Dottorato di Ricerca in Informatica - VIII Ciclo
Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Università degli Studi di Torino

Thesis of Valentino Crespi

Structural and Computational Properties of Certain Permanents

Advisors:

prof. Bruno Codenotti
Istituto di Matematica Computazionale – CNR di Pisa

prof. Giancarlo Mauri
Dipartimento di Scienze dell'Informazione – Università degli Studi di Milano

Acknowledgements

I wish to thank Bruno Codenotti for having been an important human and professional guide and a real reference point to carry out this research work with enthusiasm. He has been a source of profound ideas and offered me the opportunity to further grow up as a researcher. Special thanks also to Giovanni Resta whose contributions to this thesis have been crucial in terms of intuitions and technical skills and to Giancarlo Mauri for the continuous support and for his enlightening suggestions.

I am grateful to Cecilia Coletti for her love. I am indebted with Marco Pellegrini who helped me thousands of times in critical moments during my stay in Pisa.

I am also thankful to Gianna del Corso, Anna Bernasconi, Luciano Margara, Giovanni Manzini and Mauro Leoncini for having had enough patience to stand me. Finally, I wish to remember Nicoletta Calamita for the moral support and Filippo Basso for the beautiful discussions about algebraic methods in computer science.

Contents

Chapter 1

Introduction

1.1 History of the Problem

The origin of the function *permanent* goes back to the the beginning of the last century. According to Minc [Mi78], it seems that the term “permanent” was coined by Cauchy himself who first introduced the homonymous function, almost simultaneously to Binet [Bi812], in his *Memoires* on determinants [Ca812], appeared in 1812. Binet [Bi812] gave also formulas for computing the permanents of $m \times n$ matrices, for $m \leq 4$. Muir [Mu882] is probably responsible for the final adoption of the name *permanents* to address the “fonctions symétriques permanentes” of Cauchy.

Let $A = (a_{i,j})$ be an $m \times n$ matrix over any commutative ring, $m \leq n$. The *permanent* of A, written $\mathrm{Per}(A)$, is defined by

$$\mathrm{Per}(A) = \sum_{\sigma} \prod_{i=1}^{m} a_{i,\sigma(i)},$$

where the summation extends over all one-to-one functions from $\{1, 2, \ldots, m\}$ to $\{1, 2, \ldots, n\}$. The sequence $(a_{1,\sigma(1)}, a_{2,\sigma(2)}, \ldots, a_{m,\sigma(m)})$ is called a *diagonal* of A, and the product $a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \ldots a_{m,\sigma(m)}$ is a *diagonal product* of A. We will deal only with square matrices $(m = n)$. Minc [Mi78] denotes the permanent of a square matrix A with $\mathrm{per}(A)$ instead of $\mathrm{Per}(A)$ and so do we here too.

Looking at the above definition we notice that the “only” difference with respect to the determinant is that the sign with which the diagonal products are taken is no longer alternating but is “permanently” positive. This fact probably motivates the choice of the proposed denomination. We will see that this “only” difference makes the computation of the permanent very hard compared to that of the determinant.

After Binet and Borchardt [Bo855], Cayley [Ca859] and Muir [Mu882],[Mu] worked to establish identities involving determinants and permanents. In particular they gave formulas to express the product $\mathrm{per}(A) \cdot \det(A)$ under certain conditions on the entries of A (see [Mi78] and [Mi83] for a complete and detailed survey on the subject). Some of their results, later perfectioned and generalized by Levine [Le59] and by Carlitz & Levine [CL60], were based upon the following well-known identities. Let $A = (a_{i,j})$ be an n-square matrix. Then

$$\begin{aligned}
\mathrm{per}(A)\cdot\det(A) &= \left(\sum_{\sigma\in E}\prod_{i=1}^{n} a_{i,\sigma(i)}\right)^2 - \left(\sum_{\sigma\in F}\prod_{i=1}^{n} a_{i,\sigma(i)}\right)^2 \\
&= \sum_{\sigma\in E}\prod_{i=1}^{n} a_{i,\sigma(i)}^2 - \sum_{\sigma\in F}\prod_{i=1}^{n} a_{i,\sigma(i)}^2 + f(A) \\
&= \det(A^{(2)}) + f(A),
\end{aligned}$$

where E and F are the sets of even and odd permutations, respectively, whereas, $A^{(2)} = A * A$ is the matrix whose (i,j) entry is $a_{i,j}^2$, and $f(A)$ represents the remaining terms. That was basically all for the nineteenth century.

A significant further step was made by Muirhead [Mu03] who stated an important result that successively Hardy, Littlewood and Pólya [HLP34] extended to determine under which conditions, given two n-tuples of nonnegative integers α and β, there exists a *doubly stochastic* $n \times n$ matrix S such that $\alpha = S \cdot \beta$.

An important direction started by Pólya [Po13] concerned with the attempt of reducing permanents to determinants. In practice, it was addressed the problem of finding special transformations that would convert permanents into determinants. We have devoted an entire Section on the convertibility theory (see Section 1.4.2). For now, we can formalize the problem as follows. Given S, set of $n \times n$ matrices, find a linear transformation T on S such that $\mathrm{per}(T(A)) = \det(A)$. In particular Pólya [Po13] restricted his attention on transformations that involve only a uniform affixing of a plus or minus sign to each position in the matrix and proved a negative result for them. Later Marcus and Minc [MM61] generalized it showing that for $n \geq 3$ there exists no linear transformation T on the set of $n \times n$ matrices such that $\mathrm{per}(T(A)) = \det(A)$ (see also [MM62],[Bo67],[Bo68], [Gi71] and [Mi76] for related results). Although any hope to convert the whole class of $n \times n$ matrices was disillusioned by this last theorem, others could obtain positive results on certain restricted subclasses S of $n \times n$ matrices. As we said before we will come back to this topic in Section 1.4.2.

An entirely new approach to permanents and determinants was initiated in 1918 by Schur [Sc18]. He introduced the concept of *generalized functions* on square matrices which led him to prove that $\det(A) \leq \mathrm{per}(A)$ for A positive semi-definite hermitian, where the equality holds if and only if A is diagonal or A has a zero row. Very important is also the subsequent introduction of the methods of multilinear algebra to the study of Schur functions as in [MN62],[Ma64] and in [MM64].

About ten years later, van der Waerden [Wa26] opened a problem later to be known as the van der Waerden conjecture. Let J_n denote the matrix whose entries are all ones and let Ω_n be the set of doubly stochastic $n \times n$ matrices (that is, nonnegative matrices with row sums and column sums equal to 1). Thus it was conjectured that if $S \in \Omega_n, S \neq J_n$ then $\mathrm{per}(S) > \mathrm{per}(J_n) = n!/n^n$. For a solution people had to wait for several decades. In 1980 Egoryĉev [Eg80] finally closed the question giving a proof which starts from the Alexandrov inequality [Al38] (see [Mi83] for a complete explanation of the Egoryĉev proof, see also [MN59] and [KS82] for related results).

The research made in the context of the van der Waerden conjecture and of doubly stochastic matrices has led to new more general conjectures involving the sum of all the subpermanents of a square matrix (notably the Tverberg conjecture [Tv63] proved by Friedland [Fr82], the Doković conjecture [Do67],[Ho64] and the Marcus and Minc conjecture [MM67]), and to the establishment of interesting inequalities and bounds for the permanent function especially when applied to $(0,1)$-matrices. Also to this last matter and to permanents of $(0,1)$-matrices we have dedicated ample space in this first chapter (see Section 1.6 for details and references).

About doubly stochastic matrices we mention one last question asked by Wang [Wa78] (see also [Wa79]). Two distinct matrices $A, B \in \Omega_n$ are said to form a *permanental pair* if $\mathrm{per}(\theta A + (1-\theta)B)$ is constant for $\theta \in [0,1]$. Wang [Wa78], Brenner and Wang [BW79], and lastly Gibson [Gi80] investigated properties of permanental pairs and permanental polytopes. Finally Brualdi and Newman [BN65] proved that $\mathrm{per}(\theta I_n + (1-\theta)A) \leq \theta + (1-\theta)\mathrm{per}(A)$ for all $A \in \Omega_n$ and $\theta \in [0,1]$. Furthermore for a given $B \in \Omega_n$ they showed that $\mathrm{per}(\theta B + (1-\theta)A) \leq \theta\mathrm{per}(B) + (1-\theta)\mathrm{per}(A)$ for all $A \in \Omega_n$ if and only if $\sum_{i,j=1}^{n} b_{i,j}\mathrm{per}(A(i \mid j)) \leq \mathrm{per}(B) + (n-1)\mathrm{per}(A)$ for all $A \in \Omega_n, A \neq B$.

Since the seventies, the permanent function has given rise to much interest among the computer science community particularly in the area of complexity and algorithms because of the work of Valiant [V179], [V279]. He introduced a new class of computational problems, named #P, the class of counting problems solvable by polynomial time counting Turing machines and proved that the permanent function even when restricted to $(0,1)$-matrices is complete for that class. This means that computing the permanent of a $(0,1)$-matrix A, i.e., counting the number of perfect matchings of the bipartite graph associated with A, is as hard as counting the number of satisfying assignments to a CNF formula, or the number of accepting computations of a polynomial-time-bounded nondeterministic Turing machine. He, basically, tried to lift the P versus NP question to an algebraic platform with the perspective of reaching separation results. Unfortunately he could not go very far and the question P versus NP as well as P versus #P remains the main conjecture of this discipline. This "lifting" has also shown a singular asymmetry. Finding a perfect matching in a bipartite graph is in P and deciding the satisfiability to a CNF formula is NP-complete, whereas the respective counting counterparts have the same difficulty, since they are both #P-complete. We will say more about that in Section 1.3.

In recent times, there have been several attempts to achieve efficient approximated algorithms using randomization techniques. In this framework it is worth mentioning the two main streams (see Section 1.5.2 for a brief overview). The first one is represented by Broder [Br86], Mihail [Mi89] and Jerrum & Sinclair [JS89] who have employed the *rapidly mixing Markov chains* technique for sampling from a uniform distribution. The second one is represented by Karmarkar, Karp, Lipton, Lovász, and Luby [KKLLL93] who have devised an algorithm based upon an unbiased estimator for $\mathrm{per}(A)$ (see also [FJ95] for a further analysis of its performance).

1.2 Applications

1.2.1 The permanent as a counting function

In describing the various employments and uses of permanents we had better start from counting problems in general.

First, recall from [GJ79] that a search problem Π is characterized by a collection of instances $D_\Pi = D$ together with a function $S_\Pi[\cdot] = S[\cdot]$ that associates with each instance a set of *solutions*.

Solving a search problem means providing an algorithm that takes as input an instance $I \in D$ and outputs a solution $J \in S[I]$. Analogously a counting problem is characterized by the couple (D, S) but this time we ask for an algorithm that, given I, computes the cardinality of the set of solutions, i.e., $|\, S[I] \,|$, whereas in a decision problem, the solving algorithm is required to decide whether or not the set $S[I]$ is empty.

It is clear that if we consider the couple $\Pi = (D, S)$ then, given I, the problem of counting the number of solutions of I is at least as hard as deciding on the existence of at least one such a solution.

A particular class of counting problems are those related to the decision problems solvable by polynomial time nondeterministic Turing machines. This last class is denoted with NP and one of its possible characterizations is the following. A decision problem Π is in NP if and only if the language of its Yes-instances $L_\Pi = \{x \in D_\Pi \mid S[x] \neq \emptyset\}$ is verifiable in polynomial time, i.e., there exists a polynomial time computable relation $R(\cdot, \cdot)$ and a polynomial p such that

$$x \in L_\Pi \text{ if and only if } \exists y.(|\, y \,| \leq p(|\, x \,|) \wedge R(x, y)).$$

In other words for each yes-instance x there must exist a *proof* or better a solution y which certifies the membership of x in L_Π.

In the counting version of the problem Π we are interested in the number of all such *proves* y rather than in the existence of at least one of them.

The important fact is that if Π is in NP then the maximum possible number of solutions is $O(2^{p(n)})$, where n is the size of the input and p is a polynomial. Thus the output of the associated counting problem has length at most polynomial in the input size and it makes sense questioning about the complexity of computing that number.

Valiant [V179] introduced the class #P of the counting problems which derive from the decision problems in NP. Formally, let R be a polynomially balanced, polynomial time decidable binary relation. The counting problem associated with R is the following: given x, determine how many y there are such that $(x, y) \in R$. The required output is an integer (in binary). #P is the class of all counting problems associated with polynomially balanced, polynomial time decidable binary relations. Thus #P will be also the class of counting problems solvable by polynomial time counting nondeterministic Turing machines.

The permanent problem can be interpreted as a counting problem of the class #P. In fact, consider an $n \times n$ $(0,1)$-matrix $A = (a_{i,j})$, then $\mathrm{per}(A)$ is exactly the number of perfect matchings of the bipartite graph $G = (U, V, E)$, where $U = V = [n]$ and $e = (i, j) \in E$ if and only if $a_{i,j} = 1$. Recall that a matching in a bipartite graph G is a subset of its edges which don't share any node and a matching is called perfect if its cardinality is n, i.e., if it *marries* the set U to the set V.

If we consider $D(A)$ the digraph whose adjacency matrix is A then $\mathrm{per}(A)$ is the number of cycle covers of $D(A)$. Recall that a cycle cover C of a digraph D is a set of vertex-disjoint cycles that altogether cover the whole digraph D.

Since, given a bipartite graph G and a subset of edges M it takes polynomial time in the size of G to check whether or not M is a perfect matching of G, it follows by definition that per is in #P. (Note that the problem of finding a perfect matching in a bipartite graph is even in P.)

What renders the permanent particularly important in complexity theory and also for the applications is the remarkable result due to Valiant [V179] stating that each problem in #P can be reduced to the permanent problem. In other words, per is complete for the class #P under polynomial time many one reductions which *preserve* the number of solutions. Those reductions are said *parsimonious.*

The above discussion about NP should not mislead. In fact, as noticed before, we can find decision problems in P⊆NP whose counting versions are even #P-complete as the celebrated *matching problem.* Nevertheless there are other examples of decision problems in P whose counting versions are in #P by definition, but remain solvable in polynomial time like counting the number of distinct *spanning trees* in an input graph [HP73].

We can find several examples of interesting uses of counting problems and so of permanents in different fields and specifically in the areas of Combinatorics, Probability and Statistical Physics.

1.2.2 Combinatorics

Derangements. Consider the following question. In how many ways can a dance be arranged for n married couples, so that no husband dances with his own wife ? In other words we ask for the number of permutations of n elements that fix no element. Those are called *derangements* of n elements. This number denoted with D_n is equal to the permanent of $J_n - I_n$, where J_n is the $n \times n$ matrix whose entries are all ones. Thus

$$D_n = \mathrm{per}(J_n - I_n) = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!}\right).$$

Ménage numbers. These numbers come from the so called "*problème des ménages*". The question is: In how many ways can n couples be placed at a round table so that men and women sit in alternate places and no husband sits on either side of his wife ? Let the wives be seated in alternate places. For each such seating the husbands can be arranged in

$$U_n = \mathrm{per}(J_n - I_n - P_n) = \mathrm{per}\left(\sum_{i=2}^{n-1} P^i\right)$$

ways, where P_n is the usual permutation matrix with one in positions $(1,2),(2,3),\ldots,(n-1,n),(n,1)$. The numbers U_n are called *ménage numbers* (see Touchard [To34] for a closed formula of U_n).

Latin squares. A classical combinatorial problem is the enumeration of the Latin squares. Let S be a finite set with n elements. A *Latin rectangle* based on S is an $r \times s$ matrix $A = (a_{i,j})$ with the property that each row and each column of A contain distinct elements of S. The number r of rows and the number s of columns of the rectangle A satisfies $r \leq n$ and $s \leq n$. Usually the set S is chosen to be the set $[n] = \{1, 2, \ldots, n\}$. If $s = n$ then each row of A contains a permutation of S and these permutations have the property that no column contains a repeated element. An $n \times n$ rectangle based on the n-set S is a *Latin square* of order n. Thus in a Latin square each row and each column contains a permutation of S. Typical examples of Latin squares are given by Cayley tables of groups. In fact, let G be a group of order n whose set of elements in some order is $a_1, a_2, \dots, a_n$ and the binary operation of G be denoted by $*$. A Cayley table of G is the matrix $A = (a_{i,j})$ of order n in which $a_{i,j} = a_i * a_j$ for all $1 \le i, j \le n$. The axioms of a group imply that A is a Latin square of order n based on the set $S = [n]$. Notice that not every Latin square is a Cayley table of a group because the axiom of associativity imposes a further restriction on a Cayley table.

One of the major unsolved problems in the theory of Latin squares is the determination of the number $L(n)$ of Latin squares of order n based on the set $S = [n]$, and more generally the number $L(r, n)$ of distinct r by n Latin rectangles based on S.

Formulas for $L(n, n)$ for $n = 1, 2, \dots, 9$ are given in [R63]. For example, $L(1, n) = 1, L(2, n) = n! \cdot D_n$ where D_n is the number of derangements of S. The formula of Riordan [Ri45]

$$L(3, n) = n! \cdot \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{k} D_{n-k} D_k U_{n-2k}$$

gives the number of $3 \times n$ Latin rectangles in terms of binomial coefficients, derangements and ménage numbers U_{n-k}.

In general, let Λ_n^k be the class of $(0,1)$-matrices having exactly k nonzeros per row and column and let $m(k, n)$ and $M(k, n)$ denote any lower and upper bound, respectively, for the permanent function in Λ_n^k. Then it is possible to derive the following bound for $L(r, n)$ [Mi78]:

$$n! D_n \prod_{t=2}^{r-1} m(n-t, n) \le L(r, n) \le n! D_n \prod_{t=2}^{r-1} M(n-t, n).$$

1.2.3 Probability

Harper model. The following interpretation was suggested by L.J. Harper (see [Wi68]). Suppose we have n boxes each of which contains exactly one ball. We denote with $a_{i,j}$ the probability that the ball in the box i moves to the box j. Notice that the matrix $A = (a_{i,j})$ is doubly stochastic. Furthermore, the probability that after a simultaneous transition of the balls there is still exactly one ball in each box is $\mathrm{per}(A)$.

Network reliability. This application of counting problems is due to C.J. Colbourn [Co87] (see also [KL85]). Suppose we are given an undirected graph $G = (V, E)$, with m edges. How many of the 2^m subgraphs of G contain a path from 1 to n? The portion of subgraphs that connect the two nodes is a precise estimate of the *reliability* of the graph, that is, 2^m times the probability that the two nodes will remain connected if all edges fail independently with probability $1/2$ each [Pa94].

