Yannick Baraud
University of Nice

“ρ-estimation”

Our purpose is to present a new estimation procedure that can be used for estimating a density or a regression function (when the design points are either fixed or random with an unknown distribution) and which leads to both robust and rate optimal estimators (up to a possible logarithmic factor) in all cases we know.

In the simple linear regression setting, the estimator we get outperforms the usual least squares from many aspects. The procedure is robust to outliers and allows to estimate the coefficients at a rate which can be much faster than the usual parametric one for some distribution of errors.

In density estimation, we show that our procedure allows to recover (at least, when the number of observations is large enough) the celebrated maximum likelihood estimator when the model is regular enough and contains the true density. When the latter condition is not satisfied, we show that our procedure is robust (with respect to the Hellinger distance) while the maximum likelihood estimator is not in general.

Finally, we shall present an illustration of the performance of this new estimator on a simulation study.

(From a joint paper with Lucien Birgé and Mathieu Sart)

Jacob Bien
Cornell University

“Convex Banding of the Covariance Matrix”

We introduce a sparse and positive definite estimator of the covariance matrix designed for high-dimensional situations in which the variables have a known ordering. Our estimator is the solution to a convex optimization problem that involves a hierarchical group lasso penalty. We show how it can be efficiently computed, compare it to other methods such as tapering by a fixed matrix, and develop several theoretical results that demonstrate its strong statistical properties. Finally, we give
results on the minimum required signal strength for detection, giving insight into the effect of using a hierarchical (rather than non-overlapping) group structure in the penalty.

**Tony Cai**  
The Wharton School, University of Pennsylvania

“Recovery of High-Dimensional Low-Rank Matrices”

Low-rank structure commonly arises in many applications including genomics, signal processing, and portfolio allocation. It is also used in many statistical inference methodologies such as principal component analysis. In this talk, I will discuss some recent results on low-rank matrix estimation problems, including structured matrix completion and recovery of a low-rank matrix with rank-one measurements. Time permitting, I will also discuss optimal estimation of a spiked covariance matrix which can be viewed as a low-rank perturbation to an identity matrix.

**Venkat Chandrasekaran**  
Caltech University

“Computational and Statistical Tradeoffs via Convex Relaxation”

Modern massive datasets create a fundamental problem at the intersection of the computational and statistical sciences: how to provide guarantees on the quality of statistical inference given bounds on computational resources such as time or space. Our approach to this problem is to define a notion of “algorithmic weakening,” in which a hierarchy of algorithms is ordered by both computational efficiency and statistical efficiency, allowing the growing strength of the data at scale to be traded off against the need for sophisticated processing. We illustrate this approach in statistical denoising and in change-point detection in time series, using convex relaxation as the core inferential tool. Hierarchies of convex relaxations have been widely used in theoretical computer science to yield tractable approximation algorithms to many computationally intractable tasks. In this talk we show how to endow such hierarchies with a statistical characterization and thereby obtain concrete tradeoffs relating algorithmic runtime to amount of data.  
[Joint work with Michael Jordan and Yong-Sheng Soh]

**Christophe Giraud**  
Université Paris-Sud

“On Estimator Selection”

Accurate estimation is possible in high-dimensional settings by adapting to the (unknown) low dimensional structures of the data. Various powerful estimators have been proposed in the literature in order to adapt to various types of structures. Yet, these estimators require to know the type of structures hidden in the data and they need to be optimally tuned according to some unknown quantities (like the variance of the noise). It is then necessary to compute different estimators with different tuning parameters and then to apply an estimator selection scheme. Cross-validation technics are the more popular tools for estimator selection, yet they offer little guaranties in high-dimensional settings. We will review some alternative selection criteria which offer guaranties on the risk of the selected estimator.
Christopher Holmes
Oxford University

“Computational Decision Theory and Bayesian Methods for Exploring Sparse Structural Aberrations in Cancer Genomes”

We discuss computational approaches for exploring sparse structural changes (“copy-number aberrations”) in cancer genomes using high-throughput genotyping and sequencing data. We adopt a Bayesian mixture Hidden-Markov model to account for the state (copy-number) dependence along the genome and tumour heterogeneity. An important ex post question after fitting the model is to characterize the presence of sparse highly probable copy-number events. We show how this can be treated within a decision theoretic framework and develop computationally efficient algorithms that can deliver exact solutions. Importantly, these approaches de-couple the problem of fitting the model to data and that of exploring and representing uncertainty in highly probable (significant) sparse structures.

