# The Heat is On

## The Second International Workshop on Mathematical Methods in Computer Science

## IISc, Bangalore, India
## July 24–August 3, 2000

In recent years we have seen a remarkable qualitative change in the techniques used in the field of Theoretical Computer Science – they have seen the introduction of deep and sophisticated concepts and techniques from many different areas of mathematics – probability theory, harmonic analysis, algebra, algebraic geometry and algebraic topology, to name a few – with very dramatic results: many problems that remained open for decades were conclusively resolved.

The primary goal of the workshop was to study and understand some of the most outstanding of these methods. We also saw some examples of problems which were resolved only after the introduction of some powerful mathematical techniques. Our report is therefore organised around the principal mathematical techniques that were the focus of the two weeks of intensive activity – four two hour sessions everyday with only lunch and some relaxing evening walks (morning jogs for some more enthu ones) in between.

Actually one is being unjust to one's hosts. Keenly aware of the soporific state that the pleasant climate and environs induce in visitors to the IISC campus, they had taken elaborate measures, it seems, for non-stop live entertainment in the form of a real life soap opera. For, participants awakened one morning to the news that Veerappan the notorious bandit (who uses only one name, as CNN informed us) had made off with the darling of the Kannada masses, cine-actor Rajkumar. Reports trickling in from the outside world spoke of mayhem and carnage on the streets as an irate mob brought the high tech silicon capital of India to a standstill. For a while the scene seemed all too grim at the thought that perhaps the quarantine would apply even to incoming food stuff. Fortunately this was dispelled with the smell of fresh idli/dosa in the morning breakfast. Although a participant was lost (the intrepid Ravi who even then entertained some thoughts of talking through the Kannada mob between him in Salem and the workshop), the organisers were kind enough to ensure that the heat was off after the workshop on all fronts and we could all leave on time.

Something else was also lost in this drama. As a week of the workshop came to a close, many of us cloistered on campus felt the need for strong stimulants. And Bangalore is where India's first yuppie (as well as non-yuppie) pubs had sprung up over a decade ago, so there was no pressing need for abstinence, right? Well, there almost was! On the first hand, one has to contend with the delightful woolliness of info — most of our hosts were teetotallers (correction: they wouldn't even drink tea or coffee! Not that they objected to caffeine in a cold cola), so we had to garner info from "external sources". Which happened to be a physicist friend whom we knew to be a brilliant genius, but we should also have known him to be extremely unreliable. So Sat (29 Jul) evening had this delightful scene of two autorickshaws, filled with thirsty souls, tearing down the city roads in search of a highly recommended pub. A fruitless 60 minutes later, a call to another friend (not a physicist! but one who studies primates – and, I guess, us!!) revealed that we needed to go back to where we began, and the place was found. BUT! The pub had been de-licensed! Then began another trip to yet another place - this time we gave up the search for a pub and settled for a restaurant which also served beer. Two solid hours had been successfully diverted from beer-drinking to inane gallivanting. And, this was virtually the last beer we got: Sunday night the brigand swung into action, and all of the following week Bangalore went dry to "avoid law and order problems". As if this was what caused the problem. Oh well, if the workshop saw some clouded minds, at least it wasn't Bangalore's pub culture to blame!

## Sorting

The workshop got off to a flying start with a talk by Devdatt on sorting under partial information. Given a partial order, the information-theoretic sorting bound says that the number of comparisons needed is at least logarithmic in the number of linear extension. Can this bound be achieved for an arbitrary partially ordered set? A simple counter-example shows that the constant factor must be *greater* than 1, unlike in sorting from scratch. How much greater? One approach to bounding this has been to search for a balanced pair in the poset: an incomparable pair which when ordered halves or almost halves the number of linear extensions. The search for a balanced pair has thrown up interesting conjectures and likely counterexamples, but there's miles to go. We were taken through the details of such an analysis for width-2 posets, and also shown something rather counter-intuitive, namely that it actually pays to be lopsided rather than search for an almost balanced pair. The search for balanced pairs was also given a geometrical perspective, and once we were down to mixed volumes, an inequality due to Alexandrov and Finchel popped up and did the deed. A good overview of this approach with several proofs fully described or at least sketched is provided by a recent survey of Brightwell [8], see also [16].

Later in the workshop, we returned to the theme of sorting under partial information, when graph entropy played a key role in discovering the optimal (not necessarily balanced) pair to compare at each stage. An overview follows in this report, eventually — read on!

## Harmonic Analysis

Harmonic or Fourier analysis is an indispensable tool in many branches of Electrical Engineering, Mathematics and Physics. By passing from a function defined on one space to its Fourier transform in another one gains valuable information. To take the most visible example, we are reminded of the spectrum of sound or light frequencies in an audio or video signal.