Others. Some applications of permanents in probability theory were also developed in 1973 by Gyires [Gy73].

1.2.4 Statistical physics

Monomer-dimer systems. Here we describe the *monomer-dimer problem*. In a monomer-dimer system, the vertices of a finite undirected graph $G = (V, E)$ are covered by a non overlapping arrangement, or configuration of monomers (molecules occupying one site, or vertex of G) and dimers (molecules occupying two vertices that are neighbors in G). Typically G is a regular lattice in some fixed number of dimensions. Three-dimensional systems occur classically in the theory of mixtures of molecules of different sizes [Gu52] and in the cell-cluster theory of the liquid state [CBS55]; in two dimensions, the system is used to model the absorption of diatomic molecules on a crystal surface [Ro35] (see the seminal paper of Heilman and Lieb [HL72] for a deeper discussion).

We can identify monomer-dimer configurations with matchings in the graph G. Thus a matching of cardinality k, or a k-matching, corresponds to a monomer-dimer configuration with k dimers and $2(n-k)$ monomers, where $2n = |V|$ is the number of vertices in G. To each matching M a weight $w(M) = \lambda^{|M|}$ is assigned, where λ is a positive real parameter that reflects the contribution of a dimer to the energy of the system. The *partition function* of the system is defined as

$$Z(\lambda) = \sum_{M} w(M) = \sum_{k=0}^{n} m_k \lambda^k,$$

where m_k is the number of k-matchings in G. For a physical interpretation of the partition function see [HL72]. The partition function is a central quantity in statistical physics, and captures essentially everything one needs to know about the thermodynamics of the system, including quantities such as the free energy and the specific heat and the location of phase transitions.

Ising model. This model was introduced in the 1920s by Lenz and Ising as a means of understanding the phenomenon of ferromagnetism. An instance of the Ising model is specified by giving a set of n *sites*, a set of *interaction energies* $V_{i,j}$ for each unordered pair of sites i, j, a magnetic field intensity B, and an *inverse temperature* β. A *configuration* of the system defined by these parameters is one of the 2^n possible assignments σ of ± 1 *spins* to each site. The energy of a configuration σ is given by the *Hamiltonian* $H(\sigma)$, defined by

$$H(\sigma) = -\sum_{\{i,j\}} V_{i,j}\sigma_i\sigma_j - B\sum_{k} \sigma_k.$$

In realistic applications the sites are arranged in a regular fashion in 2- or 3-dimensional space and $V_{i,j}$ is non-zero only for adjacent sites.

A central problem in the theory is evaluating the *partition function* $Z = \sum_\sigma \exp(-\beta H(\sigma))$, where the sum is over all possible configurations σ. The significance of Z is that it is the normalizing factor in the *Gibbs* distribution, which assigns probability $\exp(-\beta H(\sigma))/Z$ to each state σ in the steady state. Other problems relate to the evaluation of the expectation of certain random variables of σ, when σ is sampled according to the Gibbs distribution, for instance, the *mean magnetic moment* and the *mean energy*.

In the *ferromagnetic* case, where $V_{i,j} \geq 0$ for all pairs $\{i, j\}$ of sites, the exact computation of the partition function Z is #P-complete [JS93].

Others. The use of permanents in physics was actually pioneered by Caianiello [Ca59] and [Ca59b] in connection with renormalization problems in quantum field theory. He used permanents and hafnians to express in algebraic form perturbation expansions of field theory for describing boson propagators, in the same way as determinants and pfaffians appear in the expansions related to fermion propagators.

1.3 The Complexity of Computing the Permanent

The computation of the permanent of a matrix is a challenging task. The problem is computationally very hard, even for $(0,1)$ matrices. In fact, Valiant proved that computing the permanent of a $(0,1)$ matrix is #P-complete (see [V179]). Thus it is extremely unlikely that there is a polynomial time algorithm for computing the permanent. Actually, the best known algorithm for computing the permanent is due to Ryser [R63] and takes $O(n\,2^n)$ operations, where n is the matrix size (this algorithm is described in Section 1.5.1).

Notice that the class #P is very powerful. Indeed Toda [To89] proved that the polynomial hierarchy PH is contained in $\mathrm{P}^{\#\mathrm{P}}$, which in turn is equal to P^{PP} [An80]. Recall that PP is the class of problems asking whether more than half of the computations of a nondeterministic machine are accepting and it is closely related to #P because given the number k of all the accepting computations of a polynomial time nondeterministic (standardized) Turing machine we can check out whether or not this number is greater than 2^{m-1}, where m is the number of steps in the computation, simply by inspecting the *first bit* of k. Clearly $\mathrm{P}^{\#\mathrm{P}} \subseteq$ PSPACE so it holds that

$$\mathrm{PH} \subseteq \mathrm{P}^{\#\mathrm{P}} = \mathrm{P}^{\mathrm{PP}} \subseteq \mathrm{PSPACE}.$$

More recently, several authors have found even stronger negative results concerning the efficiency of approximation schemes [DLMV88], as well as the hardness of computing the permanent of random matrices [FL92], and also of very sparse matrices [DLMV88]. The little hope to get efficient algorithms for matrices without a very special structure, motivated a stream of research work oriented towards analyzing the permanent of restricted classes of matrices or to develop computationally feasible approximation algorithms.

The notion of "computationally feasible approximation" algorithm can be formalized as follows. Let f be a function from input strings to natural numbers. A *fully randomized approximation scheme* (fpras) [KL83] for f is a probabilistic algorithm that takes as input a string x and a real number $0 < \epsilon < 1$, runs in time polynomial in $\mid x \mid$ and $1/\epsilon$ and outputs a number Y (a random variable) such that

$$Pr\left[\frac{f(x)}{1+\epsilon} \leq Y \leq (1+\epsilon)f(x)\right] \geq \frac{3}{4}.$$

The *confidence* probability can be boosted to $1-\delta$ for any desired $\delta > 0$ by running the algorithm $O(\log(1/\delta))$ times and taking the median of the results [JVV86].

The existence of an fpras for the unrestricted permanent, i.e., counting perfect matchings approximatively in general graphs, remains an open question although considered very unlikely. Indeed it would imply NP=RP [JVV86],[Si88]. Anyhow Jerrum and Sinclair [JS89] analyzed a probabilistic algorithm for counting the number of perfect matchings in a bipartite graph G first proposed by Broder [Br86] and proved that it runs in time polynomial in $|G|, 1/\epsilon$ and

$$t \geq \frac{|M_{n-1}(G)|}{|M_n(G)|}, \tag{1.1}$$

where M_k denotes the set of matchings composed of k edges. Thus for restricted classes of bipartite graphs and for instance for those satisfying $| M_{n-1}(G) | \ / \ | M_n(G) | < n^{O(1)}$ an fpras does exist. The problem with this approach is that we need to know in advance the value of ratio 1.1 which is hard to compute. Further investigations led to identify recognizable subclasses of matrices with polynomially bounded ratio. Let us see some examples. The *factor size* of G, f, is the maximum number of edge disjoint perfect matchings in G. Using network flow techniques, f can be computed from G in polynomial time. Dagum and Luby [DL92] showed that

$$\frac{|M_{n-1}(G)|}{|M_n(G)|} < n^{\frac{3n}{f}}.$$

From this theorem it follows that there is an fpras for the class of graphs with *large factors*, i.e., with $f \geq \alpha n$, for $\alpha > 0$. Jerrum and Sinclair [JS88] proved that the class of dense graphs (a graph G is dense if every vertex has degree at least $n/2$) has also a polynomially bounded ratio. There are also negative results. A class of bipartite graphs is said *approximation complete for the permanent* if an (ϵ, δ)-approximation algorithm (an algorithm that with probability greater than $1 - \delta$ outputs an estimate of the solution with relative error less than ϵ) for the class implies an (ϵ, δ)-approximation algorithm for all bipartite graphs. Let $\alpha < 1$ be any constant and let f be any function such that $3 \leq f(n) \leq n - 3$. Dagum and Luby [DL92] proved that the exact counting for $f(n)$-regular graphs is #P-complete and furthermore for $3 \leq f(n) \leq n^{1-\alpha}$ the class of $f(n)$-regular graphs is approximation complete for the permanent.

1.4 Permanent vs Determinant

1.4.1 Valiant

The determinant function has, among the others, a property of "universality" as Valiant showed in [V279]. To describe what he meant we need some formal definitions. Before proceeding, we point out that such a property will allow us to use the determinant for computing the permanent, even though it does not constitute an acceptable solution and the whole theory of p-completeness, started by Valiant himself, which was an attempt to move the P versus NP question from the boolean domain to the algebraic domain, brought him to crush on the wall of a new conjecture:

Valiant's hypothesis: *There exist no fast arithmetic algorithms for the $n \times n$ permanent using constants from the ground field, indeterminates, and $n^{O(1)}$ arithmetic operations $+, -, \times, /$.*

Let F be a field and $F[x_1, x_2, \ldots, x_n]$ the ring of the polynomials over the indeterminates $\{x_1, x_2, \ldots, x_n\}$ with coefficients from F. The set $\mathcal{F}$ of formulae over F is recursively defined as follows:

1. if c is a constant which denotes an element in F then
$$c \in \mathcal{F},$$
2. if x_j is an indeterminate then
$$x_j \in \mathcal{F},$$
3. if $f_1, f_2 \in \mathcal{F}$ then
$$f_1 + f_2 \in \mathcal{F}, f_1 \times f_2 \in \mathcal{F}.$$

The *size* of a formula f is the number of arithmetic operations $(+, \times)$ needed in its construction and is denoted with $|f|$. Any formula specifies a polynomial, in fact given f we can distribute the product $\times$ over the sum $+$ as long as it is possible and eventually we get a polynomial. The formula size of a polynomial p is the size of the minimal size formula that specifies it, i.e.,

$$|p| = \min\{|f| : f \text{ specifies } p\}.$$

If X is a set of indeterminates and $A \subseteq F[Y]$ then the map $\sigma : X \to A$ is a *substitution*. If $p \in F[X]$ and $\sigma : X \to A \subseteq F[Y]$ is a substitution then we denote with $p\sigma$ the polynomial obtained by applying the substitution σ to p. Furthermore if $A = Y \cup F$ then the substitution is said simple.

Definition 1 *The polynomial $q \in F[Y]$ is a projection of the polynomial $p \in F[X]$ if there exists a simple substitution $\sigma : X \to Y \cup F$ such that $q = p\sigma$.*

Let Y be the $n \times n$ matrix of indeterminates $\{y_{i,j}\}$ for $1 \leq i, j \leq n$. Let $G = (V, E)$ the digraph whose adjacency matrix is Y, i.e., $G = D(Y)$. The determinant polynomial in the indeterminates $\{y_{i,j}\}$ is defined as

$$\det(Y) = \sum_{\pi} (-1)^{\mathrm{sign}(\pi)} \prod_{i=1}^{n} y_{i,\pi(i)},$$

where the summation extends over all the $n!$ permutations π on $\{1, 2, \ldots, n\}$. Now we are ready to state the "universality" property of the determinant.

Theorem 1 (see [V279]) *Let $p \in F[X]$. Then there exists a substitution $\sigma : Y \to X \cup F$[1] such that*

$$p = \det(Y)\sigma,$$

where $Y = (y_{i,j})$ is of order $|p| + 2$. In other words each polynomial of size k is a projection of the determinant polynomial of a $(k+2)$-square matrix of indeterminates.

[1] Here we have considered Y both as a matrix and as the set of its variables. This should not arise confusion.

We sketch the proof. Given p, let f be the formula that specifies p such that $|p|=|f|$. Then we can map f onto a digraph G of $l=|f|+2$ nodes whose edges are labeled with the indeterminates and the constants in f and with the important additional property:

$$p = \det(A(G)) = \sum_{cc} W(cc),$$

where $A(G)$ is the adjacency matrix of G, the summation extends over all cycle covers cc in G and the function W returns the total weight of a cycle cover, i.e., the product of the labels of its edges. If $Y = (y_{i,j})$ is an l-square matrix of indeterminates then from $A(G)$ we can derive a substitution $\sigma : Y \to \text{var}(p) \cup F$ such that $p = \det(A(G)) = \det(Y)\sigma$.

This result is very interesting because it says that for each polynomial p of size k there exists a $(k+2)$-square matrix A such that $p = \text{per}(A) = \det(A)$.

We can apply this to establish a relation between permanents and determinants. Indeed Ryser [R63] showed that the permanent polynomial associated with an n-square matrix of indeterminates can be specified by a formula f of size $|f| = n^2 \cdot 2^n$. Thus a direct corollary of the above theorem would imply that for each n-square matrix X there exists a square matrix A of order $l = O(n^2 \cdot 2^n)$ such that $\text{per}(X) = \det(A)$.

1.4.2 Convertible Matrices

Introduction and Background. The property we are going to investigate concerns the possibility of reducing the computation of the permanent of a $(0,1)$-matrix to the determinant of the $(1,-1,0)$-matrix of the same size obtained from the former by changing the sign to some of its elements. A matrix for which this property holds is said to be *convertible*. Two equivalent characterizations of the convertibility in terms of graphs are contained in the work of Little [Li75] and Seymour and Thomassen [ST87] (see also [Th86]). In the context of circulant $(0,1)$-matrices we also mention the remarkable work of Tinsley [Ti60]. Brualdi and Shader [BS91] (see also [BR91] and [BS95] for an exaustive explanation of this theory) have instead shown the equivalence between *sign-nonsingularity* and *convertibility*. They provided a series of results which led to the development of an algorithm avoiding backtracking that receives a $(0,1)$-matrix as input and computes its *conversion* whenever it exists or gets stuck and rejects otherwise. The conversion of a $(0,1)$-matrix is the $(1,-1,0)$-matrix advocated above. That algorithm can be exponential in the worst case whereas the complexity of deciding the convertibility of a square matrix remains unknown [KLM84]. Further investigations proved that the sign-nonsingularity recognition problem is NP-hard since it reduces to the EVEN problem for digraphs [Th86] which had already been proven to be NP-hard [LP84].

Let us formalize the notions of convertibility and sign-nonsingularity.

A $(0,1)$-matrix A is said to be *convertible* if there exists a $(-1,1)$-matrix X such that $\text{per}(A) = \det(A*X)$ where $*$ denotes the componentwise product operation. The $(1,-1,0)$-matrix $A*X$ is said to be the *conversion* of A.

A matrix B is said to be *sign-nonsingular* if each matrix Y whose elements have the same sign as those of B, i.e. the same sign pattern, is nonsingular. In other words B is sign-nonsingular if and only if

$$\forall Y \quad (\text{sign}(Y) = \text{sign}(B) \Longrightarrow \det(Y) \neq 0).$$

Convertibility and sign-nonsingularity. The sign-nonsingularity property is a crucial feature in the analysis of the interrelation between the determinant and the permanent. Indeed it is possible to prove that a $(1,-1,0)$-matrix B is sign-nonsingular if and only if

$$\mathrm{per}(|B|) = |\det(B)|,$$

where $|B|$ is the matrix whose entries are the absolute values of the entries of B.
In other words a $(0,1)$-matrix A turns out to be convertible if and only if there exists a $(1,-1)$-matrix X such that $A * X$ is sign-nonsingular.

Operations on graphs and matrices. Before giving the details of the two characterizations, we introduce some basic operations on graphs and matrices.

(a) *Splitting of a vertex in a directed graph*

The splitting of a vertex u is obtained by adding a new vertex v and a new edge (u,v) and then replacing each edge (u,y) with an edge (v,y).

(b) *Splitting of an edge in a directed graph*

The splitting of an edge (u,v) is obtained by the addition of a new vertex w and the replacement of the edge (u,v) with the two edges (u,w) and (w,v).

A *subdivision* of a digraph G is a digraph G' obtained from G by a sequence of edge splittings.

A *splitting* of a digraph G is a digraph G' obtained from G by a sequence of edge splittings and vertex splittings.

(c) *Contraction of a* $(0,1)$ *matrix*

Let $X = (x_{i,j})$ be a $(0,1)$-matrix of order n. We consider a row i and a column j such that in both of them there be exactly two nonzero entries. Let them be $x_{i,j}, x_{k,j}, x_{i,l}$. Furthermore, if we let $x_{k,l} = 0$, then the contraction gives a new $(0,1)$-matrix of order $n-1$ which differs from X in that row i and column j have been eliminated and the entry $x_{k,l}$ has been set to 1.

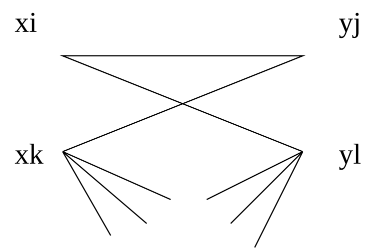
Example:

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Graph operations versus matrix operations. The contraction operation assumes different meanings depending on different possible graph interpretations. For instance we could look at either the $2n$ vertices bipartite graph or the $n \times n$ digraph associated with the matrix.

Let us clarify the relationships between operations on graphs and operations on matrices. Let us analyze the contraction of a matrix A looking first at what happens to the bipartite graphs associated respectively with A' and with its contraction A, and then to the digraphs.

	yj	yl
xi	1	1
xk	1	0



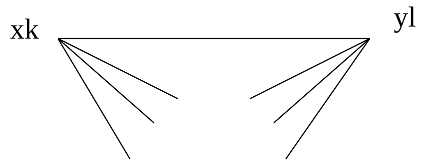


Figure 1.1: Vertices x_i and y_j are canceled and their three incident arcs are replaced by the arc (x_k, y_l)

(1) *Contraction and Bipartite Graphs*

In practice the contraction operation corresponds to two inverse splitting of edges in a bipartite graph.

A subdivision consisting of an even number of splittings is said to be even. So, the contraction operation is the inversion of an even subdivision.

(2) *Contraction and Digraphs*

Given a matrix $A = (a_{i,j})$ of order n, we define the digraph $D[A] =< V, E >$ as follows:

1. $V = \{1, 2, \ldots, n\}$;
2. $(i, j) \in E$ if and only if $i \neq j \wedge a_{i,j} \neq 0$.

