

Winter Workshop on  
Dimension Reduction and High Dimensional Inference

Department of Statistics  
University of Florida

January 17-18, 2014

# Contents

<b>Location</b>	<b>1</b>
<b>Sponsors</b>	<b>1</b>
<b>Organizing Committee</b>	<b>1</b>
<b>Invited Speakers</b>	<b>1</b>
<b>Participants</b>	<b>1</b>
<b>Acknowledgements</b>	<b>2</b>
<b>Thursday, January 16</b>	<b>2</b>
<b>Friday, January 17</b>	<b>3</b>
<b>Saturday, January 18</b>	<b>4</b>
<b>Abstracts</b>	<b>5</b>
Valid Post-Selection Inference.....	5
Computationally Efficient Minimax Adaptive Estimation of Large Covariance Matrices with Low Dimensional Structures .....	5
Ensemble Subsampling for Imbalanced Multivariate Two-Sample Tests .....	6
Prediction in Abundant High-dimensional Linear Regression.....	6
On an Additive Semi-Graphoid Model for Statistical Networks with Application to Pathway Analysis.....	7
Tensor Regression, Regularization, and Neuroimaging Data Analysis.....	7
On Estimation Efficiency in Dimension Reduction.....	8
Properties of Optimizations Used in Penalized Gaussian Likelihood Inverse Covariance Matrix Estimation.....	8
Optimal Detection of Sparse Signal Segments.....	8
Asymptotic Normality and Efficiency In Estimation of High-dimensional Graphical Models	9
<b>Poster Abstracts</b>	<b>10</b>
Interpolation of Computationally Expensive Posterior Densities with Variable Parameter Costs .....	10
Modeling Financial Volatility: An Exogenous Log-GARCH Approach.....	10
Identification of Important Regressor Groups, Subgroups, and Individuals via Regulariza- tion Methods: Application to Gut Microbiome Data.....	11
Making the Cut: Ranking and Selection Procedures for Large-Scale Inference.....	11
Covariance Estimation for Multivariate Longitudinal Data.....	12
Meta-Analysis Based Variable Selection for Gene Expression Data.....	12
Assessing Protein Conformational Sampling Methods Based on Bivariate Lag-Distributions of Backbone Angles.....	13
Dimension Reduction Using Inverse Spline Regression.....	13

Latent Supervised Learning for Survival Data.....	14
Stable Estimation in Dimension Reduction.....	14
Principal Flows.....	15
Fused Estimators of the Central Subspace in Sufficient Dimension Reduction.....	15
<b>Gainesville Restaurants</b>	<b>16</b>

## **Location**

Ail sessions will be held in Emerson Alumni Hall (EAH), 1938 W. University Avenue, Gainesville, FL.

## **Sponsors**

Info Tech, Inc.; National Science Foundation; UF Office of Research; UF College of Liberal Arts and Sciences; UF Department of Statistics.

## **Organizing Committee**

Hani Doss, Malay Ghosh, Zhihua (Sophia) Su.

## **Invited Speakers**

Andreas Buja, University of Pennsylvania  
Florentina Bunea, Cornell University  
Fisha Chen, Yale University  
Dennis Cook, University of Minnesota  
Bing Fi, Pennsylvania State University  
Fexin Fi, North Carolina State University  
Yanyuan Ma, Texas A&M University  
Adam Rothman, University of Minnesota  
Ming Yuan, University of Wisconsin-Madison  
Harrison Huibin Zhou, Yale University