In recent years, these methods have made breakthrough contributions in Combinatorics and Computer Science. In a recent commemoratory volume is an essay by Jean Bourgain [2] on "What can Combinatorics and Harmonic Analysis contribute to each other?". Among other achievements, he mentions the brilliant work of Fields Medallist Timothy Gowers in obtaining a quantitative solution to the Erdös -Turan problem of arithmetic progressions (to which Endre Szemeredi had given a combinatorial solution before). Another area where these methods are playing an increasing role is in percolation theory (much to the annoyance of some people whom I heard complaining that they lose all the intuition when it goes over into another domain!). Finally mention must be made of the surprising role they have played in the theory of lower bounds in approximation algorithms in the work of near optimal inapproximability results of Johan Hastad. The reason for the "unreasonable effectiveness" of Fourier methods in combinatorics, is perhaps that in the words of Gowers, it does not matter that we perturb the coefficients of the characteristic function of a set and hence can concentrate on the higher coefficients, a remark whose mystery was partially dispelled in this series of lectures. A good reference for this series is the course notes developed independently by Babai and Umesh Vazirani (check their home pages/ ask them for a copy).

Jaikumar led us through a charming introduction to the world of Fourier transforms, onto some interesting applications. The first application was about resilience of Boolean functions: given a function $f$ from $n$ to $m$ bits with outputs uniformly distributed over the range, we say that $f$ is $t$-resilient if this uniform distribution is preserved even after fixing any $t \leq n - m$ inputs. For instance, the XOR of $n$ bits is as resilient as they get, namely, $(n-1)$-resilient. How resilient can functions be at all? For functions from $3k$ to 2 bits, a simple construction using Fourier coefficients shows that one can have at most $2k-1$ resilience.

The second application was to do with influence of individual variables on the overall function, and sensitivity, another application mentioned by Bourgain's essay. It is easy to construct functions where the average influence of variables is low. With some more work, one can achieve low maximum influence. With clever use of the mysterious Bonami-Becker Inequality (however does one know when to pull out which such inequality? One was pulled out in the first talk while searching for balanced pairs, now another one here. We really need a nice oracle for this!), Kahn, Kim and Linial showed that if the Fourier transform of a balanced function is concentrated on large sets, then the average sensitivity of the function is high [15].

In a later talk, Meena described a proof due to Linial, Mansour and Nisan [13] that $AC^0$ functions have Fourier transforms concentrated on low sets, and thus have low average sensitivity, have good low-degree polynomial approximations over reals, are efficiently PAC-learnable, and are not good pseudorandom function generators.

Yet another very neat application Jaikumar described was using harmonic analysis to provide an upper bound on $N(H, \ell)$, the maximum number of copies of graph $H$ to be found in any graph with $\ell$ edges. Friedgut and Kahn have another proof using entropy. The harmonic analysis proof seemed much more simple and elegant to me, but it does not work for hypergraphs. Can't have it all ...

While on harmonic analysis, Pranab gave some explicit constructions for expanders. The constructions are very simple, but the proof of expansion properties needs bounds on graph eigenvalues, and the bounds are obtained by getting into matrix algebra and analyzing related Fourier transforms.

Pranab also talked about computing the quantum Fourier transform with polynomially many quantum gates. A crash intro to quantum computing, and a sketch of Shor's algorithm over cyclic groups, was followed by Kitaev's extension to Abelian groups.

## Group-theoretic techniques

Group theory crops us simply everywhere in theory of computing — a seminar on "Group-theoretic techniques in computer science" would run on into weeks. All we could hope for was a good sample. We got a wonderful one.

Arvind showed us how group theory was

used effectively to solve what looks like a purely graph-theoretic problem. The problem is to decide if two given graphs, of bounded degree (in fact most of the exposition was for graphs with max degree 3), are isomorphic. Without the degree restriction, graph isomorphism is the celebrated and well-studied "intermediate" problem — unlikely to be NP-complete, not known to be in P, but no known lower bounds better than NLOG. For bounded degree graphs, there is however a polynomial time algorithm.

The intuition behind how this algorithm works on trivalent graphs is roughly as follows: Consider the graph $G$ which is disjoint union of $G_1$ and $G_2$, and look at its automorphism group $A$ (elements are vertex relabellings which preserve adjacency; the group operation is composition of labellings). If the graphs are isomorphic, then $A$, and hence any generator set for $A$, has elements which map a vertex of $G_1$ to a vertex of $G_2$. The converse is also true. So it suffices to compute a generator set for $A$. But this may not be an easy task. How can one nonetheless use this idea?

Let $e_i$ be an edge of $G_i$, $i = 1, 2$. Subdivide both $e_1$ and $e_2$ and connect the division points $u_1, u_2$ by new edge $e$; this gives another trivalent graph $H_e$. Now $G_1$ and $G_2$ are isomorphic iff for some such $e$, there is some element of $\mathrm{Aut}(H_e)$ that flips the endpoints of $e$.

