Intersections between linear algebra and combinatorics: selection of some case studies for Spring School 2000

Bruno Codenotti*

April 15, 2000

1 Introduction

These notes contain an arbitrary selection of topics in a scientific area at the intersection between linear algebra and combinatorics. These topics arise in connection with problems in theoretical computer science, especially in the realm of lower bound issues, and are, to my personal taste, quite fascinating.

They reveal a deep connection between linear algebra and combinatorics, as well as the need to combine tools from both areas in order to attack hard computational questions.

The notes are centered around four topics, which are treated in the following four sections, i.e. Shannon capacity and theta function (Section 1.1), Rank, rigidity and Ramsey numbers (Section 1.2), Rank vs Communication Complexity (Section 1.3), and Random walks on graphs (Section 1.4). For each of the four topics, the references contain several bibliographic suggestions, which are discussed in Section 2.

Most of the papers cited here can be obtained on-line.

The papers where one of the coauthors is Codenotti can be downloaded from the site http://www.imc.pi.cnr.it/ codenotti/online.html.

The papers [2, 17] can be obtained from http://www.math.cas.cz/ pudlak/.

The Electronic Journal of Combinatorics home page is http://www.combinatorics.org/, while the Electronic Colloquium on Computational Complexity home page is http://www.eccc.uni-trier.de/eccc/

At the site http://www.math.tau.ac.il/ noga/, one can find reference [1], together with other papers by Noga Alon on theta function and Shannon capacity.

*Istituto di Matematica Computazionale del CNR, Area della Ricerca di S.Cataldo, 56010-Ghezzano, Pisa (Italy). e-mail: codenotti@imc.pi.cnr.it.

¿From http://www.cs.huji.ac.il/ noam/papers.html one can download paper [16] as well as other papers by Noam Nisan on communication complexity.

¿From http://www.cs.yale.edu/ lovasz/survey.html one can download paper [13].

1.1 Shannon capacity and theta function

The notion of capacity of a graph has been introduced by Shannon in [21], and after that was labeled as *Shannon capacity.* This concept arises in connection with a graph representation for the problem of communicating messages in a channel. One considers a graph G, whose vertices are letters from a given alphabet, and where adjacency indicates that two letters can be confused. In such a setting, the maximum number of one-letter messages that can be safely sent, i.e., sent without danger of confusion, is clearly given by the independence number of G, which we denote by $\alpha(G)$. If $\alpha(G^k)$ stands for the maximum number of k-letter messages that can be safely communicated, we immediately see that $\alpha(G^k) \geq \alpha(G)^k$. Furthermore one can readily show that equality does not hold in general (see, e.g., [11] for some examples).

The Shannon capacity of G is the number

$$\Theta(G) = \lim_{k \to \infty} \sqrt[k]{\alpha(G^k)}\,,$$

which, by the previous observations, satisfies $\Theta(G) \geq \alpha(G)$, where equality does not need to occur.

The Shannon capacity can be defined in terms of *strong graph products.* We say that two vertices are adjacent if they are either connected by an edge or they are equal. Then for two graphs G and H, we define their strong product $G \cdot H$ as the graph with vertex set $V(G) \times V(H)$, where (i, j) is adjacent to (i', j') if and only if i is adjacent to i' in G and j is adjacent to j' in H. If G^k denotes the strong product of k copies of G, we immediately see that $\alpha(G^k)$ is indeed the independence number of G^k.

It was very early recognized that the determination of the Shannon capacity is a very difficult problem, even for small and simple graphs.

