论文 Uniqueness and modified Newton method for cracks from the far field patterns with a fixed incident direction 被 INVERSE PROBLEMS 接收发表。

摘要：We consider the inverse cracks scattering problems from the far field patterns with a fixed incident direction. We firstly show that the sound-soft cracks can be uniquely determined by the multi-frequency far field patterns with a fixed incident direction. The proof is based on a low frequency asymptotic analysis of the scattered field. One important feature of the uniqueness result is that the background can even be an unknown inhomogeneous medium. A modified Newton method is then proposed for the numerical reconstruction of the shapes and locations of the cracks. Compared to the classical Newton method, the modified Newton method relaxes the dependence of a good initial guess and can be applied for multiple cracks. Numerical examples in two dimensions are presented to demonstrate the feasibility and effectiveness of the modified Newton method. In particular, the quality of the reconstructions can be greatly improved if we use the measurements properly with two frequencies or two incident directions.

论文链接：http://dx.doi.org/10.1088/1361-6420/ad904d

成果二.

张汉勤研究员与合作者的论文 ON SUBSTOCHASTIC INVERSE EIGENVALUE PROBLEMS WITH THE CORRESPONDING EIGENVECTOR CONSTRAINTS 被 SIAM JOURNAL ON MATRIX ANALYSIS AND APPLICATIONS 接收发表。

摘要： We consider the inverse eigenvalue problem of constructing a substochastic matrix from the given spectrum parameters with the corresponding eigenvector constraints. This substochastic inverse eigenvalue problem (SstIEP) with the specific eigenvector constraints is formulated into a nonconvex optimization problem (NcOP). The solvability for SstIEP with the specific eigenvector constraints is equivalent to identifying the attainability of a zero optimal value for the formulated NcOP. When the optimal objective value is zero, the corresponding optimal solution to the formulated NcOP is just the substochastic matrix that we wish to construct. We develop the alternating minimization algorithm to solve the formulated NcOP, and its convergence is established by developing a novel method to obtain the boundedness of the optimal solution. Some numerical experiments are conducted to demonstrate the efficiency of the proposed method.

论文链接：http://dx.doi.org/10.1137/23M1547305

成果三.

孙六全研究员与合作者的论文 Inference on High-dimensional Single-index Models with Streaming Data 被 JOURNAL OF MACHINE LEARNING RESEARCH 接收发表。

摘要：Traditional statistical methods are faced with new challenges due to streaming data. The major challenge is the rapidly growing volume and velocity of data, which makes storing such huge data sets in memory impossible. The paper presents an online inference framework for regression parameters in high-dimensional semiparametric single-index models with unknown link functions. The proposed online procedure updates only the current data batch and summary statistics of historical data instead of re-accessing the entire raw data set. At the same time, we do not need to estimate the unknown link function, which is a highly challenging task. In addition, a generalized convex loss function is used in the proposed inference procedure. To illustrate the proposed method, we use the Huber loss function and the negative log-likelihood of the logistic regression model. In this study, the asymptotic normality of the proposed online debiased Lasso estimators and the bounds of the proposed online Lasso estimators are investigated. To evaluate the performance of the proposed method, extensive simulation studies have been conducted. We provide applications to Nasdaq stock prices and financial distress data sets.
论文链接：https://jmlr.org/papers/volume25/22-1124/22-1124.pdf

成果四.

张世华研究员与合作者的论文 Statistical batch-aware embedded integration, dimension reduction, and alignment for spatial transcriptomics 被 BIOINFORMATICS 接收发表。

摘要：Motivation: Spatial transcriptomics (ST) technologies provide richer insights into the molecular characteristics of cells by simultaneously measuring gene expression profiles and their relative locations. However, each slice can only contain limited biological variation, and since there are almost always non-negligible batch effects across different slices, integrating numerous slices to account for batch effects and locations is not straightforward. Performing multi-slice integration, dimensionality reduction, and other downstream analyses separately often results in suboptimal embeddings for technical artifacts and biological variations. Joint modeling integrating these steps can enhance our understanding of the complex interplay between technical artifacts and biological signals, leading to more accurate and insightful results. Results: In this context, we propose a hierarchical hidden Markov random field model STADIA to reduce batch effects, extract common biological patterns across multiple ST slices, and simultaneously identify spatial domains. We demonstrate the effectiveness of STADIA using five datasets from different species (human and mouse), various organs (brain, skin, and liver), and diverse platforms (10x Visium, ST, and SliceseqV2). STADIA can capture common tissue structures across multiple slices and preserve slice-specific biological signals.

In addition, STADIA outperforms the other three competing methods (PRECAST, fastMNN, and Harmony) in terms of the balance between batch mixing and spatial domain identification, and it demonstrates the advantage of joint modeling when compared to STAGATE and GraphST.
论文链接：http://dx.doi.org/10.1093/bioinformatics/btae611

成果五.

骆顺龙研究员与合作者的论文 Quantifying incompatibility between positive operator-valued measures via negativity of the Jordan product 被 PHYSICAL REVIEW A 接收发表。

摘要：A fundamental feature of quantum mechanics is the incompatibility between quantum measurements, which leads to the Heisenberg uncertainty relations and is intrinsically related to detection of many other quantum features such as nonlocality and steering. Given the significance and ubiquity of incompatibility, it is desirable to quantify the degree of incompatibility between two quantum measurements which are mathematically described by positive operator valued measures. The commutator (Lie product) between operators has widely been used in quantifying incompatibility (noncommutativity) and plays a crucial role in quantum mechanics. In contrast, the anticommutator (Jordan product) between operators has been relatively less employed in the study of quantum mechanics. In this work, we explore the role of the Jordan product in characterizing measurement incompatibility. The key observation lies in that while the product of two non-negative classical observables is automatically non-negative, this is not the case for quantum observables. Indeed the Jordan product of two non- negative observables is not necessarily non-negative. This negativity of the Jordan product is a purely quantum phenomenon and is here exploited to quantify a kind of measurement incompatibility, which interpolates between noncommutativity and joint measurability. We reveal basic features of this approach, make some comparisons with existing quantifiers of incompatibility and suggest the issue of investigating its applications and implications in quantum information tasks.
论文链接：http://dx.doi.org/10.1103/PhysRevA.110.062206

成果六.