We are back to computing the generator set of the automorphism group of a graph. But the graph is trivalent, and hence the automorphism subgroups stabilising individual edges ($A_0$ below) are 2-groups. Furthermore, and this is crucial, the graph has a special edge $e$, a bridge between two parts, and this provides the key to an incremental construction as follows: For graphs $X_r$ of increasing "radius" centred around the edge $e$ (strictly speaking, $X_r$ consists of all length $\leq r$ paths containing $e$), figure out the ways in which these vertices/edges can be relabelled preserving adjacency. (For distance 0, only two labellings are possible: identity, and the map that flips $u_1$ and $u_2$.) From each such set of permissible relabellings, now figure out permissible relabellings for adjacent edges/vertices.

Group theoretically, construct generator sets for the automorphism groups $A_r$ of $X_r$. Each $A_{r+1}$ restricted to $X_r$ is in $A_r$, and the natural homomorphism from $A_{r+1}$ to $A_r$ has as kernel those relabellings that pointwise fix $X_r$. Given a generator set for $A_r$, we can first figure out this kernel, and then use it to find a generator set for $A_{r+1}$. Eventually we construct $A_n$, which has the information we need.

The incremental construction requires polynomial time (poly in $n$) algorithms for the following group-theoretic questions: given a group (which is a subgroup of $S_n$) presented by a generator set, (1) find the group size. (2) test for membership in the group. (3) given poly time membership testing for a "large" subgroup, find a generator set for the subgroup. In describing the algorithms, we were led through a nice exposure of solvability, orbits, stabiliser subgroups, blocks and primitivity, blocks and subgroups, block systems, Jerrum's filter, Schreir's generator, and more. One had to remind oneself periodically that this delving deep into the fascinating world of groups was being done at that point just to address graph isomorphism! The original paper describing this in detail, is, I am told, quite accessible/readable, and can be found at [17].

There was more to follow from Arvind, on connecting Galois field theory with the question of whether a polynomial is solvable by radicals (i.e. by a circuit with $+$, $\times$ and radical $x^{1/n}$ gates.) Landau and Miller showed that if this question is poly-time decidable, then there is an effective poly-time construction of the required circuit as well. However I must confess to having lost out on this part, being satiated and saturated with the isomorphism algorithm.

Incidentally, the tower of groups construction used in the GI algorithm, with sifting down the cosets to obtain the desired representation, has a fascinating history involving breaking the Iron Curtain; see Babai's article in Calude's book [9].

## Entropy

> Holy Entropy! It's boiling!
> Mr. Tomkins in [6]

There is some kind of mystique about Entropy, be it in Physics, Mathematics or Computing. It has been said [4, 7], that von Neumann suggested the term to Shannon, saying that "it will give you a great edge in debates because nobody really knows what entropy is anyway".

Indeed, our first encounter seems to be confirm the worst fears: the entropy of a (discrete valued) random variable $X$, denoted $H(X)$, is defined to be

$$H(X) := -\sum_x \Pr[X = x] \log \Pr[X = x],$$

i.e. it is (the negative of) the expected value of logarithm (in base 2, of course) of the distribution function. This is quite mysterious and certainly seems quite arbitrary as a measure of the "uncertainty in $X$", or the "amount of information obtained when $X$ is revealed", as the way this fundamental formula of Shannon is usually motivated. In actual fact, it emerges quite ineluctably as the unique measure satisfying some rather natural properties that one can axiomatically require any information measure to satisfy. Good discussions of this can be found in the classic texts of Khinchin and Renyi, where it is illustrated in discussions of the famous coin weighing problems. A good modern account is the book by Cover and Thomas [3].

But in fact, the concept of entropy – its power and versatility – can be understood and illustrated most effectively through very simple examples in combinatorics and computing. One such that Pranab began this series of talks with, is *Bregman's Theorem*: the number of perfect matchings in a bipartite graph $G = (A, B, E)$ with $|A| = |B|$ is at most $\prod_{v \in A} (d(v)!)^{1/d(v)}$. Here is the gist of the argument: a perfect matching here is simply given by specifying for each $v \in A$, it's partner $\sigma(v) \in B$. Suppose $\sigma$ is such an assignment chosen uniformly at random. The entropy of this random variable is $\log N$, where $N$ is the (unknown) number of perfect matchings. Now compute this another way: $\sigma$ is determined fully by giving the values of $\sigma(v)$ for $v \in A$, and thus $H(\sigma) \leq \sum_{v \in A} H(\sigma(v)) \leq \sum_{v \in A} \log d(v)$. This shows that $N \leq \prod_{v \in A} d(v)$. Now it seems quite silly to derive this obvious fact using entropy! However, just a slightly more careful and obvious modification of this very same argument gives Bergman's Theorem.

