Invariants of Incidence Matrices

Chris Godsil (University of Waterloo),
Peter Sin (University of Florida),
Qing Xiang (University of Delaware).

March 29 – April 3, 2009

1 Overview of the Field

Incidence matrices arise whenever one attempts to find invariants of a relation between two (usually finite) sets. Researchers in design theory, coding theory, algebraic graph theory, representation theory, and finite geometry all encounter problems about modular ranks and Smith normal forms (SNF) of incidence matrices. For example, the work by Hamada [9] on the dimension of the code generated by r-flats in a projective geometry was motivated by problems in coding theory (Reed-Muller codes) and finite geometry; the work of Wilson [18] on the diagonal forms of subset-inclusion matrices was motivated by questions on existence of designs; and the papers [2] and [5] on p-ranks and Smith normal forms of subspace-inclusion matrices have their roots in representation theory and finite geometry.

An impression of the current directions in research can be gained by considering our level of understanding of some fundamental examples.

Incidence of subsets of a finite set

Let X_r denote the set of subsets of size r in a finite set X. We can consider various incidence relations between X_r and X_s, such as inclusion, empty intersection or, more generally, intersection of fixed size t. These incidence systems are of central importance in the theory of designs, where they play a key role in Wilson's fundamental work on existence theorems. They also appear in the theory of association schemes.

Incidence of subspaces of a finite vector space

This class of incidence systems is the exact q-analogue of the class of subset incidences. The possible incidence relations are inclusion or, more generally, intersection in subspace of fixed dimension. These examples have been studied for their relation to questions in representation theory of the general linear group. In some cases, they have been applied to finite geometry and to construct error-correcting codes.

Incidence of distinguished subspaces of a vector space

In the presence of a quadratic, Hermtian or symplectic form, we may refine the above incidence systems by considering distinguished subspaces such as totally isotropic or non-singular ones. The corresponding classical group acts and there are connections to its representation theory and to the geometry of the associated polar spaces.

General Problem: Computation of invariants

Incidence matrices have invariants at several levels of rigidity. If we consider the matrices as representing linear maps over fields then we wish to compute its eigenvalues and rank in every characteristic p (p-rank for short). Since there is usually a group G acting which preserves the incidence relation, the linear map becomes a homomorphism of G-modules, raising deeper questions about the G-module structure of the domain, codomain, image and kernel of the map.

The incidence matrix is integral, and can also be regarded as the matrix of a homomorphism of free abelian groups. Thus, the invariant factors (or Smith normal form) of the matrix form a stronger set of invariants than the p-ranks, which can be deduced immediately from the former. This time the group action raises questions about representations over the integers and over p-adic rings.

Finally, incidence matrices have been used as parity check or generator matrices of codes. Then the relevant invariants are those which are preserved by automorphisms of the code, such as the minimum weight of a codeword or, more generally the weight enumerator.

Thus it can be seen that there is a multitude of natural problems, depending on the choice of incidence system and the choice of invariant. These problems share many common features but their origins and the reasons for studying some of them are very diverse, so that published work on these questions is scattered across the literature of the subdisciplines. It is no easy task just to keep track of which ones have been answered!

2 Presentation Highlights and Scientific Progress Made

Design theory

Let v, k, t and λ be integers with $v \geq k \geq t \geq 0$ and $\lambda \geq 1$. A *t-design* on v points with block size k and index λ is an incidence structure $\mathcal{D} = (X, \mathcal{B})$ with:

1. $|X| = v$,

2. each $B \in \mathcal{B}$ is a k-element subset of X,

3. for any set $T \subset X$ of t points, there are exactly λ blocks containing all points in T.