## **Participants**

Tavis Abrahamsen, Jeongyoun Ahn, Mohamed Amezziane, Andreas Artemiou, Misheck Banda, Abhishek Bhattacharjee, Nikolay Bliznyuk, Andreas Buja, Florentina Bunea, Xuan Cao, Elvan Ceyhan, Saptarshi Chakraborty, Fisha Chen, Ming Chen, Zhe Chen, Heemin Choi, Dennis Cook, Hani Doss, Isaac Duerr, Tanya P. Garcia, Malay Ghosh, Satyajit Ghosh, Nicholas Henderson, Jim Hobert, Yongho Jeon, Yeun Ji Jung, Yicheng Kang, Moumita Karmakar, Priya Kohli, Myung Hee Fee, Bing Fi, Fexin Fi, Qiefeng Fi, Antonio Finero, Anna Fittle, Yanyuan Ma, Mehdi Maadooliat, Madhurima Majumder, Nonhle Channon Mdziniso, Hunter R. Merrill, Jiyoun Myung, Kijoeng Nam, Trang Nguyen, Omosigho Maryjane Osakue, Subhadip Pal, Cheolwoo Park, Yeonhee Park, Robert Parker, Cornells Potgieter, Brett Presneli, Syed Hafizur Rahman, Andrew Rosalsky, Mark Bennett Rosenberg, Adam Rothman, Zhihua Su, Aixin Tan, Xueying Tang, Jin Tao, Susan Wei, Xuerong Meggie Wen, Samuel Wong, Johnny Wu, Wenbo Wu, Xiao Wu, Yang Wu, Zhiming Xia, Ruoxuan Xiang, Xiaofan Xu, Zhigang Yao, Ming Yuan, Bilin Zeng, Fong Zhang, Shaojun Zhang, Xin Zhang, Shishun Zhao, Xiaolong Zhong, Harrison Huibin Zhou, Guangyu Zhu

## **Acknowledgements**

The organizers thank the staff of the Department of Statistics and Center for Management Communication, William Campbell, Robyn Crawford and Tina Greenly, for their help in organizing this meeting and making it run smoothly.

## **Thursday, January 16**

6:00-7:30 p.m. Reception

Al and Judy WarringtonRoom

# Friday, January 17

8:00-8:45 a.m.	Continental Breakfast	
8:45-9:00 a.m.	CONFERENCE WELCOME Ata Sarajedini, Associate Dean for Natural Sciences & Mathematics Zhihua Su, Organizing Committee	Rm209
9:00-10:40 a.m.	SESSION 1 Chair: Zhihua Su	Rm209
	Dennis Cook      Prediction in Abundant High-dimensional Linear Regression	
	Lexin Li            Tensor Regression, Regularization, and Neuroimaging Data Analysis	
10:40-11:10 a.m.	Break/Refreshments	
11:10 a.m.- 12:00 p.m.	SESSION 2 Chair: Hani Doss	Rm209
	Ming Yuan            Optimal Detection of Sparse Signal Segments	
12:00-12:30 p.m.	Conference Photo	Entrance Stairs
12:30-2:20 p.m.	Lunch (Gator Corner Dining Center)	
2:20M-:00 p.m.	SESSION 3 Chair: Malay Ghosh	Rm 209
	Florentina Bunea      Computationally Efficient Minimax Adaptive Estimation of Large Covariance Matrices with Low Dimensional Structures	
	Adam Rothman        Properties of Optimizations Used in Penalized Gaussian Likelihood Inverse Covariance Matrix Estimation	
4:00-5:30 p.m.	Poster Session/Refreshments	Presidents Rm B EAH

# Saturday, January 18

8:00-9:00 a.m. Continental Breakfast

9:00 a.m.-10:40 a.m. SESSION 4 Rm209  
Chair: Kshitij Khare

Bing Li On an Additive Semi-Graphoid Model for Statistical Networks with Application to Pathway Analysis

Yanyuan Ma On Estimation Efficiency in Dimension Reduction

10:40-11:10 a.m. Break/Refreshments

11:10 a.m.-12:50 p.m. SESSION 5 Rm209  
Chair: Andrew Rosalsky

Harrison Zhou Asymptotic Normality and Efficiency In Estimation of High-dimensional Graphical Models