In Computer Science proper, take the mother of all examples: sorting. Every baby knows the bound $\Theta(n \log n)$, and also (hopefully!) one favourite sorting method (Quicksort, Mergesort) and a lower bound argument (decision tree) for this. However, what if you are given some additional information: say that the input values are only of $r$ distinct kinds? Or even more generally, that there are $n_1, n_2, \ldots, n_r$ of each kind respectively? What is the bound now? The beautiful answer is that it is $nH(X)$ where $X$ is a random variable taking one of the distinct input values with probabilities proportional the corresponding frequencies: $n_1/n, n_2/n, \cdots, n_r/n$ respectively. This answer includes all the cases: when no information is available *a priori*, $H(X) = \log n$ (all distinct input values) and if one knows there are $r$ distinct values (but not how many of each kind), $H(X) = \log r$ and the sorting bound is $n \log r$. Moreover, some simple modifications of the standard sorting algorithms result in an *output sensitive* performance achieving these bounds. (Incidentally, it also shows that sorting goes better when there are non-distinct elements. I once heard a speaker at a conference claim the opposite! He was speaking on a very technical result about the tail distribution of quicksort, employing some very high powered generating function and complex analytic technology ...)

Or take this curious geometrical fact: given $n$ distinct points in three dimensions, project them onto the $yz$, $zx$ and $xy$ planes to get $n_x$, $n_y$ and $n_z$ points respectively. Then $n^2 \leq n_x n_y n_z$. There is a proof by Cauchy-Schwarz, but it is always a challenge to reproduce it. On the other hand, there is a charming four line proof via entropy: take a point at random and compute the entropy two different ways, directly as $\log n$ and indirectly as the result of revealing the point via its three projections. The crucial point is intuitively, that the latter gives the full information "twice over" (since the information in each coordinate is repeated twice).

This simple example has a fairly natural generalization called *Shearer's Lemma*. The statement is somewhat technical, so I won't give it here, but the proof is almost the same as that above. But what is truly impressive is the ease with which the inequality can be applied in diverse situations to impressive effect. One example was to counting independent sets in regular bipartite graphs. But the most impressive and elegant one was the solution by Friedgut and Kahn to the following question of Erdös (is it? must be!): given a graph $H$, what is the maximum number $N(H, \ell)$ of copies of $H$ that can appear in a graph with $\ell$ edges? The answer is:

$$c_1 \ell^{\rho^*(H)} \leq N(H, \ell) \leq c_2 \ell^{\rho^*(H)},$$

where $c_1, c_2 > 0$ are constants and $\rho^*(H)$ is the *fractional cover number* of $H$ i.e. the optimum of the LP relaxation of the natural integer LP formulation for the optimum edge cover of $H$. We saw an alternative proof of the result (also due to Friedgut and Kahn) in the harmonic analysis series which the authors claim is "even more interesting" (however it doesn't extend to hypergraphs).

The next segment of talks dealt with a notion of *graph entropy* due originally to Janos Körner. I have often been motivated to look at it via the problem of prefix-free codes which is discussed in elementary algorithms courses: assign binary codes to a bunch of $n$ objects so that no two codes are prefixes of each other. We want to minimise the average (or total) length of the code words, given the frequencies of occurrence of the objects. The prefix-free condition is motivated by the need to decode online. Huffman coding comes within 1 of the optimal which is $H(X)$ where $X$ is a random variable taking the different objects with probabilities proportional to the frequencies of occurrence. Now the generalization is to do the same for objects not all of which need to be distinguished from each other. There is a graph on the underlying objects as vertices, where an edge between two objects indicates that they must be distinguished from each other. The classical case corresponds to the complete graph. Graph entropy is an approach to this problem, which seems to be tantalizingly still largely unresolved.

Starting with the seminal paper of Körner where he used graph entropy to give a very simple proof of earlier results by Fredman and Kömlos on sizes of covering families, graph entropy has been successfully applied to a variety of problems in combinatorics and complexity. So it certainly seems to be a candidate for the essential toolkit. If one looks at the literature on the subject, however, we are confronted with two seemingly completely different and both equally mystifying definitions, but nevertheless allegedly equivalent. One is a definition with a distinctly combinatorial optimization flavour: the entropy of a graph $G$ is the minimum of the convex function $-\sum_v \log a_v/|G|$ where $a := (a_v, v \in V(G))$ ranges over the *vertex packing polytope* of $G$ i.e. the convex hull of the independent sets in $G$. The other definition, on the other hand, seems hardcore information theory: it is the minimum of the *mutual information* over all possible joint distributions, of the pair $(X, Y)$, where $X$ ranges over the vertices of $G$ and $Y$ over its independent sets subject to the condition that $X \in Y$ with probability 1.

Promising to clear the mists, Jaikumar started his talks with the following simple puzzle: what is the minimum number of bipartite graphs whose union will cover the complete graph on $n$ vertices? There are a number of ways to tackle this.