Next we define a class of subset-inclusion matrices. Let X be a v-set. Let W_{tk} denote the $\binom{v}{t}$ by $\binom{v}{k}$ matrix whose rows are indexed by the t-subsets of X, whose columns are indexed by the k-subsets of X, and where the entry in row T and column K is

$$W_{tk}(T, K) = \begin{cases} 1, & \text{if } T \subseteq K, \\ 0, & \text{otherwise} . \end{cases} \tag{1}$$

With the definition of W_{tk}, it becomes clear that a t-design is nothing but a $\binom{v}{k}$ by 1 vector x with nonnegative integer entries such that

$$W_{tk}x = \lambda \mathbf{j}, \tag{2}$$

where $\mathbf{j}$ is the all-one $\binom{v}{t}$ by 1 vector. Therefore investigating the Smith normal form of W_{tk} is important for the study of t-designs. Rick Wilson in his talk spoke of his work [18] on a diagonal form of W_{tk} and various applications of this result, including application to a zero-sum Ramsey-type problem. It should be noted that the p-ranks of W_{tk} and the p'-case of the subspace-inclusion matrices were treated by representation theoretic methods in the work of Frumkin and Yakir [8]. Rick Wilson also noted that empty intersection relation between subsets and the inclusion relation are essential the same in the set case. This seemingly trivial point was made by at least three of the speakers. Navi Singhi talked about his work on tags on subsets [16], and G. B. Khosrovshahi described his work with his collaborators on special bases of the null space of W_{tk}. These treatments seek to impose orderings on the object, in other words to break their symmetry. Both talks have a strong algorithmic flavor. Vladimir Tonchev talked about his recent work with Dieter Jungnickel on counterexamples to the Hamada conjecture, which states that the geometric designs $\mathrm{PG}_d(n, q)$ and $\mathrm{AG}_d(n, q)$ are characterized as the designs of minimum p-ranks among all designs with the given parameters. It should be noted that Hamada's conjecture implies that for any prime p, the only projective plane of order p is $\mathrm{PG}(2, p)$. Previously, only a few counterexamples (with concrete parameters) to Hamada's conjecture were known. Recently Jungnickel and Tonchev [10] constructed an infinite family of counterexamples. However it should be noted that Hamada's conjecture for symmetric designs with classical parameters is neither proved nor disproved.

Related to Singhi's talk, we mention that so far attempts to define a theory of tags for vector spaces have not been fruitful. But in work of Paul Li [13] solving the conjecture of Brouwer on the 2-rank of the symplectic dual polar spaces over $\mathrm{GF}(2)$, one can clearly see similar ideas about ordering (i.e., breaking symmetry) applied to good effect.

Strongly regular graphs

A *strongly regular graph srg* (v, k, λ, μ) is a graph with v vertices that is regular of valency k and that has the following properties:

1. For any two adjacent vertices x, y, there are exactly λ vertices adjacent to both x and y.
2. For any two nonadjacent vertices x, y, there are exactly μ vertices adjacent to both x and y.

It is well known that strongly regular graphs are equivalent to two-class association schemes. Many of the problems above had already been considered in the context of strongly regular graphs. This area has been a rich source of examples and interesting problems on invariants of incidence matrices (more appropriately, adjacency matrices). There are theorems which characterize graphs by the invariants and interesting examples of non-isomorphic graphs with the same invariants. In his talk, Andries Brouwer surveyed the results from his work with Van Eijl [4]. For various graphs the SNF of the adjacency matrix is given. Kneser graphs are defined (graphs on flags of a building of spherical type, adjacent when far apart) and it is shown by examples that in the thin case these Kneser graphs often have the property that the SNF of the adjacency matrix A equals the SNF of the diagonal matrix with the spectrum of A on the diagonal, while in the thick case the SNF has only powers of p. Quite a few very interesting remarks were made in the talk. For example, Brouwer commented that it is generally easier to consider the relations of "far apart" rather than "close together". Examples of this include Kneser graphs and their q-analogues. This philosophy is borne out in the vector space setting where we know the p-ranks for r-dimensional subspaces versus s-dimensional subspaces for all r and s when the relation is zero intersection, but we know the p-ranks only when either r or s is equal to 1 if we consider inclusion. A nice open problem along these lines is to compute the integral invariants for zero intersection of r-subspaces and s-subspaces in projective space. After showing an old proof by Brouwer and Van Eijl for the p-ranks of Paley graphs and large submatrices, Brouwer commented that the "and large submatrices" part is interesting: representation-theoretic methods usually give the p-rank of the full matrix but do not give information on submatrices. Willem Haemers talked about the work of his former student Rene Peeters [14] on p-ranks and SNF of distance-regular graphs.