$D[A]$ has no self-loops, furthermore the elements on the main diagonal don't affect the structure of the digraph. Similarly the weighted digraph of A is basically $D[A]$ together with a weighting of the edges given by the entries of A:

$$\forall i, j \quad i \neq j \Rightarrow w(i, j) = a_{i,j}.$$

Now, let us consider the following $(0, 1)$ matrix A together with its associated (in the above sense) directed graph $D[A]$:

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

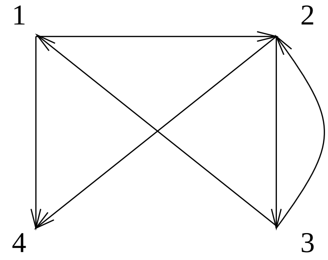


Figure 1.2: Digraph representation of the matrix A.

<u>Let us perform the splitting of edge $(3, 2)$ on $D[A]$:</u>

$$A' = \left[\begin{array}{cccc:c} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & \boxed{0} & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ \hdashline 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

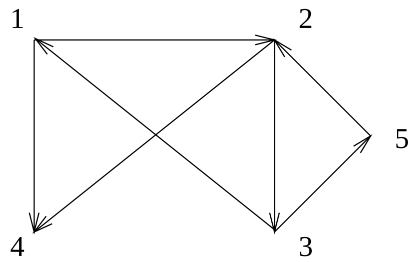


Figure 1.3: The new vertex gets the largest number (now 5)

It turns immediately out that A is obtained by contraction of A' on $a'_{5,5}$. It is very important to note that the contraction occurs on an element on the main diagonal. In fact splitting an edge *always* corresponds to contracting an element on the main diagonal.

<u>Let us perform the splitting of vertex 2 on $D[A]$:</u>

$$A' = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & \boxed{0} & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

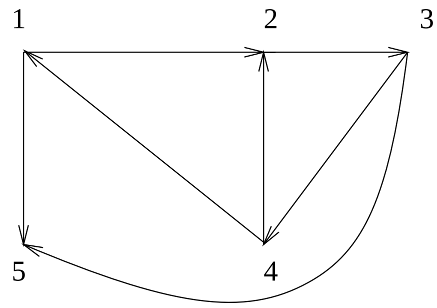


Figure 1.4: The vertices are renumbered so that the new vertex (now 3) follows the one from which it has been split (now 2)

Also in this case A can be obtained as a contraction of A'. This time the contraction is on the off diagonal element $a'_{2,3}$. Thus we have that the contraction is equivalent to the inverse operation of the splitting of either an edge or a vertex of the digraph depending on whether the nonzero entry on which the operation is performed is on or off the main diagonal.

The Characterization of Little. Little characterizes convertible matrices by working on the associated bipartite graphs.
His result can be expressed as follows:

Theorem 2 *Let A be an $n \times n$ $(0,1)$-matrix. Then A is not convertible if and only if the associated $2n$ vertices bipartite graph contains an even subdivision J of $K_{3,3}$ such that $G - V(J)$ has a perfect matching.*

□

Recalling the correspondence between even subdivisions and contractions we can restate the previous theorem in the following way:

Theorem 3 *An $n \times n$ $(0,1)$-matrix A is not convertible if and only if there exists a $(0,1)$-matrix $F \leq A$, which, after row and column permutations, can be put into the form $I \oplus Y$, where J_3 can be obtained from Y by a sequence of contractions.*

□

Kuratowski's theorem for planar graphs says that a graph G is planar if and only if it does not contain a subgraph which is isomorphic to a subdivision of $K_{3,3}$ or K_5. Thus Little's theorem implies the following

Corollary 4 *Let A be a square $(0,1)$-matrix and G be its associated bipartite graph. If A is not convertible then G is not planar.*

□

In fact if A is not convertible then G contains a spanning subgraph G' which is a subdivision of $K_{3,3}$ and so G is not planar.

The Characterization of Seymour and Thomassen. Unlike Little, Seymour and Thomassen characterize convertible matrices in terms of properties of the associated digraphs.

Given a weighted directed graph G, we define the *weight* of a directed cycle as the product of the weights of the edges belonging to it.

Let A be a sign-nonsingular $(1,-1,0)$-matrix. It is possible to show that by permuting rows and columns or by multiplying some rows by -1, we still have a sign-nonsingular matrix. So, without loss of generality, we can assume that all the entries in the main diagonal of A are equal to -1.
Bassett, Maybee and Quirk [BMQ68] have shown that a $(1,-1,0)$-matrix A whose main diagonal entries are all equal to -1 is sign-nonsingular if and only if the weight of each directed cycle of $D[A]$ is equal to -1.

It follows that if B is a $(0,1)$-matrix whose main diagonal entries are equal to 1, then B is convertible if and only if it is possible to assign to the edges of $D[B]$ weights $(-1,1)$ so that each of its directed cycle has weight -1.
A digraph D is even if each of its subdivisions contains at least a directed cycle of even length. This is equivalent to saying that D is even if for all the assignments of weights $(-1,1)$ to its edges there exists a directed cycle of weight 1. This can be proven considering that the subdivision of an edge (splitting of an edge) is equivalent to assigning a weight $(1,-1)$ to it. In particular, if we set $w(u,v) = -1$ if and only if (u,v) has been split an even number of times, we have the equivalence of the two definitions. Putting everything together, we obtain the following result:

Theorem 5 (see [BR91]) *Given a $(0,1)$-matrix B with the entries of main diagonal equal to 1, the following statements are equivalent:*

1. *B is convertible;*
2. *there exists a sign-nonsingular matrix A whose main diagonal elements are equal to -1 such that $|A|= B$;*
3. *there exists an assignment of weights $(-1,1)$ to the edges of $D[B]$ so that each directed cycle has weight -1;*
4. *$D[B]$ is not even.*

□

Thus B is convertible if and only if $D[B]$ is not even. We now show a characterization of even graphs by Seymour and Thomassen.

Let C_k^* the k-vertices directed graph obtained taking the undirected cycle of k vertices C_k and replacing each edge with a pair of opposite directed edges. It is possible to show that each splitting of C_k^* (k odd) is an even graph. The interesting result is that the structure of the splittings of C_k^* (k odd) completely captures the property of being even. Indeed we have the following:

Theorem 6 (see [ST87]) *A digraph G is even if and only if it contains a splitting of a C_k^*, for k odd.*

□

The equivalence between the two characterizations follows from the relationships between the contraction of matrices and the splitting of vertices and edges of directed graphs mentioned in the previous paragraph.

The Brualdi and Shader algorithm. Before we dealt with the problem of characterizing convertible matrices. Here we will be interested in finding an effective way to decide whether or not a given $(0,1)$-matrix is convertible and if so to compute the actual conversion matrix. The algorithm follows from two important results due to Brualdi and Shader [BS91] (see also [BS95]).

Let $A = (a_{i,j})$ be an $n \times n$ $(0,1)$-matrix. If $\sigma : \{1,\dots,n\} \to \{1,\dots,n\}$ is a permutation, then the set $\{a_{i,\sigma(i)} \mid i = 1,\dots,n\}$ is said to be a diagonal of A. The diagonal is nonzero if all of its elements are not equal to 0. The matrix A has *total support* if each of its nonzero entries belongs to at least a nonzero diagonal. This means that each nonzero entry contributes to the value of the permanent. Matrices with total support can be reduced to a special form by permuting rows and columns according to the following theorem.

Theorem 7 (see [BR91]) *Let A be a square $(0,1)$-matrix whose elements are not all zero. Then A has total support if and only if there exist two permutation matrices P, Q such that*

$$PAQ = B_1 \oplus B_2 \oplus \ldots \oplus B_t,$$

where B_i is fully indecomposable, $i = 1, 2, \dots, t$.

□

A matrix A is said to be partly decomposable if there exist two permutation matrices P, Q such that PAQ contains a $k \times (n-k)$ zero submatrix for $k \geq 1$. So A will be said to be fully indecomposable if it is not partly decomposable. Fully indecomposable matrices have an inductive structure up to rows and columns permutations.

Theorem 8 (see [BS91]) *Let A be a fully indecomposable $(0,1)$-matrix of order $n \geq 2$. There there exist permutation matrices P and Q of order n and an integer $m \geq 2$ such that PAQ has the form*

$$\begin{bmatrix} A_1 & O & O & \ldots & O & E_m \\ E_1 & A_2 & O & \ldots & O & O \\ O & E_2 & A_3 & \ldots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \ldots & A_{m-1} & O \\ O & O & O & \ldots & E_{m-1} & A_m \end{bmatrix}$$

where $A_1, A_2, \dots, A_m$ are fully indecomposable matrices and the matrices $E_1, E_2, \dots, E_m$ each contain at least one 1.

□

Also fully indecomposable matrices can be characterized within the set of matrices with total support by using bipartite graphs and digraphs.

Theorem 9 (see [BR91]) *Let A be a nonzero $(0,1)$-matrix of order n with total support, and let G be the bipartite graph whose reduced adjacency matrix is A. Then A is fully indecomposable if and only if G is connected.*

□

Theorem 10 (see [BR91]) *Let A be a nonzero $(0,1)$-matrix of order n whose main diagonal elements are equal to 1, and let $D[A]$ be the associated digraph (in which, as usual, we do not consider the self-loops). Then A is fully indecomposable if and only if $D[A]$ is strongly connected.*

□

In what follows we will recall three theorems by Brualdi and Shader and how those results can be used to actually compute the conversion of a $(0,1)$-matrix, whenever it exists.

Theorem 11 (see [BS91]) *Let A be a $(0,1)$-matrix with total support and let X and Y be sign-nonsingular matrices with $|X|=|Y|=A$. Then there exist diagonal matrices D_1 and D_2 of order n whose diagonal elements are equal to ± 1 such that $Y = D_1 X D_2$. If A is a fully indecomposable matrix, then D_1 and D_2 are unique up to a scalar factor of -1.*

□

Theorem 12 (see [BS91]) *Let A be a fully indecomposable $(0,1)$-matrix of order n, and let $X = (x_{i,j})$ be a sign-nonsingular matrix with $|X| = A$. Let B be the $(0,1)$-matrix obtained from A by replacing a 0 in position (u,v) with a 1. Then the following are equivalent:*

1. *There exists a sign-nonsingular matrix $Z = (z_{i,j})$ with $|Z| = B$.*
2. *There exists a sign-nonsingular matrix $\hat{Z} = (\hat{z}_{i,j})$ which can be obtained from X by changing x_{uv} to 1 or -1.*
3. *The matrix obtained from X by deleting row u and column v is a sign-nonsingular matrix.*

□

The next result extends the previous theorem to matrices with total support.

Theorem 13 (see [BS91]) *Let $A = (a_{i,j})$ be a $(0,1)$-matrix with total support and assume that $A = A_1 \oplus A_2 \oplus \ldots \oplus A_k \oplus A_{k+1}$ where the matrices $A_1, \ldots, A_k$ are fully indecomposable. Let $\hat{A} = \hat{A}_1 \oplus \hat{A}_2 \oplus \ldots \oplus \hat{A}_k \oplus \hat{A}_{k+1}$ be a sign-nonsingular matrix with $|\hat{A}| = A$. Let*

$$B = \begin{bmatrix} A_1 & O & O & \ldots & O & F_k \\ F_1 & A_2 & O & \ldots & O & O \\ O & F_2 & A_3 & \ldots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \ldots & A_{k-1} & O \\ O & O & O & \ldots & F_{k-1} & A_k \end{bmatrix} \oplus A_{k+1}$$

where the matrix F_i is a $(0,1)$-matrix with exactly one 1 and this 1 is in position (u_i, v_i) of F_i, $(i = 1, 2, \ldots, k)$. Then the following are equivalent:

1. *There exists a sign-nonsingular matrix $\tilde{B}$ with $\mid \tilde{B} \mid = B$.*

2. *There exists a sign-nonsingular matrix $\hat{B} = (\hat{b}_{i,j})$ with $\mid \hat{B} \mid = B$ such that $\hat{b}_{i,j} = \hat{a}_{i,j}$ for all (i,j) for which $a_{i,j} \neq 0$.*

3. *For $i = 1, 2, \ldots, k$ the matrix $\hat{A}'_i$ obtained from $\hat{A}_i$ by deleting row u_{i-1} and column v_i is a sign-nonsingular matrix (here we interpret u_0 as u_k).*

□

Furthermore if 1 of theorem 13 holds, then a conversion matrix satisfying 2 would be given by

$$\hat{B} = \begin{bmatrix} \hat{A}_1 & O & O & \ldots & O & \hat{F}_k \\ F_1 & \hat{A}_2 & O & \ldots & O & O \\ O & F_2 & \hat{A}_3 & \ldots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \ldots & \hat{A}_{k-1} & O \\ O & O & O & \ldots & F_{k-1} & \hat{A}_k \end{bmatrix} \oplus \hat{A}_{k+1}$$

for one of $\hat{F}_k = F_k$ or $\hat{F}_k = -F_k$.

Now let us see how to use theorems 12 and 13. Let B be a $(0,1)$-matrix of order n with total support and assume without loss of generality that $I_n \leq A$. Starting with the identity matrix I_n, we can obtain B by repeatedly applying the constructions of the matrices F in theorems 12 and 13 (in general after row and column permutations). Let $B_0 = I_n, B_1, B_2, \ldots, B_l = B$ be a sequence of matrices starting with $I = B_0$ and ending with $B_l = B$, where B_{i+1} is obtained from B_i via theorem 12 or theorem 13. Theorems 12 and 13 imply that given any conversion $\hat{B}_i$ of B_i there is a conversion of B_{i+1} if and only if there is a conversion $\hat{B}_{i+1}$ of B_{i+1} which extends the conversion $\hat{B}_i$. Starting with the conversion $\hat{B}_0 = I_n$ of B_0, we attempt to extend a conversion of B_i to a conversion of B_{i+1}. If at some point we are unable to do so then we are guaranteed that no conversion of B exists. Statement (3) of the two theorems gives us a way to check this. Furthermore whenever condition 3 is satisfied we know exactly which elements have to be changed in sign. Let us see an example. Suppose we are given the matrix

$$B = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

which has total support. The elements on the main diagonal are all equal to 1 so that $I_6 \leq B$. If this were not the case, the total support property will guarantee that one could force this condition, by permuting certain rows and columns. Let us start from $B_0 = I_6$ and first construct the sequence of matrices $\{B_i\}$. In a second phase, we will then compute the appropriate conversions.

$$I_6 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \oplus I_3.$$

Following the construction of the matrix F in theorem 13 we obtain

$$B_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \oplus I_3,$$

where $F_1 = F_2 = F_3 = [1]$. The first term of the direct sum is a fully indecomposable matrix. Proceeding with the construction of the matrix F in theorem 12 we get

$$B_2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \oplus I_3 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \oplus I_2,$$

where the element $(2,3)$ is set to 1. For the next step we will use the construction of the matrix F in theorem 13 to obtain

$$B_3 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \oplus I_2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where

$$F_1 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

and

$$F_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

One more application of theorem 13 will give

$$B_4 \equiv B = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix},$$

where

$$F_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, F_2 = [1], F_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

The first phase has been accomplished. We now turn to the second phase starting from the conversion of $B_0 = I_6$ which is I_6 itself. So $\hat{B}_0 = I_6, \hat{B}_1 = B_1$,

$$\hat{B}_2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \hat{B}_3 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now the problem consists of extending the conversion $\hat{B}_3$ of B_3 to get a conversion $\hat{B}_4$ for B_4. Using condition 3 of theorem 13 we first check if such an extension exists. Recalling the construction we considered when transforming B_3 into B_4, by theorem 13 we have

$$B_4 = \begin{bmatrix} A_1 & O & F_3 \\ F_1 & 1 & O \\ O & F_2 & 1 \end{bmatrix}, \hat{A}_1 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix},$$

and

$$F_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}, F_2 = [1], F_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Thus $(u_1, v_1) = (1, 1), (u_2, v_2) = (1, 1), (u_3, v_3) = (3, 1)$, which means that $(u_0, v_1) = (u_3, v_1) = (3, 1), (u_1, v_2) = (1, 1), (u_2, v_3) = (1, 1)$. To check condition 3, we have to take $\hat{A}_1$, delete from it row 3 and column 1 and look whether it is sign-nonsingular or not. This operation would give

$$\hat{A}'_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

which is not sign-singular. In fact $\text{per}(|\hat{A}'_1|) = 3$ whereas $\det(\hat{A}'_1) = 1$. One could also check that by applying directly the definition of sign-nonsigularity. We consider the matrix

$$U = \begin{bmatrix} 0 & 1 & 1 \\ 2 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

It is clear that $\text{sign}(U) = \text{sign}(\hat{A}'_1)$, but $\det(U) = 0$. The conclusion is that B is not convertible.

1.4.3 Other results

We have seen before that given a square matrix A it is always possible to reduce the computation of $\text{per}(A)$ to the determinant of an exponentially bigger matrix, say B.

Recalling the definition of projection given in Section 1.4.1 we have that the $n \times n$ permanent is a projection of the $m \times m$ determinant if there exists an $m \times m$ matrix f whose entries are constants and indeterminates $x_{i,j}$, for $1 \leq i, j \leq n$, such that $\text{per}(x_{i,j}) = \det(f)$. Now, let us denote by $p(n)$ the smallest such m. Then Valiant's theorem (see Section 1.4.1) gives an upper bound on $p(n)$, i.e., $p(n) = O(n^2 \cdot 2^n)$.

In 1987 von zur Gathen [Ga87] provided a linear lower bound on $p(n)$. In particular he proved that $p(n) > 1.06n-1$, whereas Babai and Seress [BS87] obtained $p(n) > \sqrt{2}n - 6\sqrt{n}$. Finally Meshulam [Me87] contributed with further refinements.

1.5 Existing Algorithms

1.5.1 Exact Algorithms

Ryser Algorithm. Let A be an n-square matrix and Σ be the set of all permutations of the first n integers. Let us consider n-tuples consisting of exactly one nonzero entry from each row of the matrix A. For each $J \subseteq \{1, 2, \ldots, n\}$, let H_J be the set of n-tuples whose entries are all from columns in J. Then, by the principle of inclusion-exclusion, we have that $\text{per}(A) = \sum_J (-1)^{n-|J|} |H_J|$. For each J, we have that $|H_J| = \prod_{i=1}^n p_i$, where p_i is the number of nonzero entries in row i which belong to columns in J. Thus $|H_J|$ can be computed in time $O(n^2)$. Since the number of subsets of $\{1, 2, \ldots, n\}$ is 2^n, the total running time is $O(n^2 2^n)$, which can be improved to $O(n 2^n)$.