One is the deletion method: try to maintain as large an independent set as permitted by the edges of the bipartite graphs. At the outset we have the full vertex set. Consider the first bipartite graph: now we must delete some vertices to maintain the independence property. But we can always retain the larger of the two halves of the bipartition. Thus, as we consider successive bipartite graphs, we can keep at least a half of the vertices. At the end, we have only a single vertex (a maximal independent set in the complete graph). Hence we need at least $\log n$ bipartite graphs. Now, one can tinker with this deletion process (making it probabilistic!) and we are led to the polyhedral definition. I believe this is an original route to arrive at the definition and Jaikumar should be given the credit officially somewhere (could he be persuaded to convert the notes into a full fledged survey for publication?).

The other is the entropy method: a vertex in the graph is completely specified by giving its colour in each of the bipartitions. Two different vertices cannot have the same colour in all the bipartite graphs, because the edge connecting them must appear in at least one of them. So, pick a vertex at random, and let $\chi_i$ be its colour in bipartite graph $i$. Now compute the entropy of the random vertex in two different ways as usual and we get the same result: at least $\log n$ bipartite graphs are necessary. Once again, one can improve this argument, and this leads naturally to the information theoretic definition.

Amongst the most basic and useful properties of graph entropy is its *subadditivity*: the entropy of a union of graphs is at most the sum of the entropies of the components. This lies at the heart of its nature as a measure of complexity and is repeatedly used in all the applications. A beautiful side light is that it also provides an information-theoretic characterization of Berge's *perfect graphs*: a graph $G$ is perfect iff it *splits* entropy perfectly i.e. $H(G) + H(\overline{G}) = H(K_n)$ $(= \log n)$.

The subadditivity of entropy led Ravi Boppana to give an elegant lower bound argument in parallel complexity. It was this argument that inspired Kahn and Kim to their wonderful solution to the problem of sorting under partial information mentioned earlier. The argument is very easy to outline: at the start we have the partial order $P$ identified with the underlying comparability graph $G$, and at the end we have the total order whose underlying comparability graph is $K_n$; so the sorting process leads to an *increase* in graph entropy by $H(K_n) - H(P) = \log n - H(P)$. So, if we are able to show the change in entropy resulting from a single comparison is

$\Theta(1/n)$, then the total number of comparisons required is $n(\log n - H(P))$. How does this relate to $\log e(P)$, the information theoretic bound? Kahn and Kim show that for some constant $C$,

$$n(\log n - H(P)) \geq \log e(P) \geq \max(\log(n!) - nH(P), Cn(\log n - H(P))).$$

Second they show that the required comparison always exists: in any poset, the increase in entropy after a single comparison is never more than $2/n$ but for suitable choice of the comparison is at least $c/n$ for some constant $c > 0$. Given the fact that there is a polynomial time deterministic algorithm to compute graph entropy via the ellipsoid algorithm, we then have a complete answer to the problem of sorting under partial information. Interestingly, the comparison made at any stage need not always involve the most balanced pair! The simplicity of the strategy notwithstanding, the proof is quite elaborate and involves *inter alia* a nice laminar decomposition of the poset which might have other applications. For details, you could see [14], but there's a much better exposition in the workshop lecture notes [1]. An open problem is to get an algorithm that avoids the ellipsoid algorithm.

# Probability

A large series of talks was devoted to tools for probabilistic analysis. The beginning was with something one requires in probably the very first probabilistic analysis one goes through — some nice simple randomized algorithm has an equally nice expected run time, but to be practical we need to bound the probability that the runtime significantly deviates from its mean. Fortunately, in most cases this probability is gratifyingly small (i.e. measure is concentrated around the mean), but how does one go around showing this? Devdatt described how isoperimetric inequalities provide a meta-solution to be appropriately invoked. Talagrand's isoperimetric inequality TII, used with his weird notion of convex distance, is amazingly easy to apply and often provides tail bounds better than Chernoff-Hoeffding. So far, this sounded like hard sell of something relatively unknown, so then Devdatt proceeded to give concrete evidence — a series of examples where either the (non-uniform) method of bounded differences NMBD, or the usage of hereditary functions of index sets HFIS, using TII, easily yields quite sharp bounds. This could not but convince the sceptics, who then wished to see more of why TII holds or how it gives NMBD or HFIS, and this then led us through another maze where an info-theoretic lemma for measure concentration popped up, involving the mysterious information theoretic concept of *divergence*. Janson's inequality peeked in, and Talagrand's convex distance function was partially demystified.

A most interesting aside was a surprising combinatorial result with an even more surprisingly simple proof; I offer the result here with the proof left as a puzzle to the reader. Given any number of points in the unit square, there is a tour through them such that the sum of the squares of the edges in the tour is at most 4!

Further on, Devdatt talked about qualifying dependence of variables. In analysing a random event, one often needs independence of sub-events, or at least $k$-wise independence. The theme here was that in practice, it is not always necessary to quantify the dependence; it often suffices to qualify the dependence as positive or negative. Examples followed for positive dependence. For negative dependence, there is apparently not much literature, despite the hope that negative association or regression will give sharper concentration of measure.