Representation theory

In recent years representation theory has proven itself to be an extremely powerful tool for the exact calculation of p-ranks.

New results announced at the workshop included the solution by P. Sin of the p-rank problem for point-hyperplane incidences in orthogonal geometries, originally raised in a 1995 paper of Blokhuis and Moorhouse [3], and the solution of the analogous question for hermitian geometries by P. Sin and O. Arslan. This work applies fairly sophisticated techniques of representation theory of algebraic groups in characteristic p, such as the Jantzen Sum Formula and the theory of good filtrations. The p-ranks in question turn out to be the (previously unknown) dimensions of irreducible representations and the above theory reduced the problem to some complicated but tractable combinatorics.

Representation theory is an important tool in the work Bardoe-Sin [2] describing the permutation module for $\mathrm{GL}(n, q)$ on the points of projective space. This work yields the

p-rank of many incidence systems on which $\mathrm{GL}(n, q)$ acts, including the famous Hamada formula for the p-rank of points versus subspaces of a fixed dimension. Later the p-rank for the incidence relation of zero intersection between r-subspaces and s-subspaces for any r and s was determined from detailed knowledge of this module.

The work of Chandler, Sin and Xiang [6, 7] on the p-ranks for symplectic spaces also depends heavily on representation theory. A special case of their computations which is of interest to other areas include the p-ranks for the symplectic generalized quadrangles in characteristic p, which are consequently all known now.

The success of representation theory in the above problems suggests applying the representation theory of the symmetric group to incidences of subsets of a set. In a way the elegant matrix method of Wilson serves the same purpose as representation theory. Nevertheless, it may still be an instructive exercise to recast this body of work in the language of symmetric group representations, Young tableaux, Specht modules etc. This might also throw some light on open problems such as the incidence of subsets with prescribed intersection size.

Coding theory

Recently, electrical engineers have been interested in low-density parity-check (LDPC) codes defined by incidence matrices of generalized polygons. The dimensions of the codes are 2-ranks, which are known, having been computed by eigenvalue methods and representation theory. However, one can also ask questions about weight enumerators, for which such methods cannot be used. The earliest work in this direction was by Bagchi and Sastry [1] but the subject has been dormant until the recent interest. L. Storme, J-L. Kim, K. Mellinger and others have brought new life to the subject. In this talk, Leo Storme survey his recent results with his collaborators on codewords of small weights in codes arising from projective planes, on codewords of small or large weights in codes arising from the classical generalized quadrangles. The methods used here are mainly from finite geometry. See [12, 11] for more details.

Computation

Incidence matrix problems can easily stretch computers to the limit. For example, in the process of computing the SNF of a $(0, 1)$-incidence matrix, the entries of the matrices arising from intermediate steps can get extremely large even though the entries in the original matrix are very small (here the entries are 0 or 1). Saunders gave an overview of the LinBox package, which contains efficient algorithms for computing SNF of integral matrices. In particular, Saunders explained the importance and effectiveness of probabilistic algorithms. Brouwer explained the need for ways to parallelize computations.

3 Open Problems

Many open problems and conjectures were proposed in the talks of the workshop or during informal discussions. Some were already mentioned in previous sections. Here we collect a few of them.

1. Let V be an $(n+1)$-dimensional vector space over $\mathrm{GF}(q)$, where $q = p^t$. For any i, $1 \le i \le n$, we use $\mathcal{L}_i$ to denote the set of all i-dimensional subspaces of V. For integers r, s, $1 \le s \le r \le n$, let $A_{r,s}(q)$ denote the (0,1)-incidence matrix with rows indexed by elements Y of $\mathcal{L}_r$ and columns indexed by elements Z of $\mathcal{L}_s$, and with (Y, Z)-entry equal to 1 if and only if $Z \subseteq Y$. The p-rank of $A_{r,s}(q)$ is known when $s = 1$. What is the p-rank of $A_{r,s}(q)$ when $1 < s < n$?