We have implemented Ryser algorithm as follows. The input is an $n \times n$ matrix M. Initially we set the value of the permanent to 0. For every $k = 0, 1, \ldots, 2^n - 1$, we convert k in a 0-1 vector v_k containing the binary representation of k. Let $|v_k|$ be the number of 1's in v_k. Then we multiply every row M_j of M by v_k, and we compute the product of the n values obtained. The result is multiplied by $(-1)^{n-|v_k|}$ and added to the current value of the permanent. The overall formula is

$$\text{per}(M) = \sum_{k=0}^{2^n-1} \left[(-1)^{n-|v_k|} \prod_{j=1}^{n} M_j \cdot v_k \right].$$

Our implementation exploits the internal binary representation of numbers, and does not explicitly build the v_k's. It also makes some not expensive precomputations, so that the computation of all the n values $M_j \cdot v_k$ requires at most $n\lceil \frac{n}{8} - 1 \rceil$ additions, instead of $n(n-1)$. Moreover, if one of the values $M_j \cdot v_k$ is zero, we directly skip to the next k, avoiding subsequent useless computations. This is useful especially when the matrix M is very sparse, since in that case a large fraction of the values $M_j \cdot v_k$ is zero. Our implementation results to be more than 70 times faster than a naive one.

Cummings and Wallis algorithm. This algorithm works for circulant matrices and was devised in 1977 by L.J. Cummings and J.S. Wallis [CW77].

Recall that an $n \times n$ matrix $A = (a_{i,j})$ is circulant if it has the Toeplitz structure, i.e., its entries are constant along the diagonals, enriched with the cyclic condition $a_{i,n} = a_{i+1,1}$. Note that the circulant matrix A can be expressed as the sum

$$\sum_{i=1}^{n} a_i P^{i-1},$$

where $P = (p_{i,j})$ is the cyclic permutation with $p_{i,i+1} = 1, i = 1, 2, \ldots, n-1$ and $p_{n,1} = 1$.

Given the cyclic property and remembering the formula of the permanent, Cummings and Wallis could reformulate the permanent of a cyclic matrix A as a polynomial in the indeterminates $a_1, a_2, \ldots, a_n$ and precisely as the summation over the *viable* n-tuples of indices $(i_1, i_2, \ldots, i_n)$, each ranging from 1 to n, of the monomials $a_{i_1} \cdot a_{i_2} \ldots a_{i_n}$.

The trick is enumerating the set S of all such n^n n-tuples lexicographically and discarding those which are illegal. Illegal tuples correspond to monomials composed of elements a_{j_i} some of which occurring in the same column.

The Algorithm

1. Input: circulant $n \times n$ matrix A;
2. Let $\{\alpha_k\}$ be an enumeration of S consistent with the lexicographic order. For $k = 1, 2, \ldots, |S|$ let $\alpha_k = (i_1, i_2, \ldots, i_n) \in S$, with $1 \leq i_j \leq n$, for all j. Then discard α_k if
   (i) $\sum_{j=1}^{n} i_j \not\equiv 0 \pmod{n}$, or if
   (ii) $i_{j+k} \equiv i_j - k \pmod{n}$ for any $k = 1, 2, \ldots, n-1$.
3. Compute and output the sum of all the monomials associated with the set R_n of the remaining n-tuples:
$$\mathrm{per}(A) = \sum_{(i_1, i_2, \ldots, i_n) \in R_n} a_{i_1} a_{i_2} \cdots a_{i_n}.$$

Brualdi and Newman [BN70] determined the number of formally distinct diagonal products in the permanent of a circulant matrix.

Others. We also mention the algorithm devised by Gal and Breitbart [GB74] to solve perfect matching problems and applicable to compute the permanent of a $(0,1)$-matrix. It runs in time $O(kn^3)$ where k is the permanent of the matrix, so it results efficient for large n provided that the permanent is not too large.

1.5.2 Approximation Algorithms

KKLLL algorithm. This randomized algorithm was proposed by Karmarkar, Karp, Lipton, Lovasz and Luby [KKLLL93]. It is based upon an improvement of the Godsil/Gudman estimator [GG81] and runs in time $\mathrm{poly}(n) 2^{n/2} \frac{1}{\epsilon^2}$. Subsequent analyses performed by Frieze and Jerrum [FJ95] showed that the algorithm behaves well on dense $(0,1)$-matrices being an fpras for them.

First we introduce the KKLLL estimator Z:

1. Let $A = (a_{i,j})$ the $(0,1)$-matrix of order n whose permanent has to be estimated;
2. Form an $n \times n$ matrix $B = (b_{i,j})$ as follows. Let $\{1, \omega, \omega^2\}$ be the cube roots of unity. For each pair (i,j), $1 \leq i, j \leq n$:
   (a) if $a_{i,j} = 0$ then $b_{i,j} \leftarrow 0$;

   (b) if $a_{i,j} = 1$ then choose $b_{i,j}$ independently and u.a.r. from the set $\{1, \omega, \omega^2\}$.

3. $Z \leftarrow \det(B) \cdot \overline{\det(B)}$.

The estimator can be evaluated in polynomial time. Furthermore it is unbiased, i.e., $E[Z] = \mathrm{per}(A)$ [KKLLL93].

The Algorithm

1. Input: An $n \times n$ $(0,1)$-matrix A and an integer parameter t;
2. For $i = 1, 2, \ldots, t$ compute the KKLLL estimator and set Z_i equal to the result;
3. Compute $\overline{Z} = \frac{1}{t} \cdot \sum_{i=1}^{t} Z_i$;
4. Output the estimate $\overline{Z}$.

The problem now is determining the minimum number t of trials necessary to obtain an acceptable estimate. This reduces to determining a useful expression of the variance of the estimator. Let G be a bipartite graph on vertex sets $U = V = \{1, 2, \ldots, n\}$, and let M and M' be perfect matchings in G. Denote by $c(M, M')$ the number of connected components (cycles) in $M \oplus M'$, the symmetric difference of M and M'. Define $\gamma(G) = E[2^{c(M,M')}]$ to be the expected value of the random variable $2^{c(M,M')}$, when M and M' are selected u.a.r. from the set of all perfect matchings in G (if G has no perfect matchings then define $\gamma(G) = 1$).

Theorem 14 (see [FJ95] and [KKLLL93])

$$\frac{E[Z^2]}{E[Z]^2} = \gamma(G).$$

□

Corollary 15 (see [FJ95]) *A sequence of $t = O(\epsilon^{-2}\gamma(G))$ trials with the KKLLL estimator suffices to obtain an approximation of the number of perfect matchings in G that satisfies the conditions of a randomized approximation scheme.*

We sketch the proof. Recalling the formula of the variance of an estimator and using theorem 14 we have:

$$\mathrm{Var}\left[\overline{Z}\right] = \frac{\mathrm{Var}[Z]}{t} \leq \frac{\gamma(G)(E[Z])^2}{t}.$$

Now, set $t = \lceil 4\epsilon^{-2}\gamma(G) \rceil$. Then

$$\mathrm{Var}\left[\overline{Z}\right] \leq \left(\frac{\epsilon}{2} E[Z]\right)^2.$$

Finally resuming Chebychev's inequality we have

$$
\begin{aligned}
P\left[(1-\epsilon)\mathrm{per}(A) \le \overline{Z}_t \le (1+\epsilon)\mathrm{per}(A)\right] &= P\left[|\overline{Z} - E[Z]| \le \epsilon E[Z]\right] \\
&= 1 - P\left[|\overline{Z} - E[\overline{Z}]| > \epsilon E[Z]\right] \\
&= 1 - P\left[|\overline{Z} - E[Z]| > \frac{\epsilon E[Z]}{\sigma_{\overline{Z}}} \cdot \sigma_{\overline{Z}}\right] \\
&\ge 1 - \left(\frac{\sigma_{\overline{Z}}}{\epsilon E[Z]}\right)^2 \\
&\ge 1 - \frac{1}{4} \\
&\ge \frac{3}{4}.
\end{aligned}
$$

Remember that given a random variable X, $\sigma_X^2 = \mathrm{Var}[X]$ by definition of standard deviation. Chebychev's inequality has been applied between the third and the fourth passage. □

From corollary 15 it emerges the fact that if the quantity $\gamma(G)$ is polynomially bounded for a class of bipartite graphs the algorithm is an fpras for that class. It is the case for the class of dense graphs [FJ95]. Unfortunately the best we can say in the general unrestricted case is $\gamma(G) \le 2^{n/2}$ [KKLLL93] which implies that the algorithm can be almost as bad as the Ryser exact algorithm.

Finally, consider running the above algorithm with a fixed ϵ. As it is written it requires $\Theta(2^{\frac{n}{2}} n^2)$ random bits in total, i.e., n^2 random bits per trial to randomly choose the values for B. This can be reduced to $O(n^3)$ random bits overall using standard methods of generating pairwise independent unbiased random bits [ACGS88], [Lu86]. In fact the analysis performed in corollary 15 is based upon the Chebychev's inequality and thus it holds even if the random variables $\{Z_i\}$ are just pairwise independent.

Broder Algorithm. This randomized algorithm was first proposed in 1986 by Broder [Br86] but its performance remained unknown until 1989 when Jerrum and Sinclair could analyze it by applying the Markov chain technique [JS89].

Let G be a bipartite graph of $n+n$ nodes. Let M_k be the set of matchings of cardinality k (M_n is then the set of perfect matchings) and define $m_k = |M_k|$. Thus we wish to estimate m_n.

It is easier to view the problem in the more general context of monomer-dimer systems and describe a more recent version of the Broder algorithm due to Jerrum and Sinclair. Let λ be a positive real. We introduce the *partition* function associated with the bipartite graph G:

$$Z(\lambda) = \sum_{k=0}^{n} m_k \lambda^k.$$

Notice that $Z(\lambda)$ is the generating function for matchings of G, i.e., the *matching polynomial* of G.

The problem now is, given G and λ, to randomly approximate $Z(\lambda)$ in time polynomial in $|G|$ and λ. We will see that from the solution to this problem it follows a solution to the permanent.

The strategy is to express $Z(\lambda)$ as the product:

$$Z(\lambda) = \frac{Z(\lambda_r)}{Z(\lambda_{r-1})} \times \frac{Z(\lambda_{r-1})}{Z(\lambda_{r-2})} \times \ldots \times \frac{Z(\lambda_2)}{Z(\lambda_1)} \times \frac{Z(\lambda_1)}{Z(\lambda_0)} \times Z(\lambda_0),$$

where $0 < \lambda_1 < \lambda_2 < \lambda \ldots \lambda_{r-1} < \lambda_r < \lambda$ is a suitably chosen sequence of values.

Note that $Z(\lambda_0) = Z(0) = 1$. If we could compute *suitable* estimates $\{\hat{\rho}_k\}$ for the above ratios $\{\rho_k\}$, for $k = 0, 1, \ldots, r$, we could compute a suitable estimate for the whole product by computing $\hat{\rho}_1 \times \hat{\rho}_2 \times \hat{\rho}_3 \times \ldots \times \hat{\rho}_r$. We use the sequence of values $\lambda_1 = (2\,|E|)^{-1}$ and $\lambda_i = (1 + \frac{1}{n})^{i-1}\lambda_1$ for $1 \leq i < r$. The length r of the sequence is taken to be minimal such that $\lambda_r \geq \lambda$, so the following bound holds:

$$r \leq \lceil 2n(\ln \lambda + \ln(2\,|E|)) \rceil + 1.$$

Consider for simplicity the ratio $\rho_i = \frac{Z(\lambda_i)}{Z(\lambda_{i-1})}$. To compute a *good* estimate $\hat{\rho}_i$ for ρ_i proceed as follows.

(i) First we need a system to sample matchings from the distribution

$$\pi_{\lambda_i}(M) = \frac{\lambda_i^{|M|}}{Z(\lambda_i)}.$$

To this end set up a Markov chain $\mathcal{M}(\lambda_i)$ over the set Ω of all matchings with the crucial property of having π_{λ_i} as stationary distribution. Then make a random walk on $\mathcal{M}(\lambda_i)$ until the stationary distribution is reached and take the reached state M as the sampled element;

(ii) Consider the following random variable:

$$f_i(M) = \left(\frac{\lambda_{i-1}}{\lambda_i}\right)^{|M|}.$$

It holds that $E[f_i] = \frac{Z(\lambda_{i-1})}{Z(\lambda_i)}$.

It is not hard to construct a Markov chain $\mathcal{M}(\lambda_i)$ with the right asymptotic properties. The states set is Ω and the transitions from any matching M are made according to the following experiment:

1. with probability $\frac{1}{2}$ let $M' = M$; otherwise,

2. select an edge $e = (u, v) \in E$ u.a.r. and set

$$M' = \begin{cases} M - \{e\} & \text{if } e \in M; \\ M \cup \{e\} & \text{if both } u \text{ and } v \text{ are unmatched in } M; \\ (M \cup \{e\}) - \{e'\} & \text{if exactly one of } u \text{ and } v \text{ is matched in } M \\ & \text{and } e' \text{ is the matching edge;} \\ M & \text{otherwise;} \end{cases}$$

3. go to M' with probability $\min\{1, \pi_\lambda(M')/\pi_\lambda(M)\}$.

The algorithm can be so outlined:

1. Input: a bipartite graph of $n + n$ nodes G, a parameter λ, a relative error $\epsilon > 0$.
2. Compute the sequence $\{\lambda_1, \lambda_2, \ldots, \lambda_{r-1}\}$, set $\lambda_0 = 0$ and $\lambda_r = \lambda$;
3. For each value λ_i of the sequence compute an estimate $\hat{\rho}_i$ for the ratio $\rho_i = \frac{Z(\lambda_i)}{Z(\lambda_{i-1})}$, as follows
   (i) perform $t_i(\epsilon) = t_i$ independent simulations of the Markov chain $\mathcal{M}(\lambda_i)$, each of length $l_i(\epsilon) = l_i$ obtaining an independent sample of size t_i from (almost) the distribution π_{λ_i};
   (ii) let $\overline{Y}^{t_i}$ the sample mean with respect to the random variable f_i defined as before;
4. Output the product $Y = \prod_{i=1}^{r} \left(\overline{Y}^{t_i}\right)^{-1}$.

Observe that the efficiency of the algorithm depends on the sizes t_i of the samples needed to achieve a good estimate and on the lengths l_i of the random walks to sample from a distribution *close enough* to π_{λ_i}. In particular the l_i's are related to the rate at which the Markov chain approaches stationarity.

Of course we need to quantify "closeness" to stationarity. Let $\mathcal{M}$ be an ergodic Markov chain on state space Ω with transition probabilities $P : \Omega^2 \to [0, 1]$. Let $x \in \Omega$ be an arbitrary state and denote by $P^t(x, \cdot)$ the distribution of the states at time t given that x is the initial state. Denote by π the stationary distribution of $\mathcal{M}$. Then the *variation* distance at time t with respect to the initial state x is defined to be

$$\Delta_x(t) = \max_{S \subseteq \Omega} | P^t(x, S) - \pi(S) | = \frac{1}{2} \sum_{y \in \Omega} | P^t(x, y) - \pi(y) | .$$

The rate of convergence of $\mathcal{M}$ to stationarity may be measured by the function

$$\tau_x(\epsilon) = \min\{t : \Delta_x(t') \leq \epsilon \text{ for all } t' \geq t\}.$$

Now, the issue of the sample size comes from standard results in statistics using Chebyshev inequality:

Theorem 16 (see [JS89]) *Let the sample size* t_i *be equal to* $\lceil 65e\epsilon^{-2}r \rceil$ *and the simulation length* l_i *be large enough that the variation distance of* $\mathcal{M}(\lambda_i)$ *from* π_{λ_i} *satisfies*

$$\max_{x\in\Omega} \Delta_x(t_i) \leq \frac{\epsilon}{4er}.$$

Then for the output random variable Y *it holds that*

$$P\left[(1-\epsilon)Z(\lambda) \leq Y \leq (1+\epsilon)Z(\lambda)\right] \geq \frac{3}{4}.$$

□

The key point in the proof of this theorem is the fact that the sequence of values λ_i has been chosen to render the ratio $\mathrm{Var}[Y]/E[Y]^2$ sufficiently small to apply Chebyshev inequality.

Since $r = O(n)$ by construction the above theorem says that a *small* sample size at each stage suffices to ensure a good final estimate Y, provided that samples come from a distribution that is close enough to π_{λ_i}. So the key point is determining the minimum number of simulation steps t_i needed to obtain the desired variation distance. The following theorem states that the Markov chains $\mathcal{M}(\lambda_i)$ are somewhat *rapidly mixing*:

Theorem 17 (see [JS89]) *The mixing time of the Markov chain* $\mathcal{M}(\lambda_i)$ *satisfies*

$$\tau_x(\epsilon) \leq 4\,|E|\,n\lambda'\left(n(\ln n + \ln\lambda') + \ln\frac{1}{\epsilon}\right),$$

where $\lambda' = \max\{1, \lambda\}$.

□

Theorem 17 guarantees that

$$t_i = \lceil 4\,|E|\,n\lambda_i'(n(\ln n + \ln\lambda_i') + \ln(4er/\epsilon))\rceil$$

fulfills the requirements of theorem 16. Thus the overall running time of the algorithm for monomer-dimer systems is

$$O(n^4\,|E|\,\lambda'(\ln n\lambda')^3\epsilon^{-2}).$$

Finally we turn to the original problem of estimating the number of perfect matchings of G, m_n which is the leading coefficient of the matching polynomial $Z(\lambda)$. Given λ, suppose that we have computed a good estimate $\hat{Z}(\lambda)$ of $Z(\lambda)$. Then to get an estimator of m_n sample matchings from the distribution π_λ defined as above and consider the random variable X defined as

$$X = \begin{cases} 1 & \text{if the sampled element } M \text{ is in } M_n, \\ 0 & \text{otherwise.} \end{cases}$$

Repeat the sampling process t times, pick up t independent samples $M_1, M_2, \ldots, M_t$ and define the sample mean as usual

$$\overline{X}^t = \frac{1}{t}\cdot\sum_{i=1}^{t} X_i.$$

Since $E[\overline{X}^t] = E[X] = m_n \lambda^n / Z(\lambda)$ then the estimator

$$U = X \lambda^{-n} \hat{Z}(\lambda)$$

is unbiased for m_n, i.e., $E[U] = m_n$.

Using Chebyshev inequality it follows that the sample size required to ensure a good estimate depends on the ratio $\mathrm{Var}[U]/E[U]^2 \le E[U]^{-1}$. This quantity, in turn, depends on how large λ is. In fact augmenting λ corresponds to placing very large weight on the perfect matchings so that their proportion can be estimated well by random sampling. The following result is a direct consequence of the *log-concavity* of the sequence $\{m_k\}$, i.e., $m_{k-1} m_{k+1} \le m_k^2$ for $k = 1, 2, \dots, n-1$ (see [JS89] and [HL72]).

