

DOCTORAL THESIS

Numerical algorithms for data processing and analysis

Chen, Chuan

Date of Award:
2016

[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

In this thesis, we investigate several numerical algorithms for data processing and analysis specifically the total variation (TV) based models for data processing and multi-domain data clustering.

The first part of the thesis is to study TV based methods for both image restoration and tensor decomposition problems. In Chapter 2, we introduce the image restoration problems which play an important role in many applications of sciences and engineering. Different from existing approaches, both total variation regularization and nonlinear least squares data-fitting are taken into consideration. By making use of the structure of the objective function, an efficient alternating direction method of multipliers (ADM) can be developed for solving the proposed model. Simulation study is carried out to examine the performance of the proposed method in nonlinear image restoration. The method is also applied to high dynamic range imaging for illustration.

In Chapter 3, we further consider TV based tensor decomposition data with time dimension. Based on CANDECOMP/PARAFAC (CP) decomposition method, we employ the total variation regularization term in the time dimension to enforce the time factor constraint. We also employ ADM to solve the resulting optimization problem and study its convergence. Simulation studies are also carried out to assess the performance of our proposed method. Both video data and gene expression data analysis are conducted for illustrations.

For the second part of this thesis, we mainly focus on numerical methods for multi-domain data clustering, including multi-view data and multi-relation data. In Chapter 4, we study clustering for multiple undirected graphs over the same set of objects in different views, known as multi-view data. Different from existing approaches which seek for common clusters for all graphs, we search the clustering of objects within each graph while at the same time should be highly consistent across different graphs. We propose a simple but novel block spectral clustering method for multi-view data. We demonstrate that the multiplicity of the zero eigenvalue of the

constructed block Laplacian matrix is equal to the number of connected components in multiple graphs while the corresponding eigenvectors are solutions of the relaxation of multiple graphs cut problems. Simulations are conducted to examine the performance, while both publication data and gene expression data are analysed for illustration.

In Chapter 5, we further extend the clustering method for multi-view data to handle multi-relation data, i.e., multiple graphs over different sets of objects. With inter-relation among different graphs further considered in the block spectral clustering framework, we construct block Laplacian matrix with off-diagonal blocks representing inter-relation. By making use of its eigenvectors to perform clustering, global optimal solutions are obtained in the proposed method. Experiments on synthetic data sets are given to examine the clustering accuracy and computational time required by the proposed block clustering methods. Both publication data and gene expression data are conducted for illustration.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	ix
List of Figures	xi
Chapter 1 Introduction	1
1.1 Image Restoration Problems	3
1.1.1 Nonlinear Image Restoration Problems	3
1.1.2 Contribution	4
1.2 Tensor Decomposition for Multi-dimensional Data with Time Dimension	4
1.2.1 CANDECOMP/PARAFAC Decomposition Method	5
1.2.2 High-dimensional Data with Time Dimension	6
1.2.3 Contribution	7
1.3 Clustering for Multi-view Data	9
1.3.1 Multi-view Data	9
1.3.2 Contribution	10
1.4 Clustering for Multi-relation Data	11

1.4.1	Multi-relation Data	12
1.4.2	Contribution	13
1.5	Outline of the Thesis	13
Chapter 2 Image Restoration Problem		17
2.1	Related Work	17
2.2	Numerical Method for TVNLS Model	19
2.3	Experimental Results	24
2.3.1	Nonlinear Image Restoration	25
2.3.2	The Influence of Parameter Settings	31
2.3.3	High Dynamic Range Imaging	34
2.4	Concluding Remarks	37
2.5	Appendices	37
2.5.1	Proof of Theorem 2.1	37
2.5.2	Proof of Theorem 2.2	43
Chapter 3 Tensor Decomposition for Multi-dimensional Data with Time		
	Dimension	46
3.1	Numerical Method for TVCP Model	47
3.1.1	Convergence Analysis	49
3.2	Numerical Examples	53
3.2.1	Synthetic Data	54
3.2.2	Video Data	58
3.2.3	Gene Expression Data	63
3.3	Concluding Remarks	65
3.4	Appendices	66
3.4.1	Proof of Lemma 3.1	66
3.4.2	Proof of Lemma 3.3	67
3.4.3	Proof of Theorem 3.1	67

Chapter 4 Clustering for Multi-view Data	70
4.1 Related Work	70
4.2 Block Spectral Clustering Method	72
4.2.1 Spectral Clustering	72
4.2.2 Basic Idea	72
4.2.3 Block Laplacian Matrices	73
4.2.4 Block Intra-normalized Laplacian Matrices	78
4.3 The Proposed Algorithm	81
4.3.1 Multiple Graphs Cut	82
4.3.2 The Bounds for the Optimal Values of J_1 and J_2	85
4.4 Experimental Results	87
4.4.1 Synthetic Data	88
4.4.2 Publication Data	93
4.5 Concluding Remarks	99
4.6 Appendices	101
4.6.1 Proof of Theorem 4.2	101
4.6.2 Proof of Theorem 4.3	102
4.6.3 Proof of Lemma 4.2	104
4.6.4 Proof of Lemma 4.3	104
Chapter 5 Clustering for Multi-relation Data	109
5.1 Related Work	109
5.2 Block Spectral Clustering Methods	111
5.2.1 Spectral Clustering	111
5.2.2 Spectral Clustering for Two Graphs	113
5.2.3 Multiple Graphs Spectral Clustering	115
5.3 Multiple Graph Cut Problems	117
5.3.1 Multiple Graph Cut Criteria	119
5.3.2 The Ratio Cut Approach	120

5.3.3	The Normalized Cut Approach	121
5.4	Experimental Results	122
5.4.1	Synthetic Data Sets	122
5.4.2	Real Data Sets	129
5.5	Concluding Remarks	141
5.6	Appendices	143
5.6.1	Proof of Theorem 5.1	143
5.6.2	Proof of Theorem 5.2	143
5.6.3	Proof of Theorem 5.3	145
5.6.4	Proof of Theorem 5.4	146
Chapter 6 Conclusion and Future Work		147
6.1	Conclusion	147
6.2	Future Work	148
Bibliography		150
Curriculum Vitae		162