Han Liu
Princeton University

“Transelliptical Modeling and its Applications”

We introduce a new family of robust semiparametric methods for analyzing large, complex, and noisy datasets. Our method is based on the transelliptical distribution family which assumes that the variables follow an elliptical distribution after a set of unknown marginal transformations. The transelliptical distribution family includes the nonparanormal and elliptical families as special cases and can be used to robustify a broad family of multivariate methods, including covariance estimation, principal component analysis, graphical models, discriminant analysis, regression analysis, and principal component regression. We present a hierarchical representation of the transelliptical distribution and propose a new estimation technique that combines the ideas of marginal ranks and robust M-estimator. We also develop a unified theoretical framework which shows that, even though the transelliptical family is significantly larger than the Gaussian family, the obtained estimator attains the same statistical and computational rates of convergence as the Gaussian based estimator in most applications.

This is based on joint work with Fang Han and Cun-hui Zhang.

George Michailidis
University of Michigan

“Change Point Inference in Dynamic Erdos-Renyi Random Graphs”

We investigate a model of an Erdos-Renyi graph, where the edges can be in a set of finite states (e.g. present/absent). The states of each edge evolve as a Markov chain independently of the other edges, and whose parameters exhibit a change-point behavior in time. We derive the maximum likelihood estimator for the change-point and characterize its distribution. Depending on a measure of the signal-to-noise ratio present in the data, different limiting regimes emerge. Nevertheless, a unifying adaptive scheme can be used in practice that covers all cases. Finally, for appropriate choices of the parameters of the Markov kernels, the limiting distribution of the change-point
exhibits long-range dependence. The model is illustrated on synthetic, as well as US House roll call data.

Andrew Nobel
University of North Carolina

“Hypothesis Testing and Community Detection”

Community detection is a central and well studied problem in the field of network analysis, with applications to the analysis of social, political, and biological networks. The goal of community detection is to divide the vertices of an observed network into groups (usually disjoint), called communities, such that the edge density within groups is substantially higher than the edge density between groups. In this talk I will describe an extraction based procedure for community detection that is based on ideas from multiple testing, in particular, p-values derived from a conditional configuration model. The procedure is able to handle overlapping communities, and to identify background vertices that do not belong to any community. Comparisons to several existing community detection procedures on real and simulated data will be presented.

Sofia C. Olhede
University College London

“Nonparametric Graphon Estimation”

We propose a nonparametric framework for the analysis of networks, based on a natural limit object termed a graphon. We prove consistency of graphon estimation under general conditions, giving rates which include the important practical setting of sparse networks. Our results cover dense and sparse stochastic blockmodels with a growing number of classes, under model misspecification. We use profile likelihood methods, and connect our results to approximation theory, nonparametric function estimation, and the theory of graph limits. This is joint work with Patrick Wolfe

Xiaotong Shen
University of Minnesota

“Ordinal Classification with Unstructured Predictors”

Unstructured data refers to information that lacks certain structures and cannot be organized in a predefined fashion, which involve heavily on words, graphs, objects or multimedia types of files. In this presentation, I will focus on classification for unstructured word predictors with ordered class categories, where imprecise information concerning strengths between predictors is available for predicting the class labels. However, the imprecise information is expressed in terms of a directed graph, with each node representing a predictor and directed edge containing pairwise strengths between two nodes. One of targeted applications for unstructured data arises from sentiment analysis, which identifies and extracts the relevant content or opinion of a document towards a specific event of interest. Large margin ordinal classifiers will be introduced, which integrate the imprecise predictor relations into linear relational constraints over classification function coefficients subject to quadratically many linear constraints. An application for sentiment analysis will be discussed. This work is joint with J. Wang (UIC), P. Qu (UIUC) and Y. Sun (UMN)
Alexandre Tsybakov
CREST-ENSAE

“Linear and Conic Programming Approaches to High-Dimensional Errors-in-variables Models”

We consider the regression model with observation error in the design when the dimension can be much larger than the sample size and the true parameter is sparse. We propose two new estimators, based on linear and conic programming, and we prove that they satisfy oracle inequalities similar to those for the model with exactly known covariates. The only difference is that they contain additional scaling with the $l_1$ or $l_2$ norm of the true parameter. The scaling with the $l_2$ norm is minimax optimal and it is achieved on conic programming. This is a joint work with Mathieu Rosenbaum.

