RGC Ref.: N_CityU120/14 NSFC Ref. :
(please insert ref. above)

The Research Grants Council of Hong Kong
NSFC/RGC Joint Research Scheme
Joint Completion Report

(Please attach a copy of the completion report submitted to the NSFC by the Mainland researcher)

Part A: The Project and Investigator(s)

1. Project Title

Approximation Analysis of Information Theoretic Learning and Ranking Type Learning Problems
信息論學習和排序型學習算法的逼近分析

2. Investigator(s) and Academic Department/Units Involved

	Hong Kong Team	Mainland Team
Name of Principal Investigator *(with title)*	Prof. ZHOU Dingxuan	Prof. WU Zongmin
Post	Chair Professor	Professor
Unit / Department / Institution	School of Data Science	Fudan University
Contact Information	mazhou@cityu.edu.hk	
Co-investigator(s) *(with title and institution)*		Dr. SHI Lei (Fudan University)

3. Project Duration

	Original	Revised	Date of RGC/ Institution Approval *(must be quoted)*
Project Start date	1 Jan 2015		
Project Completion date	31 Dec 2018		
Duration *(in month)*	48		
Deadline for Submission of Completion Report	31 Dec 2019		

Part B: The Completion Report

5. Project Objectives

5.1 Objectives as per original application

1. To carry out mathematical analysis of some information theoretic learning algorithms. Mathematical analysis will be developed for a minimum error entropy algorithm in the empirical risk minimization setting when the scaling parameter is small and the model is not homoscedastic. Error analysis will be carried out for the corresponding regularization schemes. Information theoretic learning algorithms induced by Wasserstein metric will be analyzed by scaling-based approximation schemes.
2. To make Fourier analysis of some ranking type learning algorithms. Mathematical analysis of ranking algorithms based on scoring functions will be conducted by characterizing a

minimizer of the associated generalization error using the generalized Fourier transform. Error analysis of some regularization schemes for metric learning and similarity learning will be provided by making Fourier analysis of a minimizer of the generalization error and estimating the approximation error. Robustness of ranking type regularization schemes will be analyzed.

3. To study some approximation theory problems arising from learning theory. Theory of approximation measured by correntropy will be developed and used for studying the corresponding regularization schemes. Essential differences between additive kernels and classical kernels will be investigated by methods from approximation theory of radial basis functions. Approximation schemes will be analyzed with respect to the Wasserstein metric.

5.2 Revised Objectives

Date of approval from the RGC: ____________________

Reasons for the change: ____

____

1.
2.
3. ....

6. Research Outcome

Major findings and research outcome
(maximum 1 page; please make reference to Part C where necessary)

Pairwise learning refers to a learning task involving a loss function depending on example pairs. It includes algorithms for ranking, metric learning, similarity learning, and AUC maximization. In [1] we study an online algorithm for pairwise learning with a least square loss function in a reproducing kernel Hilbert space. By some ideas of Fourier analysis and estimating generalization error and approximation error with a family of pairwise kernels, we establish a general convergence theorem and derive explicit convergence rates. In [2] kernel-based regularized pairwise learning methods are investigated, including the examples with the error entropy loss for information theoretic learning and ranking loss. Nice statistical robustness properties are presented. In [3] we study the online composite mirror descent algorithm, which involves a mirror map to reflect the geometry of the data and a convex loss possibly inducing sparsity. Our error analysis provides convergence rates in terms of properties of the strongly convex differentiable mirror map and the objective function. Our methodology mainly depends on a novel error decomposition in terms of an excess Bregman distance which is key in information theoretic learning. Our paper [4] is concerned with distributed spectral algorithms, for handling big data, based on a divide-and-conquer approach. We present a learning theory for these algorithms in a regression framework including nice error bounds and optimal minimax learning rates achieved by solving an approximation problem involving the integral operators induced by kernels. In [5] a regression model associated with the correntropy induced losses is studied. The correntropy, as a similarity measure in information theoretic learning, induces the maximum correntropy criterion for regression for which convergence and robustness are analyzed by solving the related approximation theory problems. In [6] a learning algorithm for finding the entire solution path for support vector machines induced by pinball losses with flexible quantile parameters is proposed. It is based on the observation that the optimal solution a linear spline with the quantile parameter which leads to an interesting approximation theory problem on splines. In [7] we study two learning algorithms generated by kernel partial least squares and kernel minimal residual methods. We propose a stopping rule for determining the number of iterations based on cross-validation, without assuming a priori knowledge of the underlying probability measure, and show that optimal learning rates can be achieved. Our novel analysis with nice bounds for the number of iterations uses technical tools and methods from approximation theory. In [8] we study the generalization ability of distributed learning equipped with a gradient descent algorithm in a reproducing kernel Hilbert space. By solving an approximation theory problem involving special spectral features of the gradient descent algorithms, we provide the optimal learning rates and partly conquer the saturation phenomenon in the literature. In [9], we study data-dependent generalization error bounds that exhibit a mild dependency on the number of classes, making them suitable for multi-class learning with a large number of label classes and information theoretic learning. Key to our analysis is new structural results for multi-class Gaussian complexities and uniform covering numbers derived by ideas from scaling-based approximation schemes. In [10] convergence of online gradient descent algorithms in reproducing kernel Hilbert spaces without regularization is considered. A necessary condition and sufficient condition for the convergence of excess generalization errors in expectation are established. The results are derived by cancelling out the variances of the involved martingales by using descent properties of the algorithm and some ideas from approximation theory. In [11], we propose and study the notion of modified Fejér sequences.By solving a related approximation theory problem, we show that it provides a unifying framework to prove convergence rates for objective function values of several optimization algorithms such as forward–backward splitting algorithm, incremental subgradient proximal algorithm, and the Douglas–Rachford splitting method.