The BK inequality due to Berg and

Kesten is an analogue of the FKG inequality for negative dependence. It applies to a *disjoint* conjunction of two events in a probability space i.e. the two events occurring for "disjoint" reasons. The inequality states that for monotone events in the discrete space $\{0,1\}^n$ equipped with the uniform measure (or product measure derived from the component spaces), the probability of the disjoint occurrence of two events is at most the product of the probabilities of the events, as is intuitive. There is a simple proof of this inequality via a coupling-like construction. However, the inequality for non-monotone events was left open and remained so for almost a decade. Finally, just a few years ago, the general case was resolved by Dave Reimer, a student of Jeff Kahn. It is quite a story. Reimer was almost at the point of leaving academia to enter Wall Street, and it wasn't clear he even had a thesis in hand. One day, perusing the stuff Reimer had left, Kahn discovered that he had proved the general conjecture, which is now the Berg-Kesten-Reimer (BKR) Theorem: Any product space satisfies the BK inequality for arbitrary events!

There is something very unusual about the proof. First one can easily reduce the general case to the case of the simple space $\{0,1\}^n$ with the uniform measure, so it's just a counting problem. Or so it seems, because no direct counting proof seems to work. Nor even can one exhibit or assert an injective mapping as required. The insight of Reimer was to transform the problem into the domain of linear algebra, where a dimension argument gives the result! The linear algebra involved is quite elementary, but nevertheless seems indispensable. Expository accounts are now available in [10].

Further on in this series, C R Subramaniam tried to outline the use of Janson's inequalities in proving properties of random graphs. For instance, I had not realised how non-trivial it is to estimate the number of $K_4$'s in a random graph.

I hope that some decent lecture notes from this workshop will materialise particularly for the probabilistic tools stuff. (Can't push this demand too aggressively, though, not having written up my own lecture notes yet!) Even better, a reference text should be written and published; royalties will be non-zero as I promise to keep a copy at my elbow as I work ... . Maybe Devdatt's book-in-the-making will fill the gap.

# Embeddings and Projections

Ramesh gave a series of talks centered around embeddings and projections in Euclidean space. These techniques have applications to a surprisingly wide variety of algorithmic problems.

Many algorithms in computational geometry, for example, suffer from the *curse of dimensionality*: they are exponential in the dimension of the ambient space in which the geometric objects such as points reside. Suppose one could somehow squash the points into a lower dimensional space: then we could be rid of the space. The *Johnson-Lindenstrauss Lemma* seeks to do precisely this. Of course nothing comes for free! What we can hope to do is to retain only some approximation of the original metric structure, and there is a tradeoff between how good the approximation is and how low a dimension we can come down to: to preserve all distances within a multiplicative factor $(1+\epsilon)$, we can still go down to $\log n/\epsilon^2$ dimensions. How does one achieve this? The Probabilist Method! Choose a random subspace of that dimension and project! What does choosing a random subspace mean? Choose the normal to the hyperplane at random, which, in turn, means that each of its components must be chosen randomly and independently. In fact, to retain spherical symmetry, which is very useful in the arguments, one is led to choose the components to be normally distributed. Thereafter, some routine calculations with Chernoff bounds show that any unit vector is distorted by no more than a multiplicative factor of $(1+\epsilon)s$ where $s$ does not depend on the points themselves. A good place to see the details is in some notes of Leonard Schulman [11].

Given an arbitrary finite metric represented by an edge-weighted graph (so that we are referring to the metric induced by the shortest path), can we realise the metric by an embedding into Euclidean space under the usual metrics? In the context of an NP-hard optimization problem on graphs called the *sparsest cut*, it turns out that one is interested in $\ell_1$ embeddings. Once again, it is not always possible to achieve a perfect embedding, but one can get an embedding with a multiplicative distortion factor of $O(\log n)$. The embedding due to Jean Bourgain is in terms of a family of subsets of the vertex set: each subset in the family corresponds to a dimension in Euclidean space, and the co-ordinate for vertex $v$ in dimension corresponding to the set $A$ is $d(v, A)$, the minimum distance of $v$ from the set $A$ (together with a scaling factor). For an appropriate choice of the family of subsets, we get the low distortion embedding. Bourgain showed that one can take the family $\bigcup_i \binom{V}{2^i}$. Of course this is useless for algorithmic purposes, because this involves an exponential number of dimensions! However, the proof shows that one doesn't need all the sets: a polynomial sized random sample in each size of sets is sufficient. One can therefore alternatively say, as Ramesh did, that one should pick a random family of sets, just use different sampling rates. An account (not the best written though!) of this technique under the title "The Geometry of Graphs" is in [12].