2. Using the notation in Problem 1, for integers r, s, $1 \le s \le r \le n$, let $B_{r,s}(q)$ denote the (0,1)-incidence matrix with rows indexed by elements Y of $\mathcal{L}_r$ and columns indexed by elements Z of $\mathcal{L}_s$, and with (Y, Z)-entry equal to 1 if and only if $Z \cap Y = \{0\}$. The p-rank of $B_{r,s}(q)$ is known from the work of P. Sin [15]. What is the SNF of $B_{r,s}(q)$?

3. In [6, 7], the symplectic analogues of Hamada's formula were given. How about orthogonal and Hermitian analogues of Hamada's formula?

4. How to define tags on subspaces of a finite dimensional vector space so that we can use them to solve p-rank and SNF problems for incidence relations between subspaces?

5. In [17], it was shown that every commutative semifield of order congruent to 1 modulo 4 gives rise to a strongly regular graph with Paley parameters (or, a pseudo-Paley graph, for short). Assume that q is an odd prime power. Let j be a nonsquare in $K = \mathrm{GF}(q)$, and let $1 \neq \sigma \in Aut(K)$. The Dickson semifield $(K^2, +, *)$ is defined by
$$(a, b) * (c, d) = (ac + jb^\sigma d^\sigma, ad + bc).$$
Let
$$D(q, \sigma) = \{(x^2 + jy^{2\sigma}, 2xy) \mid (x, y) \in K^2,\ (x, y) \neq (0, 0)\}, \tag{3}$$
i.e., D is the set of nonzero "squares" of the Dickson semifield. Then the Cayley graph $X(K^2, D(q, \sigma))$ with vertex set K^2 and connecting set $D(q, \sigma)$ is a pseudo-Paley graph. Let $q = 3^t$, let A be the adjacency matrix of $X(K^2, D(3^t, \sigma))$, and let $r_t = \mathrm{rank}_3(A)$ (i.e., the rank of A over $\mathrm{GF}(3)$). The first few terms of the sequence $(r_t)_{t \ge 1}$ were computed by David Saunders and Guobiao Weng. For example, $r_1 = 4, r_2 = 20, r_3 = 85, r_4 = 376, r_5 = 1654, r_6 = 7283, r_7 = 32064$. Based on the above data, David Saunders conjectured that
$$r_t = 4r_{t-1} + 2r_{t-2} - r_{t-3},$$
for all $t \ge 4$. The significance of the conjecture lies in that its validity immediately implies that the pseudo-Paley graph constructed from the Dickson semifield (where $q = 3^t$) is not isomorphic to the Paley graph with the same parameters.

4 Conclusion

Invariants of incidence matrices have been studied by researchers in algebraic graph theory, representation theory, design theory and coding theory. Bringing together people working on distinct but overlapping and strongly analogous problems has helped to create a clear view of what is known and what are important open questions. The lectures and informal discussions have gone a long way to clarifying the relationships between the different theories, the role of different technical approaches and how the main problems fit together into a unified scheme. The territory has been charted clearly. We can now see the holes in our knowledge which can soon be filled in by existing methods and also the cutting edge problems where new results will mark significant progress. Experimental methods have also been examined in depth, with clear evidence of the value of computer work for producing conjectures and expert discussion of the precise limitations of computers in handling incidence problems.

A good example of work exhibiting influences from many sources was the talk of David Chandler. In his work with P. Sin and Q. Xiang on the integral invariants for incidence of points and subspaces of a fixed dimension in a finite projective space, combination of character sums and p-adic methods such as Stickelberger's theorem and Wan's theorem with representation theory of the general linear group. The character sums can be considered a technique imported from the theory of difference sets. He also showed how these ideas could be applied to solve a classical problem in Galois geometry on the size of intersections of unitals.