More research findings from this project under the NSFC/RGC joint scheme are described in the research outcome and the completion report of the Mainland Team.

Potential for further development of the research and the proposed course of action *(maximum half a page)*

Analysis of information theoretic learning and ranking type learning algorithms and the raised approximation theory problems lead to more research questions and learning tasks involving big data such as mathematical foundations of deep learning schemes. We shall consider these topics in our further study.

7. The Layman's Summary

(describe in layman's language the nature, significance and value of the research project, in no more than 200 words)

Theory of learning with the classical least squares loss has been well developed in mathematics, based on probability analysis, statistics, and approximation theory. Information theoretic learning is a different learning framework using descriptors from information theory to substitute the conventional statistical descriptors of variance and covariance in the least squares method for processing non-Gaussian noise. Ranking type learning problems aim at efficient algorithms involving sample pairs, which is different from methods for regression or classification. In this project we develop rigorous mathematical analysis for some problems in these two topics by methods and ideas from approximation theory and wavelet analysis. We establish error analysis and robustness of some learning algorithms for information theoretic learning, pairwise learning, gradient descent and online learning, distributed learning for big data, and kernel methods. Some approximation theory problems arising from learning theory are investigated. The produced research output, ideas and methods are used in our error and robustness analysis of the online and offline schemes for information theoretic learning and pairwise learning.

Part C: Research Output

8. Peer-reviewed journal publication(s) arising directly from this research project

(Please attach a copy of each publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.)

The Latest Status of Publications				Author(s) *(bold the authors belonging to the project teams and denote the corresponding author with an asterisk*)*	Title and Journal/ Book *(with the volume, pages and other necessary publishing details specified)*	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this report *(Yes or No)*	Acknowledged the support of this Joint Research Scheme *(Yes or No)*	Accessible from the institutional repository *(Yes or No)*
Year of publication	Year of Acceptance *(For paper accepted but not yet published)*	Under Review	Under Preparation *(optional)*						
2016				YimingYing* and **Ding-Xuan Zhou**	Online pairwise learning algorithms, Neural Computation vol. 28, pp. 743--777.	Yes (2016)	Yes	Yes	Yes
2016				Andreas Christmann* and **Ding-Xuan Zhou**	On the Robustness of Regularized Pairwise Learning Methods Based on Kernels, J. Complexity vol. 37, 1--33.	Yes (2016)	Yes	Yes	Yes

