

## DOCTORAL THESIS

### Order determination for large matrices with spiked structure

Zeng, Yicheng

*Date of Award:*  
2019

[Link to publication](#)

#### **General rights**

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

# Abstract

Motivated by dimension reduction in regression analysis and signal detection, we investigate order determination for large dimensional matrices with spiked structures in which the dimensions of the matrices are proportional to the sample sizes. Because the asymptotic behaviors of the estimated eigenvalues differ completely from those in fixed dimension scenarios, we then discuss the largest possible order, say  $q$ , we can identify and introduce criteria for different settings of  $q$ . When  $q$  is assumed to be fixed, we propose a “valley-cliff” criterion with two versions - one based on the original differences of eigenvalues and the other based on the transformed differences - to reduce the effect of ridge selection in the criterion. This generic method is very easy to implement and computationally inexpensive, and it can be applied to various matrices. As examples, we focus on spiked population models, spiked Fisher matrices and factor models with auto-covariance matrices. For the case of divergent  $q$ , we propose a scale-adjusted truncated double ridge ratio (STD RR) criterion, where a scale adjustment is implemented to deal with the bias in scale parameter for large  $q$ . Again, examples include spiked population models, spiked Fisher matrices. Numerical studies are conducted to examine the finite sample performances of the method and to compare it with existing methods. As for theoretical contributions, we investigate the limiting properties, including convergence in probability and central limit theorems, for spiked eigenvalues of spiked Fisher matrices with divergent  $q$ .

**Keywords:** Auto-covariance matrix, factor model, finite-rank perturbation, Fisher matrix, principal component analysis (PCA), phase transition, random matrix theory (RMT), ridge ratio, spiked population model.

# Acknowledgements

First and foremost, I would like to express my deep gratitude to my supervisor and mentor Prof. ZHU Lixing for all of his inspiring guidance and enthusiastic support. His superb intuition, broad knowledge, and continuous encouragement have been indispensable throughout my PhD study and are extraordinarily beneficial in my future research career. And I learned lots of various things from him in the fields of statistics during these years. It is my great honor to be his student. I would also like to thank my co-supervisor Dr. PENG Heng, for his help and support. Last but not least, thank my family, my wife Yuhang and my friends for their love and support.

# Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	vii
List of Tables	viii
Chapter 1 Introduction	1
1.1 Introduction to Order Determination . . . . .	1
1.2 Random Matrix Theory in Statistics . . . . .	4
1.2.1 A Brief Review of Random Matrix Theory . . . . .	4
1.2.2 Selected Results for Random Matrix Theory in Statistics . . . . .	5
1.3 Our Contributions and Structure . . . . .	8
Chapter 2 Order Determination for Spiked Models with $q$ fixed: Detection of Significant Spiked Eigenvalues	9
2.1 Introduction . . . . .	9
2.2 Criterion construction and properties . . . . .	12
2.2.1 Motivation . . . . .	12
2.2.2 Valley-cliff criterion and estimation consistency . . . . .	14
2.3 Modification of the VACLE . . . . .	17
2.4 Applications to several models . . . . .	20

2.5	Numerical Studies . . . . .	25
2.5.1	Scale estimation . . . . .	25
2.5.2	Simulations about spiked population models . . . . .	26
2.5.3	Numerical studies on large dimensional auto-covariance matrix . . . . .	31
2.5.4	Numerical studies on large dimensional spiked Fisher matrix . . . . .	34
2.6	Real data example . . . . .	37
2.7	Concluding remarks . . . . .	38
2.8	Proofs . . . . .	39
Chapter 3	Limiting laws for spiked eigenvalues of large-dimensional spiked Fisher matrices with divergent $q$ . . . . .	49
3.1	Notations and Assumptions . . . . .	49
3.2	Convergence of Spiked Eigenvalues . . . . .	53
3.3	Central Limit Theorem for Spiked Eigenvalues . . . . .	66
3.3.1	Definitions of centering parameter $\theta_i$ . . . . .	66
3.3.2	The simple eigenvalue case . . . . .	68
3.3.3	The multiple eigenvalue case . . . . .	74
3.4	Some Technical Lemmas . . . . .	76
Chapter 4	Order Determination for Spiked Models with $q$ Divergent: Detection of Significant Spiked Eigenvalues . . . . .	87
4.1	Challenges for Order Determination in Large $q$ Case . . . . .	88
4.2	Scale-adjusted Truncated Double Ridge Ratios Criterion . . . . .	91
4.2.1	Scale adjustment . . . . .	92
4.2.2	Criterion construction and properties . . . . .	93
4.2.3	Weak signals detection . . . . .	97
4.3	Applications to Models . . . . .	98
4.3.1	Spiked population models . . . . .	98
4.3.2	Spiked Fisher matrices . . . . .	101
4.4	Numerical Studies . . . . .	102
4.4.1	Simulations on Spiked population models . . . . .	103
4.4.2	Simulations on weak signals . . . . .	106

4.4.3	Simulaitons on Spiked Fisher matrices . . . . .	107
4.5	Proofs . . . . .	109
4.5.1	Proof of Theorem 4.2.1 . . . . .	109
4.5.2	Proof of Theorem 4.2.2 . . . . .	110
4.5.3	Proof of Theorem 4.3.1 . . . . .	111
4.5.4	Proof of Theorem 4.3.2 . . . . .	112
4.5.5	Proof of Proposition 4.3.1 . . . . .	114
4.6	Tables of Results for Simulations . . . . .	114
	Curriculum Vitae	130