

# RANDOM MATRICES AND HIGHER DIMENSIONAL INFERENCE

organized by

Peter Bickel, Christopher Jones, Helene Massam, and Don Richards

## Workshop Summary

### Background

This workshop was sponsored jointly by AIM and SAMSI (Statistical and Applied Mathematical Sciences Institute, in the Research Triangle Park, NC). The objective of the meeting was to set out an agenda for future research in the area of large random matrices that is relevant to issues of high-dimensional statistical inference. The workshop was based on the SAMSI program of the same name held in Fall, 2006. It served as a transition of that program from the intensive research activities while many researchers were present at SAMSI. In this way, the workshop started by summarizing and assessing the SAMSI activities and then moved to synthesizing the key findings and outlining directions and problems for ongoing research.

The workshop was held over four days (rather than the typical five-day format for AIM). Given the background of the SAMSI program, this worked well as most of the participants already knew each other and had ideas about the direction of the workshop.

The SAMSI program on Random Matrices and High-Dimensional Inference brought together, during the fall of 2006, representatives of various research communities working in this area. The range of research topics varied from the study of the asymptotic behavior of the eigenvalues of large random matrices, an area of theoretical probability, to model selection for complex large high-dimensional data sets, an area of applied statistics. Though seemingly different, a number of these groups discovered that they had many common research themes.

The AIM Workshop has given the opportunity to these groups to meet again physically and thus truly cooperate and influence each other. One major achievement has been the communication between the “Regularization Group” represented by Bickel, Paul and Levina and the “Bayesian Methods/Graphical Model Group” represented by Rajaratnam and Massam. These two communities have traditionally evolved separately, identifying sparsity in seemingly different ways but there promises now to be a steady flow of ideas.

The workshop also brought together the core of the people who were stationed at SAMSI with those who were participating only through webex, such as Levina and Schartzmann, those who had attended only some workshops during the fall, such as Mingo and Wermuth, some who only came to the Opening Workshop at SAMSI such as Nadler and even some who were not affiliated with SAMSI at all such as Mylene Maida.

The AIM workshop was able to cast a wider net and its timing was perfect. The months of January to April allowed us to reflect on new concepts or concepts borrowed from the other working groups, to further the work started in the fall and thus be ready for new and deeper cooperation.

## Activities during week

Tuesday and Wednesday morning were devoted to presentations of work carried out during the SAMSI program. On Wednesday afternoon, there was a group discussion where we identified some of the issues that arose from the presentations:

- (1) Choice of a “right” metric on the cone of positive definite matrices;
- (2) Multivariate integrals that crop up in statistical problems;
- (3) Eigenvalues of sparse matrices;
- (4) Notions of sparsity;
- (5) Relationship of graphical models and factor models.

Following this discussion, the participants broke into four working groups concerned with graphical models, geometry, approximation and series expansion for integrals and wireless communications, respectively.

On Thursday morning, there were three short unplanned presentations from participants who had something they wanted to communicate to the others. Nanny Wermuth spoke on matrix representations and independencies in directed acyclic graphs. Jamie Mingo gave his view on the computation of moments of random matrices with combinatorial arguments in the framework of free probabilities. Boaz Nadler spoke on eigenvalue estimation and stochastic perturbation theory. These three speakers were all participants in the Opening SAMSI workshop who had not been able to participate in the working groups and they all gave talks that presented a different, original point of view on classical problems in the area of large random matrices. In the afternoon, an attempt was made to impose four or five working groups as was done the day before but there was “spontaneous disobedience”. The active participants had evolved beyond the common discussion stage. They now wanted to work in small groups of two or three people on precise topics and that is exactly what happened.

On Friday morning, Mylene Maida and Eitan Greenshtein gave short presentations on some of their current problems and these were followed by many questions.