Luo Xiao
Johns Hopkins University

“Estimation of Covariance Matrices with Particular Structures”

We consider estimation of covariance matrices with either low-dimensional or a banded structure. For covariance matrix with a low-dimensional structure, we show that the sample covariance matrix estimator is consistent and derive its near optimal convergence rate with respect to the operator and Frobenius norms. For covariance matrix with a banding structure, we show that Bickel and Levina's banding estimator can be operator-norm rate and we propose a Sure (Stein's unbiased risk estimation)-type method for selecting the bandwidth.

Cun-Hui Zhang
Rutgers University

“Graphlet Screening in High Dimensional Variable Selection”

Consider a linear model $Y = X\_\_ + \_\_z$, where $X$ has $n$ rows and $p$ columns and $z \_ N(0; \_\_I_n)$. We assume both $p$ and $n$ are large, including the case of $p = n$. The unknown signal vector $\_\_z$ is assumed to be sparse in the sense that only a small fraction of its components is nonzero. The goal is to identify such nonzero coordinates (i.e., variable selection).

We are primarily interested in the regime where signals are both rare and weak so that successful variable selection is challenging but is still possible. We assume the Gram matrix $G = X0X$ is sparse in the sense that each row has relatively few large entries (diagonals of $G$ are normalized to 1). The sparsity of $G$ naturally induces the sparsity of the so-called Graph of Strong Dependence (GOSD). The key insight is that there is an interesting interplay between the signal sparsity and graph sparsity: in a broad context, the signals decompose into many small-size components of GOSD that are disconnected to each other.

We propose Graphlet Screening for variable selection. This is a two-step Screen and Clean procedure, where in the first step, we screen subgraphs of GOSD with sequential $l_2$-tests, and in the second step, we clean with penalized MLE. The main methodological innovation is to use GOSD to guide both the screening and cleaning processes.
For any variable selection procedure $\hat{\beta}$, we measure its performance by the Hamming distance between the sign vectors of $\hat{\beta}$ and $\beta$, and assess the optimality by the minimax Hamming distance. Compared with more stringent criterions such as exact support recovery or oracle property, which demand strong signals, the Hamming distance criterion is more appropriate for weak signals since it naturally allows a small fraction of errors.

We show that in a broad class of situations, Graphlet Screening achieves the optimal rate of convergence in terms of the Hamming distance. Unlike Graphlet Screening, well-known procedures such as the L0=L1-penalization methods do not utilize local graphic structure for variable selection, so they generally do not achieve the optimal rate of convergence, even in very simple settings and even if the tuning parameters are ideally set.

This talk is based on joint work with Jiashun Jin and Qi Zhang

**Helen Zhang**
University of Arizona

“Selection of Interaction Effects for Ultra High-Dimensional Data”

For ultra-high dimensional data, the identification of interaction effects is an extremely challenging task. The main difficulties lie in both computational and theoretical aspects. We propose a new framework and computational algorithms for interaction selection. Both forward selection and penalization methods are covered. The new methods are feasible even when the data dimension increases exponentially fast with the sample size. The proposed methods are featured with fast computation, rigorous theoretical support, and competitive numerical performances.

**Harrison Zhou**
Yale University

“Rate-Optimal Posterior Contraction for Sparse PCA”

Principal component analysis (PCA) is possibly one of the most widely used statistical tools to recover a low rank structure of the data. In the high-dimensional settings, the leading eigenvector of the sample covariance can be nearly orthogonal to the true eigenvector. A sparse structure is then commonly assumed along with a low rank structure. Recently, minimax estimation rates of sparse PCA were established under various interesting settings. On the other side, Bayesian methods are becoming more and more popular in high dimensional estimation. But there is little work to connect frequentist properties and Bayesian methodologies for high dimensional data analysis. In this talk, we propose a prior for the sparse PCA problem, and analyze its theoretical properties. The prior adapts to both sparsity and rank. The posterior distribution is shown to contract to the truth at optimal minimax rates. In addition, a computational efficient strategy for the rank-one case is discussed.