In a famous paper of 1979, Lovàsz introduced the “theta function” $\vartheta(G)$, with the explicit goal of estimating $\Theta(G)$ [11]. The last section of that paper raises a number of interesting issues, among those the questions of determining the Shannon capacity of odd-cycles (which we will denote by C_m, with the subscript indicating the length), and of stating whether or not odd-cycles satisfy $\vartheta(G) = \Theta(G)$. He observed that the last problem *pinpoints* the following crucial difficulty: in all cases where the value of $\Theta(G)$ is exactly known, there is some k such that $\alpha(G^k) = \Theta(G)^k$. The key remark is now that, if $\vartheta(G) = \Theta(G)$ for, say, the 7-cycle, then no such k can exist, since no power of $\vartheta(C_7)$ is an integer.

Shannon capacity and Lovász theta function attracted a lot of interest in the scientific community, because of the applications to communication issues, but also due to the connections with some central combinatorial and computational questions in graph theory, like computing the largest clique and finding the chromatic number of a graph. The reader can find a wealth of different results and applications of $\vartheta(G)$ and $\Theta(G)$ in the theoretical computer science and combinatorics literature. Despite a lot of work in the field, many basic open questions remain open, notably that of evaluating the Shannon capacity of C_7, and, more in general, of odd-cycles.

In [9] it is possible to find several examples of graphs for which the theta function is explicitly known.

1.2 Rank, rigidity, and Ramsey theory

The problem of relating the rank of a matrix to its structural properties given by the pattern of its nonzero entries is a classical problem in mathematics.

An important area in which this problem arises is that of the computation of linear forms, where the key notion is provided by *matrix rigidity.*

Matrix rigidity has been introduced and used in the context of a very difficult subject in complexity theory, namely that of proving non trivial lower bounds on the length of computations. So far there have been only a very few significant advances in the above area. As an example, despite more than 20 years of research, it is still unknown whether or not the existing FFT algorithms, which run in time proportional to $n \log n$, are optimal, i.e., whether or not $n \log n$ is asymptotically a lower bound for the computation of the n-point DFT.

Motivated by such lower bound issues, several authors have introduced suitable linear algebra techniques, which allow one to view certain *computational questions* as equivalent *algebraic questions.* One of such questions is *finding matrices with high rigidity*

The *rigidity of a matrix* M is defined as the function $R_M(r)$, which for a given r gives the minimum number of entries of M which one has to change in order to reduce its rank to r or less.

The goal is to prove lower bounds on the rigidity of matrices which would imply nonlinear lower bounds on some algebraic circuits. This research goes in the direction proposed by Valiant [22], who defined the concept of rigidity of matrices and proved that lower bounds on the size of logarithmic depth circuits can be proved by constructing matrices with high rigidity. Valiant proved the following result.

Theorem 1 ([22]) *If for some $\varepsilon > 0$, the $n \times n$ matrix M_n has rigidity $R_{M_n}(\varepsilon n) \geq n^{1+\varepsilon}$, then the transformation $x \to M_n x$ cannot be computed by linear size and logarithmic depth circuits with gates computing linear functions over a given field.*

Although both a random matrix and a matrix whose entries are different indeterminates have rigidity even larger than required by Theorem 1 (close to n^2), very little is known about *explicit* matrices. The best known lower bounds on the rigidity of explicit matrices are of the form $\Omega(\frac{n^2}{r}\log\frac{n}{r})$, which gives only linear lower bounds on $R_M(\varepsilon n)$. It seems that Hadamard matrices have large rigidity over the real field, but the best bound is so far only $\Omega(n^2/r)$.

In [5], the authors make advances towards a better understanding of the notion of rigidity by analyzing some structural proprties of low rank matrices.

Let us call an *alternating cycle* an oriented graph which is a cycle, when the orientation is forgotten, and such that the orientation of the arcs on the cycle alternates with one exception (put otherwise, there is a vertex v on the cycle such that if we go around the cycle from v to v, the orientation of the edges alternates).

Let $A = (a_{ij})$ be an $n \times n$ matrix. The graph of nonzero entries of A is the directed graph with vertex set $\{1, \ldots, n\}$, where (i, j) is an arc iff $a_{ij} \neq 0$.