Theorem 18 (see [JS89]) *Let $\lambda > m_{n-1}/m_n$. Then*

$$E[U] = \frac{m_n \lambda^n}{Z(\lambda)} \ge \frac{1}{n+1}.$$

$\square$

This implies that the sample size grows only linearly with n, thus it is enough to take λ about as large as the ratio m_{n-1}/m_n and the overall running time of the algorithm will be polynomial in n, ϵ^{-1} and the ratio m_{n-1}/m_n as announced.

Others. Jerrum and Vazirani [JV92] have recently devised a new algorithm which employs the Jerrum & Sinclair algorithm but exploiting expansion properties of the input graph. The overall running time achieved in the worst case is $O(\exp(\sqrt{n} \cdot \log^2 n))$.

1.6 General Lower and Upper Bounds on the permanent

In most cases formulas involving permanents cannot be evaluated if the matrices are at all large. Bounds for permanents are then helpful to obtain approximate solutions.

1.6.1 Bounds for nonnegative matrices

For nonnegative matrices we mean matrices whose entries are all nonnegative. The following upper bound is due to Jurkat and Ryser [JR67].

Theorem 19 (see [JR67]) *Let $A = (a_{i,j})$ be a nonnegative matrix of order n with row sums $r_1, r_2, \dots, r_n$ and column sums $s_1, s_2, \dots, s_n$. Then*

$$\mathrm{per}(A) \le \prod_{i=1}^{n} \min\{r_i, s_i\}.$$

$\square$

1.6.2 Bounds for $(0,1)$-matrices

Let A be an $m \times n$ $(0,1)$-matrix, with $m \leq n$. The permanent of A satisfies the inequality

$$0 \leq \mathrm{Per}(A) \leq n(n-1)\dots(n-m+1).$$

If we have information about the number of ones per row and column then we can sensibly improve the above inequality.

The best known general bounds for this class of matrices are the Minc-Brégman upper bound [Br73],[Mi63] and the Ostrand lower bound [Os70].

Theorem 20 (see [Br73]) *Let $A = (a_{i,j})$ be a $(0,1)$-matrix of order n with row sums $r_1, r_2, \dots, r_n$ and column sums $s_1, s_2, \dots, s_n$. Then*

$$\mathrm{per}(A) \leq \min\left\{\prod_{i=1}^{n}(r_i!)^{\frac{1}{r_i}}, \prod_{i=1}^{n}(s_i!)^{\frac{1}{s_i}}\right\}.$$

□

Theorem 21 (see [Os70]) *Let $A = (a_{i,j})$ be an $m \times n$ $(0,1)$-matrix with* $\mathrm{per}(A) > 0$. *Let the row sums of A be $r_1, r_2, \dots, r_m$ and assume that the rows of A have been arranged so that $r_1 \leq r_2 \leq \dots \leq r_m$. Then*

$$\mathrm{Per}(A) \geq \prod_{i=1}^{m} \max\{1, r_i - i + 1\}.$$

□

Note that theorem 20 specializes and improves theorem 19 on $(0,1)$-matrices.

Let Λ_n^k be the set of n-square $(0,1)$-matrices whose row and column sums are equal to k. Theorem 20 and the lower bound derived from the van der Waerden-Egoryĉev theorem give:

$$n! \cdot \frac{k}{n}^n \leq \mathrm{per}(A) \leq (k!)^{n/k}.$$

Now, let

$$\begin{aligned} \beta(n,k) &= \max\{\mathrm{per}(A) \mid A \in \Lambda_n^k\}, \\ \lambda(n,k) &= \min\{\mathrm{per}(A) \mid A \in \Lambda_n^k\}. \end{aligned}$$

If k is a divisor of n then $\beta(n,k) = (k!)^{n/k}$. Indeed we can take the matrix composed of n/k blocks all equal to J_k. Furthermore if k divides n and A has k ones per row on average still $\mathrm{per}(A) \leq (k!)^{n/k}$ [BGM88].

The following bound is due to Voorhoeve [Vo79].

Theorem 22 (see [Vo79]) *For all $n \geq 3$,*

$$\lambda(n,3) \geq 6\left(\frac{4}{3}\right)^{n-3}.$$

□

1.6.3 Bounds for fully-indecomposable matrices

Foregger [Fo75] contributed with an upper bound on this class of matrices whereas Minc [Mi69] obtained a lower bound.

Theorem 23 (see [Mi69] and [Fo75]) *Let A a fully-indecomposable nonnegative integral matrix of order n, the sum of whose elements equals $\sigma(A)$. Then*

$$\sigma(A) - 2n + 2 \leq \mathrm{per}(A) \leq 2^{\sigma(A)-2n} + 1.$$

□

We close this Section with a result by Brualdi and Gibson [BG77] who provide a similar upper bound which works also if we replace the full indecomposability assumption with the assumption of total support.

Theorem 24 (see [BG77]) *Let A a fully-nonnegative integral matrix of order n with total support, and let t be the number of fully indecomposable components of A. Then*

$$\mathrm{per}(A) \leq 2^{\sigma(A)+2n+t}.$$

□

1.7 Outline of this Thesis

We have concentrated the overview part of the thesis in chapter 1 and partly in chapter 2. The original part is presented in chapters 3 and 4.

In particular in chapter 2 we dissert about the hardness of computing the permanent of very sparse matrices reformulating certain well-known results due to Karpinski, Dahlaus, Dagum and Luby. Furthermore we show some reductions from the permanent problem to #SAT and to counting the solutions to a system of equations. This last transformation will be crucial in the development of an algorithm based on algebraic geometry techniques whose construction is described in chapter 4.

In chapter 3 we focus on the permanent of special cases. In particular we provide closed formulas for some Hessenberg, Circulant and Toeplitz matrices. In addition we construct a new algorithm for circulant matrices derived from a clever application of the Laplace expansion formula. Finally, we use the results about sparse matrices mentioned in chapter 2 to reduce the permanent of arbitrary circulant matrices to block circulant matrices having at most three nonzeros per row and column. All these results are contained in [CCR96].

In chapter 4 we follow a complete different approach. Indeed we devise an algorithm for the permanent of circulant matrices with three nonzeros per row and column that, as we said, takes advantage of the reduction shown in chapter 2. Thus, given a circulant $(0,1)$-matrix A with three nonzeros per row and column we set up a system of polynomial equations whose number of solutions is precisely $\mathrm{per}(A)$. Then, applying algebraic geometry tools, thank to the particular structure of the *ideal* generated by the polynomial equations, we manipulate the system to produce a so-called *Groebner basis.* This ensures that if we keep only the leading terms of the Groebner basis and get rid of the rest we still retain a system of monomial equations whose number of solutions is $\text{per}(A)$. Then we compute that number of solutions. The resuls of this research are contained in [CCR96b]

Chapter 5 contains some conclusive remarks and considerations.

Chapter 2

Complexity issues

2.1 Hardness results for very sparse matrices

We introduce a class, denoted with $\mathcal{C}$, of sparse $(0,1)$-matrices on which the permanent problem was proven to be #P-complete by Dagum, Luby, Mihail and Vazirani (see [DLMV88]). $\mathcal{C}$ is a proper subclass of the class of matrices with three nonzeros per row and column.

Let A be a $(0,1)$-matrix having three nonzeros per row and column. Then $A \in \mathcal{C}$ if and only if there exist

- n couples (x_i, y_i), $i = 1, 2, \ldots, n$, with $m = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$, and
- a permutation $m \times m$ matrix P composed of $n \times n$ blocks:

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \ldots & P_{1,n} \\ P_{2,1} & P_{2,2} & \ldots & P_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n,1} & P_{n,2} & \ldots & P_{n,n} \end{bmatrix},$$

  where $P_{i,j}$ is a $x_i \times y_j$ matrix containing at most one nonzero,

such that

$$A = \begin{bmatrix} B & C_1 & C_2 & \ldots & C_{m+n} \\ D_1 & H_1 & O & \ldots & O \\ D_2 & O & H_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & O \\ D_{m+n} & O & \ldots & O & H_{m+n} \end{bmatrix},$$

and the structure of its blocks is determined by P and by $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ in the following way.

For $i = 1, 2, \ldots, m+n$,

$$H_i = \begin{bmatrix} 1 & 1 & 1 & & & & & & & \\ 1 & 1 & 1 & & & & & & & \\ 1 & 1 & & & & 1 & & & & \\ & & & 1 & 1 & 1 & & & & \\ & & & 1 & 1 & 1 & & & & \\ & & & 1 & 1 & & & & & \\ & & & & & \ddots & & & & 1 \\ & & & & & & & 1 & 1 & 1 \\ & & & & & & & 1 & 1 & 1 \\ & & & & & & & 1 & 1 & \end{bmatrix}.$$

The matrices H_i are all the same and of order $3(5m+2) \times 3(5m+2)$. The matrix B is of order $(2m-n) \times (2m-n)$ and is itself a block matrix $(B_{i,j})$, for $1 \leq i, j \leq n$:

$$B_{i,i} = \begin{bmatrix} 1 & & & & & * & * & \ldots & * \\ 1 & 1 & & & & * & * & \ldots & * \\ & 1 & \ddots & & & * & * & \ldots & * \\ & & 1 & 1 & & * & * & \ldots & * \\ & & & 1 & 1 & * & * & \ldots & * \\ & & & & 1 & * & * & \ldots & * \\ & & & & & 1 & 1 & & \\ & & & & & & \ddots & \ddots & \\ & & & & & & & 1 & 1 \end{bmatrix}.$$

These matrices are of order $(x_i + y_i - 1) \times (x_i + y_i - 1)$. We denote the submatrix of $B_{i,i}$ identified by the first x_i rows and the last y_i columns with $U_{i,i}$. $U_{i,i}$ has at most one nonzero and corresponds to the area above evidentiated in the definition of $B_{i,i}$ with the $*$'s. Now, for $i \neq j$, $B_{i,j}$ is a matrix of order $(x_i + y_i - 1) \times (x_j + y_j - 1)$. Similarly we denote with $U_{i,j}$ the submatrix of $B_{i,j}$ identified by the first x_i rows and the last y_j columns. For all (i, j), $B_{i,j}$ contains at most one nonzero and this must fall in $U_{i,j}$.

Furthermore U must be equal to P, where $U = (U_{i,j})$ is the block matrix composed of the submatrices $U_{i,j}$.

Finally, let us describe the structure of the C_j's and of the D_i's. They have exactly one nonzero element. Each matrix C_i is of order $(2m-n) \times (3(5m+2))$ and

$$C_i[a, b] = 1$$

if and only if $b = 3$ and

$$a = \begin{cases} \displaystyle\sum_{k=1}^{q} (x_k + y_k - 1) + 1 & \text{if } r = 1, \\ \displaystyle\sum_{k=1}^{q} (x_k + y_k - 1) + x_{q+1} & \text{if } r = 2, \\ \displaystyle\sum_{k=1}^{q} (x_k + y_k - 1) + x_{q+1} + r - 2 & \text{if } 3 \leq r \leq y_{q+1} + 1, \end{cases}$$

where q satisfies $i = \sum_{k=1}^{q}(1 + y_k) + r$, $1 \leq r \leq y_{q+1} + 1$. Each matrix D_i is of order $(3(5m + 2)) \times (2m - n)$ and

$$D_i[a, b] = 1$$

if and only if $a = 5m + 2$ and

$$b = \begin{cases} \sum_{k=1}^{q}(x_k + y_k - 1) + r & \text{if } 1 \leq r \leq x_{q+1}, \\ \sum_{k=1}^{q+1}(x_k + y_k - 1) & \text{if } r = x_{q+1} + 1, \end{cases}$$

where q satisfies $i = \sum_{k=1}^{q}(1 + x_k) + r$, $1 \leq r \leq x_{q+1} + 1$.

In [DLMV88] it was proven that each n-square $(0, 1)$-matrix S with m nonzeros can be transformed into a $[(2m - n) + 3(5m + 2)(m + n)]$-square $(0, 1)$-matrix $T \in \mathcal{C}$ such that

$$\text{per}(S) = \lfloor \frac{\text{per}(T)}{4^{(m+n)(5m+2)}} \rfloor.$$

In particular, if row i has x_i nonzeros and column j has y_j nonzeros and the block matrix $P = (P_{i,j})$, for $1 \leq i, j \leq n$, is a permutation matrix such that each $P_{i,j}$, of order $x_i \times y_j$, contains a nonzero element if and only if $S[i, j] = 1$ then T is the $(0, 1)$-matrix in $\mathcal{C}$ determined by $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ and P, as previously shown.

They have extended the following construction by Karpinski and Dahlhaus (see [DK92]). Given a bipartite graph $G = (U, V, E)$, $|U|=|V|= n, |E|= m$, contruct another bipartite graph $G' = (X, Y, F)$, $|X|=|Y|= 2m - n$ with $m + n$ vertices of degree 2 and $m - 2n$ vertices of degree 3 in each bipartition such that there is a 1-1 correspondence between perfect matchings in G' and perfect matchings in G. In terms of matrices their construction corresponds exactly to the submatrix B of T in transformation given in [DLMV88], so that $\text{per}(S) = \text{per}(B)$.

2.2 Reduction to computing the number of solutions to a system of equations

Let A be a $(0, 1)$-matrix. We define a matrix X of variables $x_{i,j}$ as

$$X(i, j) = x_{i,j} \text{ if and only if } a_{i,j} = 1.$$

Let us consider the following system of equations:

$$\begin{cases} \sum_{j=1}^{n} x_{i,j} = 1 & \text{for } i = 1, 2, \ldots, n \\ \sum_{j=1}^{n} x_{i,j} = 1 & \text{for } j = 1, 2, \ldots, n \\ x_{i,j}(1 - x_{i,j}) = 0 \end{cases}$$

Theorem 25 *Let A be a $(0,1)$ $n \times n$ matrix and let X be the matrix constructed as above. Then* $\mathrm{per}(A)$ *is equal to the number of solutions of the system of equations (2.2),(2.2) and (2.2).*

Proof. The permanent of A is the number of its nonzero permutations, where a permutation σ is nonzero if for $k = 1, 2, \ldots, n$, $a_{k,\sigma(k)} = 1$.

It is easy to see that each nonzero permutation σ uniquely corresponds to a solution of the system. We simply let $x_{i,j} = 1$ if $j = \sigma(i)$ and $x_{i,j} = 0$ otherwise. The converse is also true since equations (2.2) impose that the range of each solution must be either 0 or 1 and equations (2.2) and (2.2) select exactly one nonzero entry in each row and column of A. □

Note that the above theorem brings the problem of computing the permanent into the framework of algebraic geometry, so, one can use tools from algebraic geometry in order to find or to bound the number of solutions to a special system of equations.

We have exploited the above connection with algebraic geometry and used the MACAULAY package to compute the permanent of certain matrices.

2.3 Reduction to #-SAT

Using Boolean algebra it is possible to connect the problem of computing the permanent to #-SAT. Let us assume that the matrix A has three nonzero elements per row and per column. Then we construct a matrix X as in Section 2.2 so that the permanent corresponds to the number of $(0,1)$ solutions to the system of equations

$$\begin{cases} x_{i,i_1} + x_{i,i_2} + x_{i,i_3} = 1 & \text{for } i = 1, 2, \ldots, n \\ x_{j_1,j} + x_{j_2,j} + x_{j_3,j} = 1 & \text{for } j = 1, 2, \ldots, n \end{cases} \tag{2.1}$$

The requirements of (2.1) can be expressed as

$$\bigwedge_{i=1}^{n} [(x_{i,i_1} \oplus x_{i,i_2} \oplus x_{i,i_3}) \wedge (\overline{x_{i,i_1} \oplus x_{i,i_2} \oplus x_{i,i_3}})] \bigwedge_{j=1}^{n} [(x_{j_1,j} \oplus x_{j_2,j} \oplus x_{j_3,j}) \wedge (\overline{x_{j_1,j} \oplus x_{j_2,j} \oplus x_{j_3,j}})] \tag{2.2}$$

Finding the number of satisfying assignments to 2.2 corresponds to computing the permanent of A.

Chapter 3

Permanents of special matrices

3.1 Hessenberg Matrices

3.1.1 Preliminaries

In this chapter we analyze the problem of computing the permanent of Hessenberg matrices. We will present a fast algorithm which uses the convertibility theory previously addressed.

Formally, we say that $A = (a_{i,j})$ has *upper bandwidth* p if $a_{i,j} = 0$ whenever $j > i + p$ and *lower bandwidth* q if $a_{i,j} = 0$ whenever $i > j + q$. A banded matrix A is (upper) *Hessenberg* if $q = 1$ and $1 \leq p \leq n - 1$.

3.1.2 An Algorithm

The algorithm we present in this Section is based on the reduction of the computation of the permanent of an arbitrary square Hessenberg matrix A, to the computation of the determinant of another Hessenberg matrix of the same size by changing the sign to some of the elements of A. Furthermore we will see that the position of the elements whose sign is to be changed does not depend upon the specific matrix. This important feature allows us to develop a very fast algorithm since the determinant of an Hessenberg matrix can be efficiently computed.

Let us focus our attention on the convertibility property. First we recall a lemma by Gibson which proves that each Hessenberg $(0, 1)$-matrix with maximum number of nonzero elements can be converted. Then, we generalize the result to all possible Hessenberg matrices.

Lemma 26 (see [Gi71]) *Let $A = (a_{i,j})$ be an upper Hessenberg $(0, 1)$-matrix such that $a_{i,j} = 1$ if and only if $i - j \geq 1$. Then A is convertible and its conversion, say $\hat{A}$ is obtained by changing sign to all of the elements of the lower nonzero diagonal, i.e. to all the elements $a_{i+1,i}$.*

□

The following lemma allows us to generalize the previous result.

Lemma 27 (see [BR91]) *Let E be a $(0,1,-1)$-matrix of order n. Then*

$$\mathrm{per}(|E|) = \det(E) \textit{ iff } \forall X \in \mathbf{R}^{n\times n}\ \mathrm{per}(|E| * X) = \det(E * X),$$

where $*$ is the elementwise Hadamard product.