Lisha Chen Ensemble Subsampling for Imbalanced Multivariate Two-Sample Tests

12:50-2:20 p.m. Lunch/Free time

2:20-3:10 p.m. SESSION 6 Rm209  
Chair: Brett Presneli

Andreas Buja Valid Post-Selection Inference

# Abstracts

## Valid Post-Selection Inference

Andreas Buja  
University of Pennsylvania

It is common practice in statistical data analysis to perform data-driven variable selection and derive statistical inference from the resulting model. Such inference enjoys none of the guarantees that classical statistical theory provides for tests and confidence intervals when the model has been chosen a priori. We propose to produce valid “post-selection inference” by reducing the problem to one of simultaneous inference. Simultaneity is required for all linear functions that arise as coefficient estimates in all submodels. By purchasing “simultaneity insurance” for all possible submodels, the resulting post-selection inference is rendered universally valid under all possible model selection procedures. Importantly the inference does not depend on the truth of the selected submodel, and hence it produces valid inference even in wrong models.

## Computationally Efficient Minimax Adaptive Estimation of Large Covariance Matrices with Low Dimensional Structures

Florentina Bunea  
Cornell University

We study estimation of two different classes of population covariance matrices with low dimensional structures: that of reduced effective rank matrices and the class of semi-banded matrices.

We show that the ubiquitously used sample covariance matrix is an accurate estimator, in Frobenius and operator norm, of  $r \times r$  population matrices of reduced effective rank, even if  $p$  exceeds the sample size  $n$ . Our results are based on sharp minimax bounds on these norms. Over classes of population matrices with effective rank of appropriate growth relative to  $n$ , we study the finite-sample properties of the sample eigenvalues and eigenvectors selected by scree-plot methods. We discuss the implications of these results to PCA and fPCA.

We also introduce a class of semi-banded covariance matrices, that generalizes both banded and approximately banded matrices. We introduce a new estimator, the hierarchically banded estimator, which is the solution of a computationally feasible convex optimization problem. We show that our estimator is minimax adaptive with respect to the Frobenius norm over the general class of semi-banded matrices, and with respect to the operator norm over the class of banded matrices. Extensive simulation studies show that our procedure is a strong competitor of all previously developed methods.



## **Ensemble Subsampling for Imbalanced Multivariate Two-Sample Tests**

Lisha Chen  
Yale University

In the past decade, imbalanced data have drawn increasing attention in the machine learning community. Such data commonly arise in many fields such as biomedical science, financial economics, fraud detection, marketing, and text mining. The imbalance refers to a large difference between the sample sizes of data from two underlying distributions or from two classes in the setting of classification. We tackle the challenges of imbalanced learning in the setting of the long-standing statistical problem of multivariate two-sample tests. Some existing nonparametric two-sample tests for equality of multivariate distributions perform unsatisfactorily when the two sample sizes are imbalanced. In particular, the power of these tests tends to diminish with increasingly imbalanced sample sizes. We propose a new testing procedure to solve this problem. The proposed test is based on a nearest neighbor method and employs a novel ensemble subsampling scheme to treat the imbalance of data. We demonstrate the strong power of the testing procedure by simulation study and real data example, and provide asymptotic analysis for our testing procedure.

## **Prediction in Abundant High-dimensional Linear Regression**

Dennis Cook  
University of Minnesota

An abundant regression is one in which most of the predictors contribute information about the response, which is contrary to the common notion of a sparse regression where few of the predictors are relevant. We discuss asymptotic characteristics of methodology for prediction in abundant linear regressions as the sample size and number of predictors increase in various alignments. We show that some of the estimators can perform well for the purpose of prediction in abundant regressions. This is joint work with Liliana Forzani and Adam Rothman.

## **On an Additive Semi-Graphoid Model for Statistical Networks with Application to Pathway Analysis**

Bing Li

Pennsylvania State University

We introduce a nonparametric method for estimating non-gaussian graphical models based on a new statistical relation called additive conditional independence, which is a three-way relation among random vectors that resembles the logical structure of conditional independence. Additive conditional independence allows us to use one-dimensional kernel regardless of the dimension of the graph, which not only avoids the curse of dimensionality but also simplifies computation. It also gives rise to a parallel structure to the gaussian graphical model that replaces the precision matrix by an additive precision operator. The estimators derived from additive conditional independence cover the recently introduced nonparanormal graphical model as a special case, but outperform it when the gaussian copula assumption is violated. We compare the new method with existing ones by simulations and in genetic pathway analysis. (Joint work with Hyonho Chun and Hongyu Zhao.)