Given a matrix, we call $[2, 2]$ *configuration* a 2×2 submatrix consisting of nonzero elements. In graph theoretical terms a $[2, 2]$ configuration corresponds to either an alternating 3-cycle, which is usually called a transitive triangle, (when the 2×2 submatrix has one entry on the main diagonal), or an alternating 4-cycle (when none of the entries of the 2×2 submatrix is on the main diagonal), or a 2-cycle (when the 2×2 submatrix has two entries on the main diagonal).

In [5] the authors are interested in odd alternating cycles as subgraphs of the graph of nonzero entries of low rank matrices, because of a connection to matrix rigidity. The reason for considering *odd* lengths is that an odd alternating cycle corresponds to a configuration in the matrix where one element is on the main diagonal, while an even alternating cycle is not connected to it. In particular, the following conjecture implies large rigidity, hence nonlinear lower bounds on the size of some circuits, for certain explicitly defined matrices. Note that we have to use some nontriviality assumption, such as having nonzero elements on the main diagonal, in order to get any interesting implication from the low rank assumption.

Conjecture 1 (The Odd Alternating Cycle Conjecture) *For every field F, there exist an odd k and $\varepsilon > 0$ such that every $n \times n$ matrix M with nonzero entries on the main diagonal, and such that rank$(M) \leq \varepsilon n$, contains an alternating cycle of length k.*

These results, and the results in [17] as well, show interesting connections between the above issue and the Ramsey number $R(3, n)$.

Other results, which show the role played by the determinant in the complexity of linear forms computation can be found in [18].

1.3 Rank and Communication Complexity

In complexity theory the most famous instance of the problem of relating the rank of a matrix to its structural properties given by the pattern of its nonzero entries is the relation between the communication complexity of a $\{0, 1\}$ matrix and its rank over the field of reals [16].

The increasing importance of networking, telecommunication, and distributed computing has pointed out the significance of communication as a computational resource. In addition, communication plays a central role in theoretical studies: many lower bounds in complexity theory have been obtained by looking at the communication between different parts of a given computational task.

We consider here a very simple model of communication, consisting of two processors (say A and B) connected by a direct link, each of them receiving its own input. Their goal is to compute a value (for simplicity we might assume this to be just one bit) which is a function of both inputs. We assume that both processors have unlimited computational power and that local computation is free, whereas we charge a unit of cost for each bit transmitted from a processor to the other one. The goal is to minimize the overall number of bits transmitted.

The central notion in communication is that of a *protocol*, which is essentially a set of rules specifying the order and the meaning of the messages sent (see for example [15] for more details). Associated to each protocol, there is its complexity, i.e., the number of bits transmitted in the worst case. A protocol terminates when one of the two processors, say A, knows the "answer", and the other one knows that this is the case.

There is always the so called *trivial protocol* which consists of A sending its entire input to B, so that the real challenge is to find, whenever possible, better protocols, in particular an optimal one.

A formalization of the above model can be done in terms of a matrix, called *communication matrix*, whose rows and columns are indexed by the input variables associated with processor A and B, respectively. Thus, if A and B have n and m possible inputs, respectively, we can define the $n \times m$ communication matrix $M \equiv (m_{ij})$, where the entry m_{ij} contains the value to be determined when A (resp. B) receives the i-th (resp. j-th) of its possible inputs.

A protocol has a very simple interpretation in the matrix setting. First of all, it determines the processor which sends the first message, say processor A. The input of A determines the first bit of information to be sent to B, and thus the protocol partitions the rows of M into two classes, where the bit transmitted by A tells B which of the two classes contains the row associated to the specific input received by A. After this, the "game" is restricted to the submatrix M_1 of M corresponding to the rows belonging to the suitable class. The next bit communicated further partitions M_1 into two classes, giving rise to a submatrix M_2, and so on. If the protocol consists of the transmission of k bits, then the matrix M_k must be the union of a submatrix of all ones and a submatrix of all zeros (a matrix with this property is called *almost homogeneous*). Indeed A knows the output bit when looking at M_k if its row in M_k has constant entries; furthermore this must be true for all rows, otherwise B would not know that the protocol ended.