□

Lemma 27 means that a convertible $(0,1)$-matrix individuates a coordinate subspace of $\mathbf{R}^{n\times n}$ on which the permanent can be evaluated as a determinant. Now, let E be the $n \times n$ Hessenberg $(0,1)$-matrix of lemma 26 and $\hat{E}$ its conversion. Given a generic Hessenberg matrix $A \in \mathbf{R}^{n\times n}$, lemma 27 guarantees that

$$\mathrm{per}(A) = \mathrm{per}(A * E) = \det(A * \hat{E}).$$

In practice, it is sufficient to multiply by -1 all the elements $a_{i+1,i}$, for $i = 1, 2, \ldots, n-1$. This process would take at most $n-1$ operations. Thus we are left with the problem of computing the determinant of an upper Hessenberg matrix of upper band p, for $1 \leq p \leq n-1$. We can do this using Gaussian elimination with partial pivoting:

```
input: Square Matrix A of order n
q := 1
for k := 1 to n − 1 do
    if A[k + 1, k] > A[k, k] then
        exchange row k and row k + 1
    endif
    if A[k, k] ≠ 0 then
        A[k + 1, k] := A[k+1,k] / A[k,k]
        for j := k + 1 to min(k + q, n) do
            for i := k + 1 to min(k + p, n) do
                A[i, j] := A[i, j] − A[i, k] × A[k, j]
            endo
        endo
    endif
endo
det := ∏_{k=1}^{n} A[k, k]
output: det
```

The partial pivoting variation provides numerical stability and also guarantees that the algorithm does not fail.

Let us count the number of arithmetic operations of the algorithm:

$$\begin{aligned}
\sum_{k=1}^{n-1}\left(1 + \sum_{j=k+1}^{k+q}\sum_{i=k+1}^{k+p+1} 2\right) + n - 1 &= \sum_{k=1}^{n-1}(1 + 2(p+1)q) + n - 1 \\
&= n - 1 + 2(p+1)q(n-1) + n - 1 \\
&= n - 1 + 2(p+1)n - 2(p+1) + n - 1 \\
&= (2(p+1) + 2)n - 2(p+1) - 2.
\end{aligned}$$

Thus we have proven the following

Theorem 28 *Let A be an upper Hessenberg matrix of band p. Then it is possible to compute its permanent with at most $2(p+1)n+3n$ operations. Observe that in the tridiagonal case ($p=q=1$) it takes $7n$ operations whereas in the case of largest band ($p=n-1$) it takes $2n^2+3n$ operations.*

□

3.2 Circulant Matrices

3.2.1 Preliminaries

In this Section we will study the permanent of certain $(0,1)$-circulant matrices. In particular we will give a formula for circulants of the type $P^i + P^j$, we will introduce an algorithm based upon Minc's linear recurrence formulas, and we will show other results on circulants of the type $I + P^i + P^j$.

Let P_n denote the $(0,1)$ $n \times n$ matrix with 1's only in positions $(i, i+1)$, $i = 1, 2, \dots, n-1$, and $(n,1)$. A matrix $P^{t_1} + P^{t_2} + \ldots + P^{t_k}$, where $0 \le t_1 < t_2 < \ldots < t_k < n$, is called a $(0,1)$-circulant of type $(t_1, t_2, \dots, t_k)$. Once n is specified, the $n \times n$ $(0,1)$-circulant of type $(t_1, t_2, \dots, t_k)$ is completely determined. Since permanents are invariant under multiplication by any power of P, we can assume, without loss of generality, that $t_1 = 0$. We denote the type of $(0,1)$-circulants of the form

$$I + P^{t_2} + P^{t_3} + \ldots + P^{t_k}$$

by the symbol $\langle z \rangle$, where z is the integer

$$1 + 2^{t_2} + 2^{t_3} + \ldots + 2^{t_k}.$$

For example, the circulant whose first row is $[1,1,0,1,0,0,1,0]$ is of type $\langle 75 \rangle$. The $n \times n$ $(0,1)$-circulant of type $\langle z \rangle$ is denoted by $A_n\langle z \rangle$.

Metropolis, Stein and Stein (see [MSS69]) used a combinatorial argument to obtain linear recurrence formulas for the permanents of $(0,1)$-circulants of type $(0,1,2,\dots,k-1)$. Minc (see [Mi85]) extends their method and their main results to a wider class of $(0,1)$-circulants.

These recurrence formulas express the permanent of a $(0,1)$-circulant $A_n\langle z \rangle$ in terms of the permanents of circulants $A_{n-i}\langle z \rangle$, $i = 1, 2, 3, \dots, 2^{t_k} - 2$, where $t_k = [\log_2 z]$.

3.2.2 An algebraic approach

The following theorem by Minc gives the formulas mentioned before:

Theorem 29 (see [Mi85]) *Let $A_n\langle z \rangle$ be the $n \times n$ $(0,1)$-circulant*

$$I + P^{t_1} + P^{t_2} + \ldots + P^{t_k},$$

$0 < t_2 < t_3 < \ldots < t_k < n, n \geq 2^{t_k} + 2t_k - 2$. *Then*

$$\mathrm{per}(A_n\langle z\rangle) = -\sum_{i=1}^{m} c_i \mathrm{per}(A_{n-i}\langle z\rangle) + c,$$

where $m = 2^{t_k} - 2$, *the* c_i *are integers independent of* n, $c = 2 + 2\sum_{i=1}^{m} c_i$, *and the* $A_s\langle z\rangle$ *are* $s \times s$ *circulants of type* $\langle z\rangle = (0, t_2, t_3, \ldots, t_k)$.

The proof is constructive and will now be outlined.

Consider $A_n = A_n\langle z\rangle$, for $n \geq 2t+1$, where t is a shortcut for t_k. By Laplace's expansion theorem, expand its permanent by the first $n-t$ rows:

$$\mathrm{per}(A_n) = \sum_{\omega \in Q_{n-t,n}} \mathrm{per}(A_n[1,2,\ldots,n-t \mid \omega])\mathrm{per}(A_n[n-t+1, n-t+2, \ldots, n \mid \omega']), \quad (3.1)$$

where ω' is the complement of ω with respect to the increasing sequence $1, 2, \ldots, n$.

Now, as shown in fig. 3.2.2, columns $t+1, t+2, \ldots, n-t$ of matrix $A_n[n-t+1, n-t+2, \ldots, n \mid 1, 2, \ldots, n]$ are all zero. So the nontrivial terms of the expansion are those in which ω does not contain indices $t+1, t+2, \ldots, n-t$. There are $\binom{2t}{t}$ such terms and this number does not depend on n. Furthermore the values of $\mathrm{per}(A_n[n-t+1, n-t+2, \ldots, n \mid \omega'])$ are also fixed and independent on n. In fact, as n increases, the number of zero columns increases whereas the nontrivial ones remain the same. Let $R_n\langle z\rangle$ be the vector whose entries are the subpermanents of the nontrivial $(n-t)$-square submatrices of $A_n[1, 2, \ldots, n-t \mid 1, 2, \ldots, n]$ ordered lexicographically by column indices. Let $C\langle z\rangle$ be the constant vector whose entries are the subpermanents of the nontrivial (t)-square submatrices of $A[n-t+1, n-t+2, \ldots, n \mid 1, 2, \ldots, n]$ ordered anti-lexicographically in column indices. Then formula 3.1 can be restated as the scalar product between two $\binom{2t}{t}$-tuples:

$$\mathrm{per}(A_n\langle z\rangle) = R_n\langle z\rangle \cdot C\langle z\rangle \quad (3.2)$$

for any $n > 2t$. We are now going to show how to obtain a linear recurrence formula for vectors $R_n\langle z\rangle$. If we apply once more Laplace's expansion theorem to the last row of matrices $A_n[1, 2, \ldots, n-t \mid \omega]$ we can observe that all the permanents of the submatrices of $A_n[1, 2, \ldots, n-t \mid 1, 2, \ldots, n]$ can be expressed as a sum of subpermanents of submatrices of $A_{n-1}[1, 2, \ldots, n-t-1 \mid 1, 2, \ldots, n-1]$.

Now, partition the set of nontrivial $(n-t)$-square submatrices of $A_n[1, 2, \ldots, n-t \mid 1, 2, \ldots, n]$ into classes, according to the position of the block of fixed columns $t+1, t+2, \ldots, n-t$. Each class is characterized by the two parameters t and s. In fact s represents the number of columns chosen on the left of the fixed block and so the value $t-s$ will be the number of columns chosen on the right of the block and will be called the width of the class.

Each class of width $t-s$ can be further partitioned into subclasses by choosing a particular t_uple of s indices (the columns on the left).

Let $U^{(n)} = U^{(n)}(i_1, i_2, \ldots, i_s)$ denote the column vector whose entries are the permanents of matrices of the subclass of width $t-s$, whose s chosen columns are $i_1, i_2, \ldots, i_s$.

It turns out that there exists a $\binom{t}{s}$-square $(0,1)$-matrix $\Pi_{t-s} = \Pi_{t-s}\langle z\rangle$ such that

$$U^{(n)} = \Pi_{t-s}U^{(n-1)}$$

The matrix Π_{t-s} is called transformation matrix and is the same for each subclass of the same class. In practice, it depends only on the number s of columns chosen on the left of the fixed block. Furthermore it does not depend on the dimension n.

So for $n > 2t$ we have:

$$U^{(n)} = \Pi_{t-s}^{n-2t}U^{2t}.$$

Now, let

$$f(\lambda) = \prod_{s-1}^{t-1} \det(\lambda I - \Pi_{t-s}) = \lambda^M + \sum_{i=1}^{M} c_i\lambda^{M-i},$$

the product of the characteristic polynomials of all the transformation matrices where $M = 2^t - 2$. By the Cayley-Hamilton theorem each matrix satisfies its own characteristic polynomial and so

$$\Pi_{t-s}^M = -\sum_{i=1}^{M} c_i\Pi_{t-s}^{M-i}$$

for any $n \geq 2^t + 2t - 2$ and then

$$U^{(n)} = -\sum_{i=1}^{M} c_iU^{(n-i)}$$

for each subvector $U^{(n)}$ of R_n associated with a subclass of submatrices and finally

$$R_n = -\sum_{i=1}^{M} c_iR_{n-i}.$$

Let $\tilde{R}_n$ and $\tilde{C}$ be the tuples obtained from R_n and C_n by discarding their first and last entries, each of which is equal to 1. Let

$$P_n = \tilde{R}_n \cdot \tilde{C}_n.$$

Clearly $P_n = \mathrm{per}(A_n) - 2$. Putting everything together we get

$$P_n = -\sum_{i=1}^{M} c_iP_{n-i},$$

i.e.,

$$\mathrm{per}(A_n) - 2 = -\sum_{i=1}^{M} c_i[\mathrm{per}(A_{n-i}) - 2].$$

We can therefore conclude that

$$\mathrm{per}(A_n) = 2f(1) - \sum_{i=1}^{M} c_i\mathrm{per}(A_{n-i}).$$

Structure of Transformation Matrices

Minc also investigates the structure of the transformation matrices. His results are summarized by the following three theorems:

Theorem 30 (see [Mi85]) *The transformation matrix Π_1 of width 1 for the $n \times n$ $(0,1)$-circulant*

$$I + P^{t_2} + P^{t_3} + \ldots + P^{t_k}, \tag{3.3}$$

$n > 2t$, is given by the matrix whose first column entries are the first t entries of the last row of the circulant and whose remaining columns have 1 in position $(j-1, j)$, $j = 2, 3, \ldots, t$, and zeros elsewhere.

Theorem 31 (see [Mi85]) *The transformation matrix Π_{t-1} for the circulant defined by (3.3) is given by the matrix whose first row entries are the first t entries in column t of the circulant and whose remaining rows have a 1 in position $(j+1, j)$, $j = 1, 2, \ldots, t-1$, and zeros elsewhere.*

If A_n is the circulant defined by (3.3), then the circulant

$$I + P^{t_k - t_{k-1}} + P^{t_k - t_{k-2}} + \ldots + P^{t_k - t_2} + P^{t_k}$$

is called the dual of A_n and is denoted by A_n^D. It holds that $A_n^D = P^{t_k} A_n^T$ and hence

$$\mathrm{per}(A_n^D) = \mathrm{per}(A_n).$$

The previous theorem asserts that

$$\Pi_{t-1} = (\Pi_1^D)^T.$$

Theorem 32 gives a characterization of all the other transformation matrices in terms of Π_1.

Definition 2 *Let M be an $n \times n$ matrix, and let r be an integer, $1 \leq r \leq n$. Then the rth permanental compound of M, denoted by $L_r(M)$, is the $\binom{n}{r}$-square matrix whose entries are $\mathrm{per}(M[\alpha \mid \beta])$ arranged lexicographically in $\alpha \in Q_{r,n}$ and $\beta \in Q_{r,n}$.*

Theorem 32 (see [Mi85]) *The transformation matrix Π_r of width r for the $(0,1)$-circulant matrix A_n is the rth permanental compound of the transformation matrix Π_1 of width 1 for A_n.*

3.2.3 Fast Computation of the Permanent of some Circulant Matrices

Since Minc's recurrence formulas are linear an immediate way to use them for computing the permanent is defining a linear lower triangular system of size $n - 2t$ whose indeterminates are $\mathrm{per}(A_{2t}), \mathrm{per}(A_{2t+1}), \ldots, \mathrm{per}(A_n)$. To solve this system we need to precompute the first $2^t - 2$ permanents $\mathrm{per}(A_{2t}), \mathrm{per}(A_{2t+1}), \ldots, \mathrm{per}(A_{2^t+2t-3})$, respectively.

Theorem 33 *Let A be an n-square circulant $(0,1)$-matrix of type $(0, t_2, t_3, \ldots, t_k)$ such that $t = t_k$ satisfies $t \leq \log_2(n) - 1$. Then the permanent of A corresponds to the $(n-2t)$th component of the solution of the $(n-2t) \times (n-2t)$ linear system*

$$Bx = b,$$

where

$$B = \begin{bmatrix} 1 & 0 & \ldots & & & & & & 0 \\ c_1 & 1 & 0 & \ldots & & & & & 0 \\ c_2 & c_1 & 1 & 0 & \ldots & & & & 0 \\ & & \vdots & \ldots & & & & & \vdots \\ c_{2^t-3} & c_{2^t-4} & \ldots & c_1 & 1 & 0 & \ldots & & 0 \\ c_{2^t-2} & c_{2^t-3} & \ldots & c_2 & c_1 & 1 & 0 & \ldots & 0 \\ & & \vdots & & & & \ldots & & \vdots \\ 0 & \ldots & c_{2^t-2} & c_{2^t-3} & \ldots & c_3 & c_2 & c_1 & 1 \end{bmatrix}, b = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_{2^t-3} \\ c \\ \vdots \\ c \end{bmatrix},$$

$$x = \begin{bmatrix} \mathrm{per}(A_{2t}) \\ \mathrm{per}(A_{2t+1}) \\ \mathrm{per}(A_{2t+2}) \\ \vdots \\ \mathrm{per}(A_{2^t+2t-3}) \\ \mathrm{per}(A_{2^t+2t-2}) \\ \vdots \\ \mathrm{per}(A_n) \end{bmatrix}.$$

and where $c_1, c_2, \ldots, c_{2^t-2}$ are the coefficients of Minc formula for the type $\langle z \rangle$ and $a_0 =$ $\mathrm{per}(A_{2t}), a_k = \mathrm{per}(A_{2t+k}) + \sum_{i=1}^{k} c_i \mathrm{per}(A_{2t+i-1})$, for $k = 1, 2, \ldots, 2^t - 3$.

Proof. Let $z = 1 + 2^{t_2} + 2^{t_3} + \ldots + 2^{t_k}$. Since $t \leq \log_2(n) - 1$ it follows that $2^t - 2 + 2t \leq n$ and so we can apply Minc construction to get

$$\mathrm{per}(A) + \sum_{i=1}^{2^t-2} c_i \mathrm{per}(A_{n-i}) = c. \tag{3.4}$$

where A_{n-i}'s denote $(n-i)$-square circulant $(0,1)$-matrices of the same type $\langle z \rangle$. Now, the coefficients c_i's of (3.4) allow us to construct the Toeplitz triangular matrix B. About vector b we have to precompute the values $\mathrm{per}(A_{2t}), \mathrm{per}(A_{2t+1}), \ldots, \mathrm{per}(A_{2^t+2t-3})$ from which it is immediate to derive the needed elements $a_0, a_1, \ldots, a_{2^t+2t-3}$. The structure of the linear recurrence formula (3.4) guarantees that the last component of the solution of the system $Bx = b$, denoted by $(B^{-1}b)_{n-2t}$, will provide us with the value of $\mathrm{per}(A)$. □

In order to develop a fast algorithm to compute the permanent of $(0,1)$-circulant matrices, which exploits, when possible, the theorem above, we have to be able to compute the entries of the matrix B, i.e., the coefficients of the Minc's formula for A (i.e. for type $\langle z \rangle$). The core of the construction is constituted by the determination of the characteristic polynomials of the transformation matrices. Even though the theorems cited in the previous Section tell us how to get all the transformation matrices, it is not known how the coefficients of the characteristic polynomial of the rth permanental compound of a matrix are related to those of the characteristic polynomial of the matrix itself. Anyhow, as we will see, the computational obstacles are given by the precomputation of the initial values for the recurrence formula.

The Algorithm

1. Input: Matrix $A_n\langle z\rangle$ satisfying the assumptions of theorem 33;
2. Compute coefficients c_i's and c of Minc's formula for type $\langle z\rangle$ circulants;
3. Compute $\mathrm{per}(A_{2t}), \mathrm{per}(A_{2t+1}), \ldots, \mathrm{per}(A_{2t+2^t-3})$;
4. Compute $a_0, a_1, \ldots, a_{2t+2^t-3}$;
5. Solve system $Bx = b$ as in theorem 33 and let x' be the solution;
6. Output $(x')_{n-2t}$.

We now evaluate the time performance of the above algorithm.

Phase 1: Computation of $c_1, c_2, \ldots, c_{2^t-2}, c$

We know from the previous Section that the coefficients c_i's and c are those of the product of the characteristic polynomials of the transformation matrices. To determine them we need to determine those polynomials and multiply them among each other. Each $\Pi_r = L_r(\Pi_1)$, for $r = 2, 3, \ldots, t-2$, is a $\binom{t}{r}$-square matrix whose computation requires the evaluation of $\binom{t}{r}^2$ permanents of $r \times r$ matrices and that takes $O\left(r \cdot 2^r \cdot \binom{t}{r}^2\right)$ operations. To compute its characteristic polynomial it takes $O\left(\binom{t}{r}^3\right)$. So all the characteristic polynomials require

$$\sum_{r=2}^{t-2}\left(r \cdot 2^r \cdot \binom{t}{r}^2 + \binom{t}{r}^3\right) \leq 3 \cdot 2^{3t}$$

operations. The product of them requires

$$\sum_{k=1}^{t-4} k\binom{t}{t/2}\log\left[k\binom{t}{t/2}\right] = O(t^{\frac{5}{2}} \cdot 2^t).$$

It follows that phase 1 takes $O(2^{3t})$ operations.