## **Tensor Regression, Regularization, and Neuroimaging Data Analysis**

Lexin Li

North Carolina State University

Classical regression methods treat covariates as a vector and estimate a corresponding vector of regression coefficients. Modern applications in medical imaging generate covariates of more complex form such as multidimensional arrays (tensors). Traditional statistical and computational methods are compromised for analysis of those high-throughput data due to their ultrahigh dimensionality as well as complex structure. In this talk, I will discuss a new class of tensor regression models that efficiently exploit the special structure of tensor covariates. Under this framework, ultrahigh dimensionality is reduced to a manageable level, resulting in efficient estimation and prediction. Regularization, both hard thresholding and soft thresholding types, will be carefully examined. The new methods aim to address a family of neuroimaging problems, including using brain images to diagnose neurodegenerative disorders, to predict onset of neuropsychiatric diseases, and to identify disease relevant brain regions or activity patterns.

## **On Estimation Efficiency in Dimension Reduction**

Yanyuan Ma  
Texas A&M University

We investigate the estimation efficiency of the central (mean) subspace in the framework of sufficient dimension reduction. We derive the semiparametric efficient score and study its practical applicability. Despite the difficulty caused by the potential high dimension issue in the variance component, we show that efficient and/or locally efficient estimators can be constructed in practice. We conduct simulation studies and a real-data analysis to demonstrate the finite sample performance and efficiency gain of the proposed estimators in comparison with several existing methods. This is joint work with Liping Zhu.

## **Properties of Optimizations Used in Penalized Gaussian Likelihood Inverse Covariance Matrix Estimation**

Adam Rothman  
University of Minnesota

We establish necessary and sufficient conditions for the existence of inverse covariance matrix estimates obtained by minimizing the negative Normal log-likelihood plus a weighted ridge or weighted L<sub>1</sub> penalty. A new algorithm to solve this optimization with the weighted ridge penalty is developed and its convergence is established. This algorithm combines the majorize minimize principle with minorize minimize acceleration attempts. Numerical experiments show this algorithm is superior to its only competitor and that ridge penalization is useful within quadratic discriminant analysis.

## **Optimal Detection of Sparse Signal Segments**

Ming Yuan  
University of Wisconsin-Madison

In many applications involving high dimensional data, we are interested in identifying sparse segments of signal in a noisy background. The signals can be copy number variation, or objects in video surveillance among others. We study the relationship between the detectability of a signal, and its shape as well as duration; and the implications of such a relationship on detection strategies.

# **Asymptotic Normality and Efficiency In Estimation of High-dimensional Graphical Models**

Harrison Huibin Zhou  
Yale University

In this talk we will first introduce an asymptotically normal and efficient result for estimation of high-dimensional Gaussian graphical model under a sparseness assumption, which is shown to be not only sufficient, but also necessary, then present some preliminary analogous results for Ising model.