A slight variation of the above definition of protocol consists of ending it in the presence of a monochromatic (either all zeros or all ones) matrix, instead of an almost homogeneous one.

It is immediate to notice that a protocol can be analyzed in terms of ranks of the submatrices detected as the rounds proceed; at each step, the maximum rank of submatrices decreases by at most a factor of 2, so that one obtains the following easy but important result.

Theorem 2 *[15] Let $c(M)$ be the communication complexity associated to the $\{0,1\}$ matrix M, i.e., the minimum number of bits that must be transmitted in any protocol associated to M. Then $c(M) \geq \log_2(rank(M))$.*

¿From Theorem 2 we can immediately derive the following corollary.

Corollary 3 *If M has full rank, then the trivial protocol, which consists of transmitting a row or column index, is optimal.*

It is worthwhile to mention that the rank of the communication matrix is an upper bound on communication complexity. In fact, it is not difficult to prove the following theorem.

Theorem 4 *$c(M) \leq rank(M)$.*

Theorems 2 and 4 show that communication complexity can be bounded in terms of rank, and has focussed much of the current research on understanding whether the upper bound can be sharpened. In particular, the following conjecture has been raised.

Conjecture 2 (See [14, 19]) *$c(M) \leq [\log_2(rank(M))]^c$, for a positive constant c.*

Conjecture 2 is closely related with the accuracy of bounds on the chromatic number of a graph obtained in terms of the rank of its adjacency matrix. Indeed, Lovázs and Saks proved in [14] that Conjecture 2 is true if and only if there exists a constant c such that the chromatic number of any graph G does not exceed $\exp(\log^c(rank(A_G)))$, where A_G denotes the adjacency matrix of G.

Nontrivial communication protocols can thus only arise for low rank matrices. In these cases, the truth of Conjecture 2 would intuitively imply that a protocol could end quickly by taking advantage of certain general properties of submatrices of low rank matrices. Indeed one of these properties is the existence, in any low rank matrix, of a very large monochromatic submatrix (see [6] for a review of these properties and related questions) .

1.4 Random walks on graphs, eigenvalues and expanders

We consider here random walks on graphs. They have a lot of applications:

- approximation/estimation problems: estimating permanent, volumes;

- random selection;
- derandomizing randomized algorithms.

A random walk on a graph can be defined as follows: Start at some initial vertex. Pick a neighbor at random and go there, etc. Equivalently, one can think of a random walk as currently being on some edge heading in some direction. Then move to a random neighboring edge in forward direction.

Some interesting quantities to look at are the following:

- H_{uv}, hitting time from u to v: the expected number of steps to reach v starting at u;
- C_{uv}, commute time: the expected number of steps to go from u to v, and then back to u;
- C_u, cover time from u: the expected time to visit the entire graph when starting at u.

A random walk on a graph is a special case of random walk on a Markov Chain. From this fact, one can derive a connection with eigenvalues.

The main issues behind the theory of random walks on graphs are surveyed in [13], where it is also analyzed the role played by graph eigenvalues in the investigation of the above, as well as of other related quantities.

2 A guided tour of the references

For each of the subjects there is a master reference, which is a review paper that should guide the study, and give an orientation for the analysis of the other papers.

Concerning Shannon capacity and theta function, Donald Knuth has written a very nice and clear review [9] which is actually self-contained and complete at the point that the other references ([21, 11]) have been included mostly for completeness and for providing the students with the original papers by Shannon and Lovász.

Matrix rigidity has been introduced in [22]. A survey of both results and open questions is [7]. References [5, 17, 2] contain results on properties of low rank matrices with applications to matrix rigidity and show the connections between such notion and the explicit construction of Ramsey graphs.