孙六全研究员与合作者的论文 Large-scale survival analysis with a cure fraction 被 BIOMETRICS 接收发表。

摘要：With the advent of massive survival data with a cure fraction, large-scale regression for analyzing the effects of risk factors on a general population has become an emerging challenge. This article proposes a new probability-weighted method for estimation and inference for semiparametric cure regression models. We develop a flexible formulation of the mixture cure model consisting of the model-free incidence and the latency assumed by the semiparametric proportional hazards model. The susceptible probability assesses the concordance between the observations and the latency. With the susceptible probability as weight, we propose a weighted estimating

equation method in a small-scale setting. Robust nonparametric estimation of the weight permits stable implementation of the estimation of regression parameters. A recursive probability-weighted estimation method based on data blocks with smaller sizes is further proposed, which achieves computational and memory efficiency in a large-scale or online setting. Asymptotic properties of the proposed estimators are established. We conduct simulation studies and a real data application to demonstrate the empirical performance of the proposed method.
论文链接：http://dx.doi.org/10.1093/biomtc/ujae138

成果七.

张世华研究员与合作者的论文 Uncovering topologically associating domains from three-dimensional genome maps with TADGATE 被 NUCLEIC ACIDS RESEARCH 接收发表。

摘 要： Topologically associating domains (TADs) are essential components of three-dimensional (3D) genome organization and significantly influence gene transcription regulation. However, accurately identifying TADs from sparse chromatin contact maps and exploring the structural and functional elements within TADs remain challenging. To this end, we develop TADGATE, a graph attention auto-encoder that can generate imputed maps from sparse Hi-C contact maps while adaptively preserving or enhancing the underlying topological structures, thereby facilitating TAD identification. TADGATE captures specific attention patterns with two types of units within TADs and demonstrates TAD organization relates to chromatin compartmentalization with diverse biological properties. We identify many structural and functional elements within TADs, with their abundance reflecting the overall properties of these domains. We applied TADGATE to sparse and noisy Hi-C contact maps from 21 human tissues or cell lines. That improved the clarity of TAD structures, allowing us to investigate conserved and cell-type-specific boundaries and uncover cell-type-specific transcriptional regulatory mechanisms associated with topological domains. We also demonstrated TADGATE's capability to fill in sparse single-cell Hi-C contact maps and identify TAD-like domains within them, revealing the specific domain boundaries with distinct heterogeneity and the shared backbone boundaries characterized by strong CTCF enrichment and high gene expression levels.
论文链接：http://dx.doi.org/10.1093/nar/gkae1267

成果八.

常晋源研究员与合作者的论文 On the Modeling and Prediction of High-Dimensional Functional Time Series 被 JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 接收发表。

摘 要： We propose a two-step procedure to model and predict high-dimensional functional time series, where the number of function-valued time series p is large in relation to the length of time series n. Our first step performs an eigenanalysis of a

positive definite matrix, which leads to a one-to-one linear transformation for the original high-dimensional functional time series, and the transformed curve series can be segmented into several groups such that any two subseries from any two different groups are uncorrelated both contemporaneously and serially. Consequently in our second step those groups are handled separately without the information loss on the overall linear dynamic structure. The second step is devoted to establishing a finite-dimensional dynamical structure for all the transformed functional time series within each group. Furthermore the finite-dimensional structure is represented by that of a vector time series. Modeling and forecasting for the original high-dimensional functional time series are realized via those for the vector time series in all the groups. We investigate the theoretical properties of our proposed methods, and illustrate the finite-sample performance through both extensive simulation and two real datasets. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.
论文链接：http://dx.doi.org/10.1080/01621459.2024.2413201

成果九.

王勇研究员与合作者的论文 Integrative study of lung cancer adeno-to-squamous transition in EGFR TKI resistance identifies RAPGEF3 as a therapeutic target 被 NATIONAL SCIENCE REVIEW 接收发表。

摘要：

Although adeno-to-squamous transition (AST) has been observed in association with resistance to epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) in clinic, its causality, molecular mechanism and overcoming strategies remain largely unclear. We here demonstrate that squamous transition occurs concomitantly with TKI resistance in PC9-derived xenograft tumors. Perturbation of squamous transition via DNp63 overexpression or knockdown leads to significant changes in TKI responses, indicative of a direct causal link between squamous transition and TKI resistance. Integrative RNA-seq, ATAC-seq analyses and functional studies reveal that FOXA1 plays an important role in maintaining adenomatous lineage and contributes to TKI sensitivity. FOXM1 overexpression together with FOXA1 knockout fully recapitulates squamous transition and TKI resistance in both PC9 xenografts and patient-derived xenograft (PDX) models. Importantly, pharmacological inhibition of RAPGEF3 combined with EGFR TKI efficiently overcomes TKI resistance, especially in RAPGEF3^{high} PDXs. Our findings provide novel mechanistic insights into squamous transition and therapeutic strategy to overcome EGFR TKI resistance in lung cancer.

论文链接：http://dx.doi.org/10.1093/nsr/nwae392