## Poster Abstracts

### Interpolation of Computationally Expensive Posterior Densities with Variable Parameter Costs

Nikolay Bliznyuk  
University of Florida

Markov Chain Monte Carlo (MCMC) is nowadays a standard approach to numerical computation of integrals of the posterior density  $\pi$  of the parameter vector  $\eta$ . Unfortunately, Bayesian inference using MCMC is computationally intractable when the posterior density  $\pi$  is expensive to evaluate. In many such problems, it is possible to identify a minimal subvector  $\beta$  of  $\eta$  responsible for the expensive computation in the evaluation of  $\pi$ . We propose two approaches, DOSKA and INDA, that approximate  $\pi$  by interpolation in ways that exploit this computational structure to mitigate the curse of dimensionality. DOSKA interpolates  $\pi$  directly while INDA interpolates  $\pi$  indirectly by interpolating functions, e.g., a regression function, upon which  $\pi$  depends. Our primary contribution is derivation of a Gaussian processes interpolant that provably improves over some of the existing approaches by reducing the effective dimension of the interpolation problem from  $\dim(\eta)$  to  $\dim(\beta)$ . This allows a dramatic reduction of the number of expensive evaluations necessary to construct an accurate approximation of  $\pi$  when  $\dim(\eta)$  is high but  $\dim(\beta)$  is low.

We illustrate the proposed approaches in a case study for a spatio-temporal linear model for air pollution data in the greater Boston area.

### Modeling Financial Volatility: An Exogenous Log-GARCH Approach

Ming Chen  
University of Texas at Dallas

In this article, we develop new methods for financial volatility estimation and forecast by including exogenous variables in a semiparametric log-GARCH model. With additional information, we expect an increased prediction power. We propose a quasi maximum likelihood procedure with Gaussian likelihood function via spline smoothing technique. Consistent estimation and asymptotic normality are obtained under mild regularity conditions. Simulation experiments provide strong evidence that corroborates the asymptotic theory. An application to S&P 500 index data demonstrates the advantage of our model over GARCH model and log-GARCH model.

## **Identification of Important Regressor Groups, Subgroups, and Individuals via Regularization Methods: Application to Gut Microbiome Data**

Tanya P. Garcia  
Texas A&M University

Gut microbiota can be classified at multiple taxonomy levels. Strategies to use changes in microbiota composition to effect health improvements require knowing at which taxonomy level interventions should be aimed. Identifying these important levels is difficult, however, because most statistical methods only consider when the microbiota are classified at one taxonomy level, not multiple. Using  $L_1$  and  $L_2$  regularizations, we developed a new variable selection method that identifies important features at multiple taxonomy levels. The regularization parameters are chosen by a new, data-adaptive, repeated cross-validation approach which performed well. In simulation studies, our method outperformed competing methods it more often selected significant variables and has small false discovery rates and acceptable false positive rates. Applying our method to gut microbiota data, we found which taxonomic levels were most altered by specific interventions or physiological status.

Joint work with Samuel Müller, Raymond J. Carroll and Rosemary L. Walzem.

## **Making the Cut: Ranking and Selection Procedures for Large-Scale Inference**

Nicholas Henderson  
University of Wisconsin - Madison

Identifying leading measurement units from a large collection is a common task in various domains of large-scale inference. Testing approaches, which measure evidence against a null hypothesis rather than effect magnitude, tend to overpopulate lists of leading units with those having low measurement error. By contrast, maximum likelihood approaches tend to favor units with high measurement error. In addition, many commonly used Bayesian and empirical Bayesian ranking procedure that populates the list of top units in a way that maximizes the expected overlap between the true and reported lists for all list sizes. Several examples from genomics are used to illustrate our approach.

## **Covariance Estimation for Multivariate Longitudinal Data**

Priya Kohli  
Connecticut College

Efficient modeling of covariance matrix is challenging mostly due to the positive-definiteness constraint and high-dimensionality of the parameter space where the number of parameters grows quadratically with the number of time points and outcomes in the multivariate longitudinal studies. Such studies are common in a variety of fields including business and economics, health care, epidemiology and environmental, physical and social sciences. We develop a data-based method which provides an unconstrained reparameterization for the covariance matrix with reduced dimensions. We propose iterative algorithms based on the score functions and Fisher information matrices to obtain the maximum likelihood estimators for the unconstrained covariance parameters. The performance of the estimators is illustrated through extensive simulation studies and an application to real data.

Joint work with Mohsen Pourahmadi and Tanya Garcia.