## Research working groups

*Group I Graphical models* had a detailed discussion on the relationship between graphical models and spiked models. Nouredine El-Karoui and Debashis Paul joined the group though they had initially chosen another group and were very helpful in focusing the discussion and shared their technical expertise. This level of cooperation could only have happened at that time and not earlier; it happened because of the work done during the fall and because of the level of exposure to each other’s problems that had been had during the first thirty-six hours of the AIM workshop. In a workshop such as this one, presentations play an important part since there are the medium for communication and communication is essential in an area that deals with complex real problems studied by several research communities. It should also be emphasized that the breakaway group discussions that were held were crucial because they allowed a greater depth, understanding of different participants’ take on the problems at hand and getting down to detailed solutions. It is clear that new ideas were born during these discussions.

*Group II Multivariate distributions* New directions charted included the following topics:

- (1) Explicit bounds on rates of convergence in deconvolution density estimation for diffusion tensor imaging, where it is desired to develop non-parametric estimators of the density function of a population of positive definite (symmetric) random matrices.
- (2) Central Limit Theorems and Gaussian distributions on the space of positive definite matrices.
- (3) Distributions of eigenvalues of random positive definite matrices when the matrices are in special position, e.g., when two eigenvalues are coincident.

New ideas raised included:

- Derivation of explicit descriptions of probability distributions arising in Central Limit Theorems on the space of positive definite (symmetric) matrices.
- Applications of free probability to the development of explicit series expansions for the noncentral Wishart distributions.
- Applications of the theory of singular perturbations to study systems of partial differential equations for hypergeometric functions of matrix argument, particularly to cases in which the matrix argument is in special position.

*Group III Regularization* The notion of sparsity of matrices came up repeatedly in terms of:

- (1) A small number of sparse eigenvector corresponding to a small number of large eigenvalues well separated from the rest;
- (2) Matrices that can be well approximated up to a permutation by banded matrices;
- (3) Matrices whose inverses can be well approximated similarly.

A conjecture that 2) and 3) are, in some sense, equivalent was confirmed by Marko Lindner who pointed to a book he authored, published by Springer in 2006, which showed that the closure of the class of bounded operators was closed under taking inverses. Unfortunately, the argument is entirely non-constructive and he indicated interest in collaboration to obtain bounds. This is viewed as an important problem in his domain of functional analysis as well. The problem of independent component analysis for large sparse matrices was raised as important and completely open. Methods of estimation of the so-called mixing matrix have been developed by engineers and statisticians and properties established for fixed  $p$  and large  $n$ . These methods are highly nonlinear and their large  $p$  analysis presents a significant challenge. The problem of analysis of methods for choosing the regularization parameter, which is crucial in practice, was also singled out. Potential collaboration on these points includes at least Boaz Nadler, Peter Bickel, Nouredine el Karoui, and Debashis Paul.

### **Some collaborations that started during the workshop**

As mentioned above, numerous collaborations were initiated during the workshop.

- Rajaratnam and Nadler on eigenvalues estimation, Rajaratnam and Levina on  $L_1$  Lasso penalization for covariance selection models.
- Paul and with Raj Rao: on the use of random matrices in signal processing.
- Massam and Wermuth on factor analysis and graphical models, Massam and Schartzmann on the analysis of data such that data points are positive definite matrices
- During the conference, there arose the potential for collaboration with numerous attendees, including: work of Richards with Schwartzman on Central Limit Theorems on heat equations on the space of positive definite matrices, Massam with Nadler on

the topic of singular perturbations, and Massam with Mingo on applications of free probability to the non-central Wishart distributions.

The AIM workshop also provided the opportunity for discussions between Massam and Mingo on Wishart moment computations, a topic that both have tackled but from different points of view. Mingo is going to spend the Fall of 2007 at the Fields Institute in Toronto and they will carry on teaching each other their respective techniques and points of view.

## **Conclusion**

In a practical way, the impact of this workshop will be measured by the number of papers that will acknowledge the help of the American Institute of Mathematics. In a less tangible way, it will be measured by the connections that it helped establish or consolidate between the various communities working on notions of sparsity, covariance estimation, random matrices moment computations, eigenvalues behavior, and integral approximations.

We all viewed this meeting as highly successful and would encourage AIM to consider holding a follow-on meeting in a year. The extent of collaborations stemming from this workshop will be evident at that time, and the best way to measure the impact of this meeting is to have people report on their progress on the papers they will have written alone or jointly.