Phase 2: Computation of the initial values

Using Ryser's method to compute $\mathrm{per}(A_{2t}), \mathrm{per}(A_{2t+1}), \ldots, \mathrm{per}(A_{2t+2^t-3})$ it takes

$$T = \sum_{k=2t}^{2t+2^t-3} k \cdot 2^k.$$

Observe that

$$2t \cdot \sum_{k=2t}^{2t+2^t-3} 2^k \leq T \leq (2t + 2^t - 3) \cdot \sum_{k=2t}^{2t+2^t-3} 2^k$$

and so

$$2t \cdot (2^{2t+2^t-2} - 2^{2t}) \leq T \leq (2t + 2^t - 3) \cdot (2^{2t+2^t-2} - 2^{2t}).$$

This means that if $t = \Omega(log(n))$ as assumed in theorem 33 then phase 2 will require an exponential number of operations. For the algorithm to be efficient we need to strengthen the constraint on t by assuming $t < \log\log n$. In that case the overall number of operations needed to accomplish phase 2 as a function of n would result:

$$O(n(\log n)^2).$$

whereas for phase 1 it would be $O((\log n)^3)$ which is negligible compared to phase 2.

□

The above results can be summarized as follows.

Theorem 34 *Let A be a circulant $(0,1)$ matrix of type $(0, t_1, t_2, \ldots, t_k)$. If $t_k \leq \log\log n$ then $\mathrm{per}(A)$ can be computed in $O(n \cdot \log^2 n)$ operations.*

Finally we consider a particular case in which it is possible to save some operations. First of all call a $(0,1)$-circulant palintropic if $A_n = A_n^D$. It is clear that circulant of the form 3.3 is palintropic if and only if $t_i + t_{k-i+1} = t_k$, $i = 1, 2, \ldots, k$. It is worth noting that all the circulants considered by Metropolis, Stein and Stein are palintropic.

If A_n happens to be palintropic, then transformation matrices Π_i and Π_{t-i} have the same characteristic polynomial for $i = 1, 2, \ldots, t-1$, and thus only half as many permanents are required in the recurrence formula if t is odd and somewhat more if t is even.

3.2.4 Circulants of the form $P^i + P^j$

Since the permanent is invariant under multiplication by any power of P, then the permanent of $P^i + P^j$ is equal to the permanent of $I + P^{n-i+j}$, so that we can deal with matrices of the form $I + P^d$ w.l.o.g.

Let $A = I + P^d$ be the matrix to be considered. It is well known that its permanent coincides with the number of cycle covers of the digraph whose adjacency matrix is A. Let us denote such a digraph by $D(A)$. Consider now the graph $D(A - I)$, i.e., $D(P^d)$. It is easy to see that the number of cycle covers of $D(A)$ is equal to 2^k, for k the number of cycles of $D(P^d)$. In fact such cycles are all disjoint and if we add self-loops to $D(P^d)$ we have that for each of these cycles there are exactly two ways to cover its nodes. Either we take all the self-loops or we take the cycle itself. So, since $D(P^d)$ admits only one cycle cover of say k disjoint cycles the permanent of A will be 2^k.

The problem consists of determining the value of k. This value, in general, depends upon d and the size n of the matrix A.

Theorem 35 *Let $G = D(P^d)$ be a digraph whose adjacency matrix is P^d for $1 \leq d \leq n-1$. Then G has a cycle cover of $\gcd(n,d)$ cycles.*

Proof. Since the matrix consists of a nonzero diagonal then each cycle, if there are many, has the same length. Suppose to start from a generic vertex, say, a. Then the cycle to which a belongs will be

$$a, a+d, a+d+d, a+d+d+d, \ldots, a+d+d+\ldots+d = a.$$

This means that to know the length of this cycle (which is the same as of the others) we have to determine the minimum h which satisfies

$$h \cdot d \equiv 0 \pmod{n}.$$

Equivalently

$$h \cdot d = n \cdot t$$

for some integer t. The number of cycles will be given by n/h.

Let us consider two cases:

- $d \mid n$.

  For $t = 1$ we have $h = \frac{n}{d}$ and so $k = d = \gcd(n,d)$.

- $d \not| \, n$.

  Let $u = \gcd(n,d), n = u \cdot n', d = u \cdot d'$. Then

  $$h = \frac{u \cdot n' \cdot t}{u \cdot d'} = \frac{t \cdot n'}{d'}$$

  and for $t = d'$ we have $h = n'$ and so $k = u = \gcd(n,d)$.

□

Corollary 36 $\mathrm{per}(P^i + P^j) = \mathrm{per}(I + P^{n+j-i}) = 2^{\gcd(n,|j-i|)}$.

3.2.5 Circulants of the form $I + P^{d_1} + P^{d_2}$

A combinatorial interpretation

We show that the problem of computing $\mathrm{per}(I + P^s + P^t)$ can be stated in purely combinatorial terms.

The permanent of $I + P^s + P^t$ is the number of permutation matrices that can be "generated" from the three cyclic permutation matrices

$$I, P^s, P^t.$$

This can be restated using the classical notation for permutations, so that I corresponds to $(1, 2, 3, \ldots, n)$, P^s to $(s+1, s+2, \ldots, 1, 2, \ldots, s)$ and P^t to $(t+1, t+2, \ldots, n, 1, 2, \ldots, t)$. Thus we can reformulate the problem as follows. Given the three cyclic permutations,

$$\begin{aligned}&(1, 2, 3, \ldots, n),\\&(s+1, s+2, s+3, \ldots, n, 1, 2, \ldots, s),\\&(t+1, t+2, t+3, \ldots, n, 1, 2, \ldots, t),\end{aligned}$$

we consider the set of sequences of the form

$$(\alpha_1, \alpha_2, \ldots, \alpha_n),$$

where $\alpha_h \in \{h, (s+h) \bmod n, (t+h) \bmod n\}$.

There are 3^n sequences of the above form. The problem is to count the number of permutations within the above set S of sequences.

Note that one could try to do the counting, using the following probabilistic method:

1. Generate a random (uniformly distributed in S) sequence $s \in S$.
2. Compute $Pr[s \text{ is a permutation}]$.
3. Compute $\text{per}(I + P^s + P^t)$ as

$$3^n \cdot Pr[s \text{ is a permutation}].$$

Let us see which are the practical limitations of the above approach.

We set $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, where

$$\alpha_h = (h + x_h \cdot s + y_h \cdot t) \bmod n,\ \ h = 1, 2, \ldots, n,$$

and where x_h and y_h are random variables satisfying the following conditions:

1. $x_h = y_h = 0$ with probability $\frac{1}{3}$,
2. $x_h = y_h = 1$ with probability 0,
3. $x_h = 1$ and $y_h = 0$ with probability $\frac{1}{3}$,
4. $x_h = 0$ and $y_h = 1$ with probability $\frac{1}{3}$.

Now, if we let $h \neq k$, then

$$Pr[\alpha \text{ is a permutation}] = Pr[\alpha_h \neq \alpha_k, \text{ for all } \ h, k].$$

The only method that can be employed to evaluate $Pr[\alpha_h \neq \alpha_k, \text{ for all } \ h, k]$ is the method of conditional probabilities. In fact, we should compute:

$$Pr\left[\bigcap_{h,k}\{\alpha \mid \alpha_h \neq \alpha_k\}\right],$$

and the different probabilities are not independent. Note that the only probabilities that one could easily evaluate are those of the type $Pr[\alpha_h \neq \alpha_{h+1}]$.

An efficient algorithm

We show an algorithm for computing the permanent of $(0,1)$-circulants with three nonzero entries per row which takes advantage of the convertibility of some of their submatrices.

Before stating the result we need some simple Lemmas and Definitions.

Lemma 37 *Let A be a square $(0,1)$ matrix such that $G[A]$ is planar. Then the bipartite graph associated with any square submatrix of A is planar.*

Lemma 38 *Let A be a square $(0,1)$ matrix such that $a_{ij}=1$. Then*

$$\operatorname{per}(A)=\operatorname{per}(A-E_{ij})+\operatorname{per}(A(i|j)),$$

where E_{ij} denotes the matrix whose only nonzero entry is in position (i,j), and $A(i|j)$ denotes the matrix obtained by deleting the i-th row and j-th column of A.

Definition 3 *Let us denote with $\mathcal{P}_{k,n}$ the collection of all k-subsets of the n-set $\{1,2,\ldots,n\}$. Let A be a $(0,1)$ $n\times n$ matrix. Then, for $\alpha,\beta\in\mathcal{P}_{k,n}$, we denote with $A[\alpha,\beta]$ the $k\times k$ submatrix of A determined by rows $i\in\alpha$ and columns $j\in\beta$. Then* $\operatorname{per}(A[\alpha,\beta])$ *is called a* permanental k-minor *of A and we define $p_k(A)$ as the sum of all the permanental k-minors of A, i.e.,*

$$p_k(A)=\sum_{\alpha\in\mathcal{P}_{k,n}}\sum_{\beta\in\mathcal{P}_{k,n}}\operatorname{per}(A[\alpha,\beta]). \tag{3.5}$$

$p_k(A)$ counts the number of different selections of k ones in A, such that each row and column has at most a nonzero entry.

Lemma 39 *Let A be a $(0,1)$ $n\times n$ matrix, and let $a_{ij}=1$. Then $p_k(A)=p_k(A-E_{ij})+p_{k-1}(A(i|j))$, for $k\geq 2$, and $p_1(A)=p_1(A-E_{ij})+1$.*

Proof. Equality $p_1(A)=p_1(A-E_{ij})+1$ is clearly true. From the definition of p_k, and separating the submatrices that contain the element a_{ij} from those that do not contain it, for $k\geq 2$, we have

$$p_k(A)=\sum_{\substack{\alpha\in\mathcal{P}_{k,n}\\ i\in\alpha}}\sum_{\substack{\beta\in\mathcal{P}_{k,n}\\ j\in\beta}}\operatorname{per}(A[\alpha,\beta])+\sum_{\substack{\alpha\in\mathcal{P}_{k,n}\\ i\notin\alpha\,\vee\, j\notin\beta}}\sum_{\beta\in\mathcal{P}_{k,n}}\operatorname{per}(A[\alpha,\beta]).$$

By Lemma 38, we can write

$$\begin{aligned}\operatorname{per}(A[\alpha,\beta]) &= \operatorname{per}(A[\alpha,\beta]-E_{ij})+\operatorname{per}(A[\alpha-\{i\},\beta-\{j\}])\\ &= \operatorname{per}((A-E_{ij})[\alpha,\beta])+\operatorname{per}(A[\alpha-\{i\},\beta-\{j\}]),\end{aligned}$$

if $i\in\alpha$, $j\in\beta$, while, if $i\notin\alpha\vee j\notin\beta$, we clearly have $\operatorname{per}(A[\alpha,\beta])=\operatorname{per}((A-E_{ij})[\alpha,\beta])$. Thus we obtain

$$\begin{aligned}p_k(A) &= \sum_{\substack{\alpha\in\mathcal{P}_{k,n}\\ i\in\alpha}}\sum_{\substack{\beta\in\mathcal{P}_{k,n}\\ j\in\beta}}\operatorname{per}((A-E_{ij})[\alpha,\beta])+\sum_{\substack{\alpha\in\mathcal{P}_{k,n}\\ i\in\alpha}}\sum_{\substack{\beta\in\mathcal{P}_{k,n}\\ j\in\beta}}\operatorname{per}(A[\alpha-\{i\},\beta-\{j\}])\\ &+ \sum_{\substack{\alpha\in\mathcal{P}_{k,n}\\ i\notin\alpha\,\vee\, j\notin\beta}}\sum_{\beta\in\mathcal{P}_{k,n}}\operatorname{per}((A-E_{ij})[\alpha,\beta]).\end{aligned}$$

From the definition of p_k it follows that the sum of the first and the third terms of the last formula is indeed $p_k(A - E_{ij})$, while the second term corresponds to $p_{k-1}(A(i|j))$, and the thesis follows. □

Lemma 40 *Let $A = (a_{ij})$ be an $n \times n$ $(0,1)$ matrix, and let $z(A)$ denote the number of different $(0,1)$ matrices $M = (m_{ij})$ with at most one nonzero entry in each row and column, satisfying $M \leq A$, i.e., $m_{ij} \leq a_{ij}$, for all pairs (i,j). Then, for each nonzero entry a_{ij}, we have*

$$z(A) = \sum_{k=1}^{n} p_k(A) \tag{3.6}$$

$$z(A) = z(A - E_{ij}) + z(A(i|j)), \tag{3.7}$$

and, in general, if the matrix A has k nonzero entries, then $k+1 \leq z(A) \leq 2^k$.

Proof. Equality (3.6) easily follows from the definitions of $z(A)$ and $p_k(A)$, while (3.7) follows from (3.6) and from Lemma 39. □

We now prove a theorem that will be instrumental to the definition of an efficient algorithm for the computation of the permanent of circulants of type $(0, d_1, d_2)$.

We first need the following.

Lemma 41 *The permanent of an $n \times n$ convertible matrix A for which $G[A]$ is planar can be computed in $O(n^\gamma)$ time, $\gamma < 3$.*

Proof. The proof follows, e.g., from the results of [VV89], where it is shown that, if $G[A]$ is planar, then the overall running time for the computation of the permanent of A is dominated by the determinant computation. □

Theorem 42 *Let A, B, and C be $n \times n$ $(0,1)$ matrices such that $A = B + C$, and let $G[B]$ be planar. Then* $\operatorname{per}(A)$ *can be computed in $O(z(C)n^\gamma)$ time, $\gamma < 3$.*

Proof. The proof is by induction on the number k of ones in C.

- If $k = 0$, then $A = B$, $z(C) = 1$, and, since $G[A]$ is planar, then $\operatorname{per}(A)$ can be computed in time $O(n^\gamma)$, $\gamma < 3$, by Lemma 41.
- If $k > 0$, let us consider a one of C in position, say, (i,j), and let $C' = C - E_{ij}$. Then, by Lemma 38, we have

  $\operatorname{per}(B+C) = \operatorname{per}(B+C-E_{i,j}) + \operatorname{per}((B+C)(i|j)) = \operatorname{per}(B+C') + \operatorname{per}((B(i|j)+C(i|j))$

  and by Lemma 40

  $$z(C) = z(C') + z(C(i|j)). \tag{3.8}$$

  The matrices C' and $C(i|j)$ are short of a nonzero entry with respect to C, while $G[B(i|j)]$ is planar by Lemma 37. Hence, by induction, we can claim that $\operatorname{per}(B+C')$

and $\mathrm{per}(B(i|j)+C(i|j))$ can be computed in $O(z(C')n^\gamma)$ and $O(z(C(i|j))(n-1)^\gamma)$ time, respectively. Summing up the two time bounds and using equality 3.8, we obtain

$$O(z(C')n^\gamma)+O(z(C(i|j))(n-1)^\gamma)=O([z(C')+z(C(i|j))]n^\gamma)=O(z(C)n^\gamma)\,,$$

from which the thesis follows.

□

Lemma 43 *The bipartite graphs $G[I+Q^i+Q^j]$ and $G[I+Q^i+(Q^T)^j]$ are planar.*

Proof. Let $A=I_n+Q_n^i+(Q_n^T)^j$. We assume, w.l.o.g., that $\gcd(i,j)=1$ (see the previous Section). For simplicity, consider first the case $n=2(i+j)$. The matrix A can be written as

$$A=\begin{bmatrix} U & B \\ C & V \end{bmatrix},$$

where $U=V=I_{\frac{n}{2}}+P_{\frac{n}{2}}^i$, $B=(Q_{\frac{n}{2}}^T)^j$ and $C=Q_{\frac{n}{2}}^i$. Since $\gcd(i,j)=1$, then $\gcd(i,i+j)=1$, and $\gcd(i,\frac{n}{2})=1$. This means that both $G[U]$ and $G[V]$ are cycles of n nodes, so that $G[A]$ is composed of two identical cycles connected by the $\frac{n}{2}$ edges in $G[B]$ and $G[C]$. To avoid edge intersections, we first draw the two cycles in a concentric way, e.g., $G[U]$ inside $G[V]$. We show that the edges that connect them never cross each other. Let $h=\frac{n}{2}$. The nodes of $G[U]$ and $G[V]$ can be labeled consistently with the traversal direction of the cycles as

$$(1_r)^{in},(1+i)_c^{in},(1+i)_r^{in},(1+2i)_c^{in},(1+2i)_r^{in},\ldots,(1+(h-1)i)_r^{in},1_c^{in},$$

and

$$(1_r)^{ou},(1+i)_c^{ou},(1+i)_r^{ou},(1+2i)_c^{ou},(1+2i)_r^{ou},\ldots,(1+(h-1)i)_r^{ou},1_c^{ou},$$

respectively. (The symbols r, c, in, and ou are used to recall *row*, *column*, *inner cycle*, and *outer cycle*, respectively.)

Note that the edges in B and C can be drawn without crossovers starting from node $(1_r)^{ou}$ in the following way:

$$\{(1_r)^{ou},(1+i)_c^{in}\},$$

$$\{(1+i)_r^{in},(1+2i)_c^{ou}\},$$

$$\{(1+2i)_r^{ou},(1+3i)_c^{in}\},$$

$$\vdots$$

$$\{(1+(h-1)i)_r^{ou},(1)_c^{in}\}.$$

The above construction can be easily generalized to handle an arbitrary number of cycles, i.e., $n=k\cdot(i+j)$, k integer, by drawing them one inside the other in a concentric way, and then applying the same strategy as above.

The planarity of $G[I_n + Q_n^i + Q_n^j]$ follows from the fact that it is a subgraph of $G[I_{n+i} + Q_{n+i}^{j-i} + (Q_{n+i}^T)^i]$. □

We are now ready to state our result.

Theorem 44 *Let $A = I_n + {P_n}^i + {P_n}^j$. Then $\mathrm{per}(A)$ can be computed in time $O(2^{i'+j'} n^{O(1)})$, where i' and j' are the two smallest numbers among $\{i, j, n-i, n-j\}$.*

Proof. The matrix A can be viewed as a Toeplitz matrix containing the identity and 4 diagonals of lengths $\{i, j, n-i, n-j\}$. It is thus possible to write $A = B + C$, where C contains the two shorter diagonals, of lengths i' and j', and B consists of the other three diagonals. By Lemma 43, $G[B]$ is planar and we can apply Theorem 42 to get the time bound $O(z(C)n^\gamma)$. The thesis follows since $z(C) \le 2^{i'+j'}$. □

From Theorem 44 we have that $\mathrm{per}(I_n + {P_n}^i + {P_n}^j)$ can be computed in polynomial time if i and j are either smaller than $O(\log n)$ or greater than $n - O(\log n)$.