For other graph problems, one requires something stronger: not just to preserve the length of edges, but perhaps to preserve the "volume" of simplices of vertices. An example of this kind is the NP-hard problem of *bandwidth minimization*: given a graph, embed it on a straight line so as to minimize the maximum stretch of an edge. This is the problem for which Feige was led to consider a generalization that would embed a graph so that the "volume" of any $k$ vertices of the graph is preserved. As a measure of the "volume" of $k$ vertices, Feige proposes the minimum spanning tree induced by those vertices, and seeks an embedding such that the Euclidean volume of the embedding of any $k$ vertices reflects the MST of those vertices faithfully. Feige achieves an approximation factor which is polynomial in $k$ and $\log n$. The idea of the embedding is inspired by the Bourgain technique but is considerably more involved with some deep geometric insights such as a Brunn-Minkowski type inequality. Feige told me that he was looking forward to comments by the referees but instead the editor accepted the paper without waiting for them! Both a testimony to the importance of the paper as well as an injunction for us to study it carefully. See the paper [5].

# Algebra

Saugata Basu gave a series of 3 talks on algebraic curves and codes. Most of the time was devoted to an exposition of what the Riemann-Roch Theorem means; towards the end, its use in constructing Goppa codes was described. The amount of complex analysis covered was probably enough for half an advanced-level course, and thus was not all digestible, but one could appreciate the complexity involved.

K V Subrahmanyam, in a series of 4 talks, described the recent approach of Ketan Mulmuley, Milind Sohoni et al in using the framework of algebraic geometry to pose and hopefully answer central questions from complexity theory. Again, the material was far too esoteric for me to digest, but at least we got an idea (however faint) of how a complexity question could be phrased as a problem in standard algebraic geometry. One then entertains the hope that the powerful machinery of algebraic geometry might finally be able to resolve some of our long standing conundrums. It seems likely therefore that one may soon need to understand much of this better.

Vinay gave a talk on the *Singular Value Decomposition* (SVD) and its applications in Computer Science. Starting from the simplest case - diagonal matrices, "homework exercise" - to square matrices, he proceeded in a model of systematic exposition to reveal how any matrix $A$ has a decomposition of the form $A = U\Sigma V^T$ for orthogonal matrices $U$ and $V$ and a diagonal matrix $\Sigma$. The diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq \sigma_{r+1} = 0 = \cdots = \sigma_n$ — the so called *singular values* — are of the utmost importance in revealing the properties of the matrix. For example, $r$ is the rank, $\sigma_1$ the 2-norm and $\sum_i \sigma_i^2$, the *Frobenius norm* of the matrix. The fundamental *Hoffman-Weilandt inequality* states that $||A-B||_F^2 \geq \sum_i (\alpha_i - \beta_i)^2$, where $\alpha_i, \beta_i$ are the singular values of $A, B$ respectively in sorted order. This immediately implies that the best rank $k$ approximation to a matrix $A$ is to take the singular value decomposition and ignore ALL except the first $k$ singular values. This is a fundamental technique employed in *latent semantic indexing* used in classifying documents for informational retrieval. The technique has, unsurprisingly, migrated to the Internet and Vinay later gave us a live demonstration of *Manjara*, a meta search engine that does classification of search results based on the SVD. (Check it out at `http://cluster.cs.yale.edu`. If you forget the URL, no problem; google returns this site on top if you query Manjara. If you even forget the name Manjara, still no problem; visit Vinay's homepage. Vinay who??) Another cute application of the SVD was in circuit complexity to prove a well-known bound due to Krapchenko. Actually one can prove a slightly stronger bound: the smallest formula size for a function $f$ is at least $\sigma_1^2(Q)$, where the matrix $Q$ is formed as follows: the rows are indexed by inputs for which the function takes value 1, the columns by inputs for which the function evaluates to 0, and the $(i,j)$ entry is 1 if the corresponding inputs differ in exactly one bit (and zero otherwise).

In a second talk Vinay talked about *Szemeredi's regularity lemma*, which has increasingly played an important role in recent times in graph theory and related Computer Science applications. However, the statement of the lemma itself is so involved — hardly has one $\epsilon$ been dealt with than a new $\delta$ (depending on the just mentioned $\epsilon$ pops up — that time ran out with just about the statement of the lemma completed. Perhaps a just topic for a sequel.

# Others

There were some guest talks: Vijay Chandru talked about linear constraints and logic. Subhashis Bannerjee gave a nice exposition of the difficulties in using algebro-geometric methods in computer vision (even though one is dealing with fairly small matrices and dimensions); the exposition was enlivened with pictures of 3D panoramic scenes reconstructed using some of these methods. Sundarrajan talked about coding theory, in particular Euclidean space codes and trellises.

For sheer fervour and direct contact with the audience in presenting material, he was unmatched at this (and many other) workshops! It was good to hear an exposition of how coding is moving towards reaching Shannon's limit etc from an engineer (he is from the EE Dept) - none of the cynical "Ah, so that's been done now, is it?" and lots of the "how wonderful, such a simple idea but it works". At the end, he mentioned several open problems and urged us computer scientists to find the solutions. Such heartening faith in us!