## **Meta-Analysis Based Variable Selection for Gene Expression Data**

Quefeng Li  
University of Wisconsin-Madison

Recent advance in biotechnology and its wide applications have led to the generation of many high-dimensional gene expression data sets that can be used to address similar biological questions. Meta-analysis plays an important role in summarizing and synthesizing scientific evidence from multiple studies. When the dimensions of datasets are high, it is desirable to incorporate variable selection into meta-analysis to improve model interpretation and prediction. According to our knowledge, all existing methods conduct variable selection with meta-data in an all-in-or-all-out fashion, i.e., a gene is either selected in all of studies or not selected in any study. However, due to data heterogeneity commonly exist in meta-data, including choices of biospecimens, study population, and measurement sensitivity, it is possible that a gene is important in some studies while unimportant in others. We propose a novel method called meta-lasso for gene and pathway selection with high-dimensional meta-data. Through a hierarchical decomposition on regression coefficients, our method not only borrows strength across multiple data sets to boost the power to identify important genes and pathways, but also keeps the selection flexibility among data sets to take into account data heterogeneity. We show that our method possesses the gene selection consistency, i.e., when sample size of each data set is large, with high probability, our method can identify all important genes and remove all unimportant genes. Simulation studies demonstrate that our method is superior than some popular methods. We applied our meta-lasso method to a meta-analysis of five cardiovascular studies. The analysis results are clinically meaningful.

Joint work with Sijian Wang, Chiang-Ching Huang, Menggang Yu and Jun Shao.

## **Assessing Protein Conformational Sampling Methods Based on Bivariate Lag-Distributions of Backbone Angles**

Mehdi Maadooliat  
Marquette University

Despite considerable progress in the past decades, protein structure prediction remains one of the major unsolved problems in computational biology. Angular-sampling-based methods have been extensively studied recently due to their ability to capture the continuous conformational space of protein structures. The literature has focused on using a variety of parametric models of the sequential dependencies between angle pairs along the protein chains.

In this work, we present a thorough review of angular-sampling-based methods by assessing three main questions: What is the best distribution type to model the protein angles?

What is a reasonable number of components in a mixture model that should be considered to accurately parameterize the joint distribution of the angles? and What is the order of the local sequence-structure dependency that should be considered by a prediction method?

We assess the model fits for different methods using bivariate lag-distributions of the dihedral/planar angles. Moreover, the main information across the lags can be extracted using a technique called Lag singular value decomposition (LagSVD), which considers the joint distribution of the dihedral/planar angles over different lags using a nonparametric approach and monitors the behavior of the lag-distribution of the angles using singular value decomposition. As a result, we developed graphical tools and numerical measurements to compare and evaluate the performance of different model fits. Furthermore, we developed a web-tool (<http://www.stat.tamu.edu/~madooliat/LagSVD>) that can be used to produce informative animations.

Joint work with Xin Gao and Jianhua Z. Huang

## **Dimension Reduction Using Inverse Spline Regression**

Kijoeng Nam  
University of Maryland - College Park

In high dimensional data analysis, we often want to reduce the number of predictors without eliminating variables which are related to the response of interest. Inverse regression methods use the response variable when performing dimension reduction so that information regarding the relation between covariates and the response is not lost. However, it is common to assume that the inverse regression function is linear or to use some other ad hoc approach. Instead, we propose a new dimension reduction methods which models the inverse regression function as a spline. We develop asymptotics for our approach and demonstrate its performance is better than existing inverse regression based methods, especially when the dimension reduction space is a nonlinear manifold such as Swill roll example of Roweis and Saul.



## **Latent Supervised Learning for Survival Data**

Susan Wei

University of North Carolina at Chapel Hill

Latent supervised learning is a machine learning technique for performing binary classification using a surrogate variable for the unobserved training label. We extend latent supervised learning to the case when the surrogate variable is a right-censored survival time. A motivating application for the proposed methodology is to stratify patients into two risk groups given a set of biomarkers. Sieve maximum likelihood estimation is employed for model estimation with special care taken to account for censoring. Consistency of the proposed estimator is established. Simulations show that the proposed estimator is accurate under a range of settings. Applications to real data examples demonstrate its advantages over a competing method; the proposed method produces more significant separation in survival on both training sets and held-out independent test sets.