In organizing this workshop our goals were to provide a general context for a broad range of analogous problems which previously may have appeared isolated, and to publicize certain methods, approaches and problems from the component subdisciplines which were either unknown or had never been tried by researchers with other backgrounds. Our scientific aims will have been achieved if the workshop has accelerated the adoption of new methods, provoked interest in open problems and provided a framework for future collaborative work between different subdisciplines in which invariants of incidence matrices are important.

References

[1] B. Bagchi, N. S. Narasimha Sastry, *Codes associated with generalized polygons*, Geom. Dedicata **27** (1988), 1–8.

[2] M. Bardoe and P. Sin, *The permutation modules for* $\mathrm{GL}(n+1, \mathbf{F}_q)$ *acting on* $\mathbf{P}^n(\mathbf{F}_q)$ *and* $\mathbf{F}_q^{n+1}$, J. London Math. Soc. **61** (2000), 58-80.

[3] A. Blokhuis and E. Moorhouse, *Some* p*-ranks related to orthogonal spaces*, J. Algebraic Combin. **4** (1995), 295–316.

[4] A. Brouwer, C. A. van Eijl, *On the* p*-rank of the adjacency matrices of strongly regular graphs* J. Algebraic Combin. **1** (1992), 329–346.

[5] D. B. Chandler, P. Sin, Q. Xiang, *The invariant factors of the incidence matrices of points and subspaces in* $\mathrm{PG}(n,q)$ *and* $\mathrm{AG}(n,q)$, Trans. Amer. Math. Soc. **358** (2006), 4935–4957.

[6] D. B. Chandler, P. Sin, Q. Xiang, *The permutation action of finite symplectic groups of odd characteristic on their standard modules*, J. Algebra **318** (2007), 871–892.

[7] D. B. Chandler, P. Sin, Q. Xiang, *Incidence modules for symplectic spaces in characteristic two*, preprint.

[8] A. Frumkin, A. Yakir, *Rank of inclusion matrices and modular representation theory*, Israel J. Math. **71** (1990), 309–320.

[9] N. Hamada, *The rank of the incidence matrix of points and* d*-flats in finite geometries*, J. Sci. Hiroshima Univ. Ser. A-I Math. **32** (1968), 381–396.

[10] D. Jungnickel and V. D. Tonchev, *Polarities, quasi-symmetric designs, and Hamada's conjecture*, Des. Codes Cryptogr. **51** (2009), 131–140.

[11] J.-L. Kim, K. E. Mellinger, L. Storme, *Small weight codewords in LDPC codes defined by (dual) classical generalized quadrangles*. Des. Codes Cryptogr. **42** (2007), 73–92.

[12] M. Lavrauw, L. Storme, P. Sziklai, G. Van de Voorde, *An empty interval in the spectrum of small weight codewords in the code from points and k-spaces of PG(n,q)*, J. Combin. Theory (A), **116** (2009), 996-1001.

[13] P. Li, *On the universal embedding of the* $\mathrm{Sp}_{2n}(2)$ *dual polar space*, J. Combin. Theory Ser. A **94** (2001), 100–117.

[14] R. Peeters, *On the* p*-ranks of the adjacency matrices of distance-regular graphs*, J. Algebraic Combin. **15** (2002), 127–149.

[15] P. Sin, *The* p*-rank of the incidence matrix of intersecting linear subspaces*, Des. Codes Cryptogr. **31** (2004), 213–220

[16] N. Singhi, *Tags on subsets*, Discrete Math. **306** (2006), no. 14, 1610–1623.

[17] G. Weng, W. Qiu, Z. Wang and Q. Xiang, *Pseudo-Paley graphs and skew Hadamard difference sets from presemifields*, Des. Codes and Cryptogr., **44** (2007), 49–62.

[18] R. M. Wilson, *A diagonal form for the incidence matrices of* t*-subsets vs.* k*-subsets*, Europ. J. Combin. **11** (1990), 609–615.