## Winding up

The two weeks at Bangalore provided quite a stimulating and exciting diet: elegant theorems, deep inequalities, wonderful results. Pity all the more that the follow up has once again left much wanting: no notes have been assembled (with one honourable exception – the TIFR group's fine collection on entropy [1]), no serious collaborations were initiated and no follow-up study seems to have been undertaken in the aftermath. It was once again a case of too many other things overtaking our lives. How can we set that right the next time?

Talking of next time, we feel that this series of workshops should be continued apace. Plenty of topics remain to be explored fully (some form this workshop, others conspicuous left-outs like Markov chain techniques) and this is, we have to repeat yet again, the best opportunity for Indian researchers to gather together and discuss the frontiers of research.

## We're not ingrates ...

The workshop was organised by Vinay and Ramesh, who did a wonderful job despite being tremendously overwhelmed by pressing simputer (`http://simputer.org`) deadlines and TV appearences on BBC ClickOnline. Three cheers for them and the rest of the IISc team. We are told that there was a major last-minute glitch in the funding, and the IISc-industrial collaboration met the shortfall admirably. Much thanks are due for all the organisation, monetary and otherwise.

## Newsflash!!!

As this report goes to press, the notorious brigand Veerappan (He Who Uses Only One Name) has relented and released the cine-star Rajkumar (who never quite made it to those high echelons where the media would call him a *thespian*). The release deal is shrouded in secrecy with all the elements of a classical suspense thriller — there was a mysterious woman (not quite veiled, alas) who played a key role, there were political extremist groups, there was an unexplained escape of one hostage before the big release, ... last heard of, on TV, Rajkumar said he had a good idea for a forthcoming film. (Truth writes better screenplay scripts than fiction!) Watch out for the release of "120 days with Veerappan"...

## References


[1] Jaikumar Radhakrishnan and Pranab Sen, "Notes on Entropy and Graph Entropy", lecture notes for the MTTCS00 workshop. (contact `jaikumar@tcs,tifr.res.in` for a copy.)

[2] V.I. Arnold, M. Atiyah, P. lax and B. Mazur (eds): *Mathematics, Frontiers and Perspectives*, American Mathematical Society, 1999.

[3] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley 1991.

[4] K. Denbigh, "How subjective is entropy?", in *Maxwell's Demon, Entropy, Infromation, Computing*, H.S. Leff and A.F. rex (eds), A. Hilger, Bristol 1990.

[5] U. Feige, "Approximating the bandwidth via volume respecting embeddings", *J. Comp. and Systems Sciences* 60(3) 510–539, 2000.

[6] G. Gamov, *Mr. Tomkins in Paperback*, Cambridge University Press, Cambride 1965.

[7] M. Tribus, *Boelter Anniversary Volume*, McGraw Hill, New York 1963.

[8] Graham R. Brightwell, *Balanced Pairs in Partial Orders*, Technical report LSE-CDAM-98-14 `http://www.cdam.lse.ac.uk/Reports/reports98.html`.

[9] Lá'szlo Babai, *The Forbidden Sidetrip*, in *People and Ideas in theoretical Computer Science*, Springer 1999 (editor C. S. Calude): 1–31.

[10] Borgs, C., Chayes, J. T. and Randall, D., *The Van den Berg-Kesten-Reimer inequality.* in *Perplexing Problems in Probability: Festschrift in honor of Harry Kesten*, Birkhauser 1999 (M. Bramson and R. Durrett, editors): 159–173.

[11] *Johnson-Lindenstrauss Embedding*, Lecture notes for course on Randomized Algorithms, Fall 1999, lecture 4, at `http://www.cc.gatech.edu/people/home/schulman/ra99.html`.

[12] Linial, London, and Rabinovich, "The Geometry of Graphs and Some of its Algorithmic Applications", *Combinatorica*, 15, 1995.

[13] Nathan Linial, Yishay Mansour, Noam Nisan: "Constant Depth Circuits, Fourier Transform, and Learnability". J.ACM 40(3): 607-620 (1993).

[14] Jeff Kahn, Jeong Han Kim, "Entropy and Sorting". JCSS 51(3): 390-399 (1995).

[15] Jeff Kahn, Gil Kalai, Nathan Linial, "The Influence of Variables on Boolean Functions (Extended Abstract)". FOCS 1988: 68-80.

[16] Nathan Linial, "The Information-Theoretic Bound is Good for Merging". SIAM J. Comput. 13(4): 795-801 (1984).

[17] Eugene M. Luks, "Isomorphism of Graphs of Bounded Valance Can Be Tested in Polynomial Time", Journal of Computer and System Sciences 25(1) (1982) 42–65.

Aneem Najaham, Dodo Piadico
Contact addresses:
⟨meena@imsc.ernet.in⟩,
⟨dubhashi@cs.chalmers.se⟩