## **Stable Estimation in Dimension Reduction**

Wenbo Wu

University of Georgia

Many dimension reduction methods provide estimation results with one or more tuning parameters involved. These results could be instable due to the sensitiveness to the tuning parameter values. We introduce stable estimation procedures in different aspects of dimension reduction. We first propose stable methods in estimating structural dimension which only selects the correct directions in central subspace with no false positive selection. We then propose a general Grassmann manifold estimation approach to give sparse estimation of basis directions of central subspace. For obtaining non-sparse estimation of basis directions of central subspace, we develop an ensemble methods based on subsampling. We found that similar ensemble approach could also be used to stabilize the choice of number of slices to use in the sliced inverse methods. Theoretical supports are established and effectiveness of proposed stable methods are demonstrated by real and simulated data.

Joint work with Xiangrong Yin.

## Principal Flows

Zhigang Yao

Swiss Federal Institute of Technology (EPFL)

We revisit the problem of extending the notion of principal component analysis (PCA) to multivariate data sets that satisfy non-linear constraints, therefore lying on Riemannian manifolds. Our aim is to determine curves on the manifold that retain their canonical interpretability as principal components, while at the same time being flexible enough to capture non-geodesic forms of variation. We introduce the concept of a principle flow, a curve on the manifold passing through the mean of the data, and with the property that, at any point of the curve, the tangent velocity vector attempts to fit the first eigenvector of a tangent space PCA locally at the same point, subject to move along a path of maximal variation of the data, up to smoothness constraints. The rigorous definition of a principal flow is given by means of a Lagrangian variation problem, and its solution is reduced to an ODE problem via the Euler-Lagrange method. Conditions for existence and uniqueness are provided, and algorithm is outlined for the numerical solution of the problem. Higher order principal flows are also defined. It is shown that principal flows can yield the usual principal components on a Euclidean space. By means of examples, it is illustrated that the principal flow is able to capture patterns of variation that can escape other manifold PCA methods.

Joint work with Victor M. Panaretos and Tung Pham.

## Fused Estimators of the Central Subspace in Sufficient Dimension Reduction

Xin Zhang

University of Minnesota

When studying the regression of a univariate variable  $Y$  on a vector  $X$  of predictors, most existing sufficient dimension reduction (SDR) methods require the construction of slices of  $Y$  in order to estimate moments of the conditional distribution of  $X$  given  $Y$ . But there is no widely accepted method for choosing the number of slices, while a poorly chosen slicing scheme may produce miserable results. We propose a novel and easily implemented fusing method that can mitigate the problem of choosing a slicing scheme and improve estimation efficiency at the same time. We develop two fused estimators called FIRE and DIRE based on an optimal inverse regression estimator. The asymptotic variance of FIRE is no larger than that of the original methods regardless of the choice of slicing scheme, while DIRE is less computationally intense and more robust. Simulation studies show that the fused estimators perform effectively the same as or substantially better than the parent methods. Fused estimators based on other methods can be developed in parallel: fused SIR, fused CSS-SIR and fused LAD are introduced briefly. Simulation studies of the fused CSS-SIR and fused LAD estimators show substantial gain over their parent methods. A real data example is also presented for illustration and comparison.