2017				Yunwen Lei* and **Ding-Xuan Zhou**	Analysis of online composite mirror descent algorithm, Neural Computation, vol. 29, 825–860.	Yes (2016)	Yes	Yes	Yes
2017				Zhengchu Guo, Shao-Bo Lin*, and **Ding-Xuan Zhou**	Learning theory of distributed spectral algorithms, Inverse Problems vol. 33 074009 (29pp).	Yes (2016)	Yes	Yes	Yes
2015				Yunlong Feng*, Xiaolin Huang, **Lei Shi**, Yuning Yang and Johan AK Suykens	Learning with the maximum correntropy criterion induced losses for regression. J. Mach. Learn. Res., vol 16: 993-1034.	Yes (2016)	Yes	Yes	Yes
2017				Xiaolin Huang*, **Lei Shi** and Johan AK Suykens	Solution path for pin-SVM classifiers with positive and negative τ values. IEEE Trans. Neural Netw. Learn. Syst. Vol. 28, 1584-1593.	Yes (2016)	Yes	Yes	Yes
2018				Shao-Bo Lin* and **Ding-Xuan Zhou**	Optimal Learning Rates for Kernel Partial Least Squares, J. Fourier Anal. Appl. Vol. 24 908--933.	No	Yes	Yes	Yes
2018				Shao-Bo Lin* and **Ding-Xuan Zhou**	Distributed kernel-based gradient descent algorithms, Constructive Approximation vol.47, 249—276.	No	Yes	Yes	Yes
2019				Yunwen Lei* , Ürün Dogan, **Ding-Xuan Zhou**, and Marius Kloft	Data-dependent generalization bounds for multi-class classification, IEEE Trans. Inform. Theory vol. 65, 2995-3021.	No	Yes	Yes	Yes
2018				Yunwen Lei, **Lei Shi**, and Zheng-Chu Guo*	Convergence of Unregularized Online Learning Algorithms, Journal of Machine Learning Research vol. 18, 1-33.	No	Yes	Yes	Yes
2018				Junhong Lin, Lorenzo Rosasco, Silvia Villa*, and **Ding-Xuan Zhou**	Modified Fejér sequences and applications, Computational Optimization and Applications vol. 71, 95—113.	No	Yes	Yes	Yes

9. Recognized international conference(s) in which paper(s) related to this research project was/were delivered *(Please attach a copy of each delivered paper. All listed papers must acknowledge RGC's funding support by quoting the specific grant reference.)*

Month/Year / Place	Title	Conference Name	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this report *(Yes or No)*	Acknowledged the support of this Joint Research Scheme *(Yes or No)*	Accessible from the institutional repository *(Yes or No)*
11/2015/Shanghai	Distributed learning with big data	the First International Conference on Data Science: Foundation and Applications	Yes (2016)	No	Yes	Yes
11/2016/Nanjing	Analysis of Distributed Learning Algorithms	Workshop on Machine Learning and Applications 2016	Yes (2016)	No	Yes	Yes
12/2016/Hangzhou	Approximation Analysis of Distributed Learning and Online Learning	International Conference on Some Mathematical Approximation Approaches in Data Science	Yes (2016)	No	Yes	Yes
05/2017/Hangzhou	Distributed Learning	Mini-symposium on Modern Regularization Techniques in Data-based Learning, the 9th Applied Inverse Problem Conference	No	Yes	Yes	Yes
06/2017/Tianjin	Analysis of Distributed Learning	International Workshop on Computational Harmonic Analysis	No	Yes	Yes	Yes
10/2017/Beijing	Analysis of Distributed Learning	International Conference on Harmonic Analysis and Its Applications	No	Yes	Yes	Yes
12/2017/Sanya	Theory of Distributed Learning	From Approximation Theory to real World Applications Workshop	No	Yes	Yes	Yes
10/2017/Tel Aviv	Distributed Learning	The Third CityU-TAU Joint Workshop on Mathematical Analysis and Applications	No	Yes	Yes	Yes
09/2017/Zhuhai	Distributed Learning	2017 Workshop on Mathematics for Data Sciences	No	Yes	Yes	Yes
04/2017/Hangzhou	Distributed Learning for Handling Big Data	Zhejiang University Forum on Mathematics	No	Yes	Yes	Yes

10. Student(s) trained *(Please attach a copy of the title page of the thesis.)*

Name	Degree registered for	Date of registration	Date of thesis submission/ graduation
Martin Boissier	PhD	01-09-2013	31-08-2016
Bo Zhang	PhD	01-09-2014	31-12-2017

11. Other impact *(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, etc.)*

With the funding from NSFC, the PI and the Mainland Team co-organized the International Workshop on Mathematical Aspects of Data Science 2016 at Fudan University during May 20 - 23, 2016 and the International Conference on Computational Harmonic Analysis 2017 at Fudan University during May 24-28, 2017. The PI was rated as a Highly-cited Researcher (Mathematics) by Clarivate Analytics in 2016 and 2017. Some research work from this project contributed to this honor.