Communication complexity and the corresponding main open questions is reviewed in [12]. A special attention to the underlying algebraic issues is given in [6]. The original paper on the issue is [15], while the exploration of the connections between rank, communication complexity by using linear algebra tools can be found in [16].

The main aspects of the theory of random walks on graphs are surveyed in [13], where it is also described the connection with graph eigenvalues and its practical use.

References

[1] N. Alon, *Explicit Ramsey graphs and orthonormal labelings*, Electronic Journal of Combinatorics 1 (1994), R12.

[2] N. Alon, P. Pudlak: Constructive lower bounds for off-diagonal Ramsey numbers, submitted.

[3] N. Alon, M. Szegedy, *Large sets of nearly orthogonal vectors*, preprint.

[4] V. Brimkov, B. Codenotti, V. Crespi, M.Leoncini, "On the Lovász number of certain circulant graphs", Proc. CIAC 2000 (Rome, 2000).

[5] B. Codenotti, P. Pudlák, G. Resta "Some structural properties of low rank matrices related to computational complexity", Theoretical Computer Science, Vol: 235 (1) (2000), pp. 89-107.

[6] B.Codenotti, G. Del Corso, G. Manzini, "Matrix Rank and Communication Complexity", Linear Algebra and its Applications 304 (1-3) (2000), 193–200.

[7] B.Codenotti, "Matrix Rigidity", Linear Algebra and its Applications, 304 (1-3) (2000), 181–192.

[8] P. Erdös and R. Rado, *Intersection theorems for systems of sets,* J. London Math. Soc. 35, 1960, 85-90.

[9] D. E. KNUTH. *The Sandwich Theorem,* Electronic J. Combinatorics, Vol. 1, 1994.

[10] A. Kotlov, L. Lovász, *The rank and size of graphs*, J. of Graph Theory 23(1) (1996), 185-189.

[11] L. LOVÁSZ. *On the Shannon capacity of a graph,* IEEE Transactions on Information Theory, Vol. 25, pp. 1-7, 1979.

[12] L. LOVÁSZ. *Communication complexity: a survey,* in: *Paths, flows, and VLSI-Layout*, (ed. B. Korte, L. Lovász, H. J. Prömel, A. Schrijver), Springer, 235–265.

[13] L. LOVÁSZ. *Random Walks on Graphs: A Survey*, Combinatorics, Paul Erdös is Eighty, Bolyai Society Mathematical Studies (Volume 2) 1993, pp.1-46.

[14] L. Lovász and M. Saks, *Lattices, Möbius functions and communication complexity*, Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science (1988), 81–90.

[15] K.Mehlhorn and E.M. Schmidt, *Las Vegas is better than determinism in VLSI and distributed computing*, Proceedings 14th Annual ACM Symposium on the Theory of Computing (1982), 330–337.

[16] N. Nisan and A. Wigderson, *On Rank vs Communication Complexity*, Proc. 35th IEEE FOCS (1994), 831-836.

[17] P. Pudlak: Cycles of nonzero elements in low rank matrices, submitted.

[18] P. Pudlák. *A Note on the Use of Determinant for Proving Lower Bounds on the Size of Linear Circuits,* Electronic Colloquium on Computational Complexity (1998) TR98-042.

[19] R. Raz and B. Spiker, *On the log-rank conjecture in communication complexity*, Combinatorica, Vol. 15, 1995. (Preliminary version in Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science (1993), 168–176).

[20] M. Rosenfeld, *Almost orthogonal lines in* E^d, DIMACS Series in Discrete Math. 4 (1991), 489-492.

[21] C.E. Shannon. *The zero-error capacity of a noisy channel,* IRE Trans. Inform. Theory, vol. IT-2 (3), pp. 8–19 (1956).

[22] L.G. Valiant, *Graph-theoretic arguments in low level complexity,* Proc. 6th MFCS, Springer-Verlag LNCS 53 (1977), 162-176.