## Gainesville Restaurants

\*: Indicates that the restaurant is within walking distance from Emerson Alumni Hall

Name	Address	Phone
Applebee's Neighborhood Grill & Bar	1005 NW 13th St	(352)335-0150
Ballyhoo Grill	3700 Newberry Road	(352)373-0059
Bistro 1245	1245 W. University Ave*	(352)376-0000
BJs Restaurant & Brewhouse	6611 Newberry Road (Oaks Mall)	(352) 331-8070
Boca Fiesta	232 SE 1st Street	(352)336-8226
Bonefish Grill	3237 SW 35th Blvd (Archer Rd)	(352)377-8386
Burrito Brothers Taco Co.	16 NW 13th St*	(352) 378-5948
Carrabas Italian Grill	3021 SW 34th St	(352) 692-0083
Chili's Grill & Bar	3530 SW Archer Road	(352)373-3010
Chipotle Mexican	1432 W. University Ave*	(352)372-5330
Chopstix Café	3500 SW 13th St	(352)367-0003
Civilization	1511 NW 2nd St	(352) 380-0544
Copper Monkey Restaurant	1700 W University Ave*	(352) 374-4984
David's Real Pit BBQ	5121 NW 39th Ave	(352) 373-2002
Dragonfly Sushi & Sake Company Inc	201 SE 2nd Ave	(352)371-3359
Emiliano's Café	7 SE 1st Ave	(352)375-7381
Farah's on the Avenue	1120 W University Ave	(352) 378-5179
Francescas Trattoria	4410 NW 25th Place	(352) 378-7152
TGI Friday's	3598 SW Archer Road	(352)336-0033
Fuji Hana	3720 NW 13th Street	(352)337-0038
Gators Dockside	3842 Newberry Road	(352) 338-4445
Harry's Seafood Bar & Grille	110 SE 1st Street	(352) 372-1555
Ivey's Grill	3303 W. University Ave	(352)371-4839
Jimmy John's	2220 SW Archer Rd	(352) 271-7600
Jimmy John's	1724 W Univ. Ave*	(352)375-7222
Farry's Giant Subs	1122 N Main St*	(352)376-1210
Farry's Giant Subs	1620 W Univ. Ave*	(352)271-7977
Fas Margaritas Mexican Restaurant	4401 NW 25th Place	(352)374-6699
Feonardo's 706	706 W. University Ave	(352)378-2001
Feonardos Pizza by the Slice	1245 W. University Ave*	(352)375-2007
Fiquid Ginger	101 SE 2nd Place	(352)371-2323
Romanos Macaroni Grill	6401 Newberry Rd. Oaks Mall	(352) 331-8070

*Continued on next page*

<b>Name</b>	<b>Address</b>	<b>Phone</b>
Manuels Vintage Room	6 S. Main Street	(352) 375-7372
Maude's Classic Cafe	101 SE 2nd Place, Suite 101	(352)336-9646
Mildred's Big City Food	3445 W University Ave	(352)371-1711
Napolatano's	606 NW 75th St	(352) 332-6671
New Deal Café	3445 W. University Ave	(352)371-4418
Olive Garden	3440 SW Archer Road	(352)335-5354
Paramount Grill	12 SW 1st Ave	(352) 378-3398
Pita Pit	1702 W. University Ave, Gator Plaza*	(352) 692-4400
Red Lobster	6910 W Newberry Road	(352) 331-2670
Ruby Tuesday Restaurant	Oaks Mall	(352)331-0033
Sonny's Real Pit Bar-B-Q	2700 NE Waldo Rd	(352) 378-5161
Sonny's Real Pit Bar-B-Q	9213 NW 39th Ave	(352) 381-7333
Sonny's Real Pit Bar-B-Q	3635 SW Archer Rd	(352)375-6667
Stonewood Grill & Tavern	3812 Newberry Road	(352) 379-5982
Swamp Restaurant, The	1642 W University Ave*	(352) 377-9267
Tatu	1702 W. University Ave, Gator Plaza*	371-1700
Texas Roadhouse	3830 SW Archer Road	(352)377-2820
The Top	30 N. Main Street	(352)337-1188
Tijuana Flats	1720 University Ave*	(352) 692-3093
Tony and Pat's Pizza & Subs	3501 SW Archer Road	(352) 377-7400
Virtually Cuban	2409 SW 13th Street	336-4125
Warehouse Restaurant & Lounge	502 S Main Street	(352) 240-6432
Yamato Japanese Restaurant	526 NW 60th St	(352) 332-4466