In Section 3.2.5 we will use some properties of the bipartite graph $G[A]$ in order to strengthen the above result.

Permanent versus Determinant

The results of Section 3.2.5 can be used to describe some structural properties of the permanents of circulant matrices with three nonzeros per row, and, in particular, the relationship between these permanents and the determinants of Toeplitz matrices with at most three nonzeros per row.

Unlike the Toeplitz matrices of Section 3.6, $(0, 1)$-circulant matrices of type $(0, d_1, d_2)$ are not convertible, except for very special cases, e.g., when $d_1 = 1$ and $d_2 = 2$ and the matrix has even size. Nevertheless their permanents bear some interesting connections with suitable determinants. These connections depend on the convertibility of the Toeplitz matrices of Section 3.6.

In fact, if one applies Laplace expansion to, say, a matrix of type $(0, 1, t)$, most of the submatrices induced after some steps, are the convertible Toeplitz matrices analyzed in Section 3.6.

The Laplace expansion for the permanent of $I + P + P^2$, shown in figure 3.1 for $n = 6$, outlines the close relationship of this matrix with the Fibonacci matrix $T_n[1, 1] = I + Q + Q^T$, whose permanent is $F(n + 1)$. In particular one can see that $\mathrm{per}(I + P + P^2) = F(n) + 2F(n-1) + 2$. By comparing the expansion for the permanent and the determinant of $I + P + P^2$, and recalling that the matrix $T_n[1, 1]$ is convertible (and specifically that $\mathrm{per}(I + Q + Q^T) = \det(I - Q + Q^T)$) one can immediately see that, if n is even, then the matrix $I + P + P^2$ is convertible (see also [ST87]). In particular one obtains $\mathrm{per}(I + P + P^2) = \det(I + Q - (Q^T)^{n-1} - Q^2 + (Q^T)^{n-2})$.

Note that, if n is odd, $I + P + P^2$ is not convertible; nevertheless its permanent satisfies $\mathrm{per}(I + P + P^2) = 2 + \det(I + P - P^2)$, as one can readily check.

The same kind of analysis can be performed for the permanent of $I + P + P^3$. In this case, one of the possible ways to carry out Laplace expansion, leads to 14 submatrices, the sum of whose permanents is equal to $\mathrm{per}(I + P + P^3)$. In particular, one obtains two

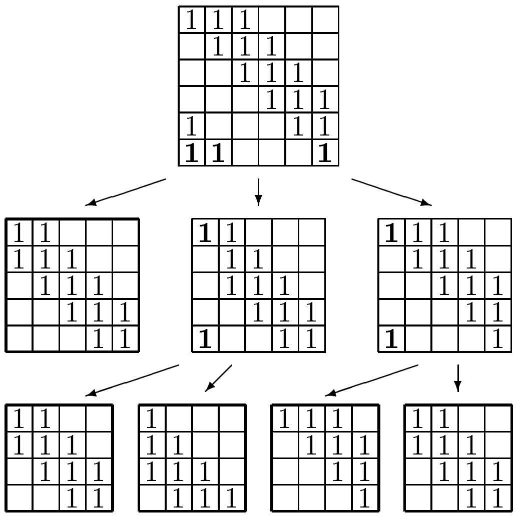


Figure 3.1: Laplace expansion for the permanent of $I+P+P^2$ for $n=6$. The bold entries are those used in the elimination.

triangular matrices, and 12 convertible Toeplitz matrices, of sizes between $n-2$ and $n-6$, of the form $I+Q^2+Q^T$ and $Q+(Q^T)^2+Q^T$. If T_1 and T_2 denote the permanent of these (convertible) matrices, then

$$\begin{aligned} \mathrm{per}(I+P+P^3) &= 2+T_1(n-2)+3T_1(n-3)+T_1(n-4) \\ &\quad + T_2(n-2)+2T_2(n-3)+T_2(n-4)+2T_2(n-5)+T_2(n-6) \end{aligned} \tag{3.9}$$

Summarizing, both $\mathrm{per}(I+P+P^2)$ and $\mathrm{per}(I+P+P^3)$ can be conveniently expressed in terms of a few determinants of Toeplitz matrices. These results can be generalized, although the corresponding formulas become more complicated.

Reduction

Our first result allows one to transform certain permanent computation into the computation of powers of permanents of smaller matrices.

Lemma 45 *Let* $A_n = I_n + P_n^{d\cdot a} + P_n^{d\cdot b}$.

1. *If* $d \mid n$ *then*

$$\mathrm{per}(A_n) = \mathrm{per}(I_{\frac{n}{d}} + P_{\frac{n}{d}}^a + P_{\frac{n}{d}}^b)^d,$$

2. *if* $\gcd(d,n)=1$ *then*

$$\mathrm{per}(A_n) = \mathrm{per}(I_n + P_n^a + P_n^b),$$

where the subscripts indicate the matrix size.

Proof.
(1) The idea of the proof is that by simultaneous rows and columns permutations we can obtain a block matrix C_n of d square blocks of the type $B_{\frac{n}{d}} = I_{\frac{n}{d}} + P^a_{\frac{n}{d}} + P^b_{\frac{n}{d}}$. To show that, we consider the digraph $D(A_n)$ and we prove that it is isomorphic to the digraph $D(C_n)$. First, each cycle of $D(A_n)$ can be represented by a solution of the following equation

$$d \cdot a \cdot x_1 + d \cdot b \cdot x_2 \equiv 0 \pmod{n}. \tag{3.10}$$

For instance, a cyclic path starting from, say, node x can be represented by the sequence of visited nodes, namely:

$$x, x + t_1, x + t_1 + t_2, \ldots, x + t_1 + t_2 + \ldots + t_j$$

where for each $i = 1, 2, \ldots, j$, t_i is either $d \cdot a$ or $d \cdot b$ and $t_1 + t_2 + \ldots + t_j \equiv 0 \pmod{n}$. If there are x^* occurrences of $d \cdot a$ and y^* occurrences of $d \cdot b$ in $\sum_i t_i$ then (x^*, y^*) is a solution of 3.10. Note that not all the solutions of 3.10 represent cycles of the graph. A solution (x', y') represents a cycle of the graph if and only if we can arrange the $l = x' + y'$ elements $d \cdot a$ and $d \cdot b$ on a sequence $s_1, s_2, \ldots, s_l$ so that

1. for each other solution (x'', y'') such that $0 \leq x'' \leq x'$, $0 \leq y'' \leq y'$ and $0 < h = x'' + y''$ there is not a subsequence of the type $s_k, s_{k+1}, \ldots, s_{k+h-1}$, for $1 \leq k \leq l - h + 1$, containing exactly x'' elements $d \cdot a$ (and so y'' elements $d \cdot b$),
2. $\displaystyle\sum_{i=1}^{l} s_i \equiv 0 \pmod{n}$.

On the other hand the following Diophantine equation represents, instead, the cycles of $D(C_n)$ which are the cycles of $D(B_{\frac{n}{d}})$ replicated d times:

$$a \cdot x_1 + b \cdot x_2 \equiv 0 \pmod{\frac{n}{d}}. \tag{3.11}$$

Since $d \mid n$ it follows that equation 3.10 and equation 3.11 are the same. So the two graphs $D(A_n)$ and $D(C_n)$ have the same connected components and the same cycles. It is then possible to define an isomorphism between them, by mapping connected components by connected components. This proves that there exists a permutation matrix Q_n such that $A_n = Q_n \cdot B_n \cdot Q_n^{-1}$ and thus $\mathrm{per}(A_n) = \mathrm{per}(C_n) = \mathrm{per}(B_{\frac{n}{d}})^d$.

(2) In this case we want to prove that $D(A_n)$ is isomorphic to $D(C_n)$, with $C_n = I_n + P^a_n + P^b_n$. Since $\gcd(d, n) = 1$, d admits multiplicative inverse, denoted by d^{-1}, in the ring $\mathbf{Z}_n$ of the residue classes modulo n. Let $\psi : \mathbf{Z}_n \to \mathbf{Z}_n$ be the function defined as follows

$$\psi(x) = x \cdot d^{-1} \bmod n.$$

It turns out that it is an isomorphism between the two digraphs $D(A_n)$ and $D(C_n)$. Surjectivity and injectivity are immediate. We have to show that ψ preserves the structure of the digraphs, i.e., for each pair of nodes $(x, y) \in \mathbf{Z}_n^2$, $(x, y) \in \mathrm{E}(D(A_n))$ if and only if

$(\psi(x), \psi(y)) \in \mathrm{E}(D(C_n))$. In fact, if $(x, y) \in \mathbf{Z}_n^2$ then without loss of generality $y = x + d \cdot a \bmod n$. So

$$\begin{aligned} \psi(y) &= x \cdot d^{-1} + d \cdot a \cdot d^{-1} \bmod n \\ &= x \cdot d^{-1} + a \bmod n \\ &= \psi(x) + a \bmod n. \end{aligned}$$

Thus $(\psi(x), \psi(y))$ is an edge in $D(C_n)$. The converse follows analogously. □

As an example of application of lemma 45 we prove, in the following Section, a formula for the permanent of symmetric circulant matrices, i.e., matrices of the form $I + P^i + P^{n-i}$.

The bipartite graph $G[A_n]$

Let us consider $n \times n$ circulant matrices of type $(0, d_1, d_2)$, for n prime.

To analyze the permanent of the matrix $A_n = I + P^{d_1} + P^{d_2}$, we take into account the bipartite graph associated with the matrix $B_n = P^{d_1} + P^{d_2}$.

Definition 4 *Let $G = (X, Y; E)$ be a bipartite graph, where $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$. We say that a pair of nodes is symmetric if it is of the form $\{x_i, y_i\}$. Let M be a perfect matching of G. We denote by $X(M)$ and $Y(M)$ the set of nodes of X and Y on which the edges of M are incident, respectively. A symmetric matching of cardinality $m \le n$ is a matching $M = \{e_1, e_2, \dots, e_m\}$ of cardinality m such that for all $u \in X(M)$ there exists $v \in Y(M)$ such that u and v are symmetric.*

We have the following simple lemma, which shows that the problem of computing $\mathrm{per}(A_n)$ can be reduced to the computation of the number of symmetric matchings in $G[B_n]$, and to $\mathrm{per}(B_n) = 2^{\gcd(n, |d_1 - d_2|)} = 2$, due to the primality of n.

Lemma 46 *Let A_n and B_n be as above and let*

$$m_k = \#\{\sigma \mid \sigma \textit{ is a symmetric matching of cardinality } n-k \textit{ in } G[B_n]\}. \tag{3.12}$$

Then

$$\mathrm{per}(A_n) = 1 + \mathrm{per}(B_n) + \sum_{k=1}^{n-1} m_k. \tag{3.13}$$

Proof. Equality (3.13) is based upon the observation that each perfect matching of $A_n = I + B_n$ is either a perfect matching of B_n or a perfect matching of A_n which contains k symmetric pairs of nodes for some k, $1 \le k \le n$. On the other hand a perfect matching of A_n containing exactly k pairs of symmetric nodes is a symmetric matching of B_n of cardinality k. □

We will show below how in some special cases it is possible to find a formula for $\mathrm{per}(A_n)$ by evaluating the m_k's.

Note that a lemma similar to Lemma 46 could actually be stated in terms of $D(A_n)$ and $D(B_n)$. In this case we would get a special case of the known expansion

$$\mathrm{per}(\lambda I_n + B_n) = \sum_{i=0}^{n} a_i \lambda^{n-i},$$

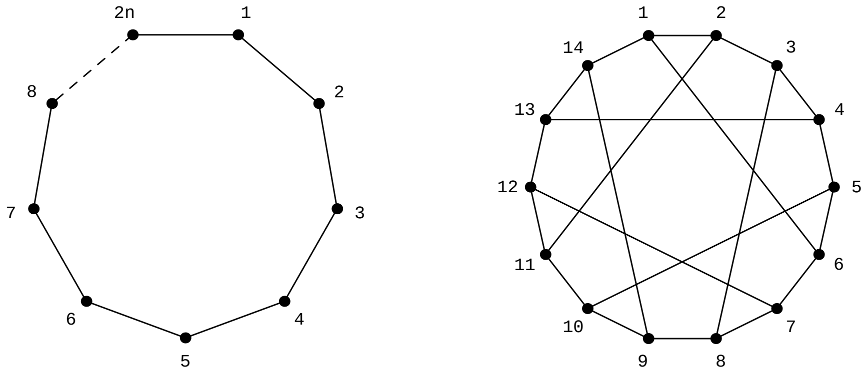


Figure 3.2: Clockwise numbering of $2n$ nodes and the graph for $n = 7$ and $D = 5$.

where the a_i's denote the number of cycle covers for subgraphs of $D(B_n)$ with i nodes (see, e.g., [CDS79] page 34).

Note that $G[B_n]$ is the disjoint union of two perfect matchings, since P^{d_1} and P^{d_2} represent two cyclic permutations of the nodes. As a result of such a union we get a finite set of $\gcd(n, |d_1 - d_2|)$ disjoint rings. So in our case, due to the primality of n, all the nodes of $G[B_n]$ belong to a unique ring and thus each pair of symmetric nodes are connected by a simple path of length $D = D(n, d_1, d_2)$.

The nodes of $G[B_n]$ can be drawn on a polygon. Let us choose one node of the polygon and label it with 1. Then we proceed clockwise and number the second node with 2, the third one with 3, and so on until we use the label $2n$ (see Figure 3.2).

We now determine the value of D. Starting from, say, $u \in X$ we reach its symmetric node $v \in Y$ after an odd number of *moves* along a path on the ring. Note that moving on the ring from a node of X to a node of Y corresponds to adding d_1 modulo n whereas moving from a node of Y to a node of X corresponds to adding $n - d_2$ modulo n. In fact, to move from Y to X we need to consider the inverse of the cyclic permutation represented by P^{d_2}, i.e. the cyclic permutation represented by $(P^{d_2})^{-1} = P^{n-d_2}$. Thus, if x is the index of both u and v, we obtain the following equation

$$x + d_1 + (n - d_2) + d_1 + (n - d_2) + \ldots + d_1 \equiv x \pmod{n}$$

i.e.,

$$d_1 + h \cdot d_1 - h \cdot d_2 \equiv 0 \pmod{n}, \tag{3.14}$$

where $h = (D - 1)/2$. Equation (3.14) leads to the linear congruence

$$h \cdot (d_1 - d_2) \equiv -d_1 \pmod{n} \tag{3.15}$$

which has solutions if and only if

$$\gcd(n, d_1 - d_2) \text{ divides } d_1. \tag{3.16}$$

In this case $\gcd(n, |d_1 - d_2|) = 1$ and certainly divides d_1. Furthermore, since n is prime, $\mathbf{Z}_n$ is a field and we can express h as

$$h = d_1 \cdot (d_2 - d_1)^{-1}, \tag{3.17}$$

so that

$$D = 2d_1 \cdot (d_2 - d_1)^{-1} + 1, \tag{3.18}$$

where both inversion and multiplication are computed in $\mathbf{Z}_n$. Note that we can express D as

$$D(n,i,j) = \min\{1+2h, 2n-2h-1\} \pmod{2n}$$

(In fact D and $2n-D$ play exactly the same role, when interpreted in the bipartite graph.)

To compute m_k, we can choose k pairs of symmetric nodes on $G[B_n]$, connect them two by two with k horizontal edges and count the number of perfect matchings of cardinality $n-k$. Then m_k will be given by the sum of such matchings over all possible choices of k pairs. Pictorially, this translates into drawing k chords connecting two nodes of the polygon at distance D. Moreover, according to the new numbering, chords starting from odd nodes end up to even nodes moving clockwise by D edges, whereas starting from even nodes they end up to odd nodes moving counterclockwise by D edges (see Figure 3.2).

It is easy to see that, for each choice of k chords, there is either one or zero matchings. This amounts to saying that once we have removed from the polygon the nodes matched by the chords and their adjacent edges there is a matching if we are not left with simple paths of even length (odd number of nodes).

By analyzing the graph $G[B_n]$ according to Lemma 46, we have determined the following formula, which holds if $D(n,i,j) = n-2$, i.e., for $i=1$ and $j=3$, (and also if $D(n,i,j)=5$).

$$\mathrm{per}(A_n) = 3 + \sum_{k=1}^{n-1} m_k, \tag{3.19}$$

where

$$m_k = \begin{cases} \frac{n}{(k/2)!} \prod_{i=1}^{k/2-1} (n-k-i), & \text{if } k \text{ is even,} \\ \frac{n}{k!\cdot 2^{k-1}} \prod_{i=1}^{k-1} (n-k-2i), & \text{if } k \text{ is odd.} \end{cases} \tag{3.20}$$

The correctness of (3.20) can be shown by comparison with a formula that can be derived from a known recurrence [Mi85] for $I+P+P^3$, i.e., for $n \geq 12$,

$$\mathrm{per}(A_n) = \mathrm{per}(A_{n-1}) + \mathrm{per}(A_{n-2}) + \mathrm{per}(A_{n-3}) - \mathrm{per}(A_{n-4}) - \mathrm{per}(A_{n-5}) - \mathrm{per}(A_{n-6}) + 2.$$

Solving this linear recurrence, we obtain

$$\mathrm{per}(A_n) = 2 + \lfloor \frac{1}{2} + \alpha^n + \beta^n \rfloor, \tag{3.21}$$

where α and β are the real solutions of the equations x^3-x-1 and x^3-x^2-1, respectively[1], i.e.,

$$\alpha = \sqrt[3]{\frac{9+\sqrt{69}}{18}} + \sqrt[3]{\frac{9-\sqrt{69}}{18}},$$

[1] In particular, $\alpha = 1.3247\ldots$and $\beta = 1.4655\ldots$, whereas the other solutions of the correspondent equation $x^6 - x^5 - x^4 - x^3 + x^2 + x + 1 = 0$ are complex and their absolute values are less than 1.

and

$$\beta = \frac{1}{3} + \frac{1}{3}\sqrt[3]{\frac{29+3\sqrt{93}}{2}} + \frac{1}{3}\sqrt[3]{\frac{29-3\sqrt{93}}{2}}.$$

One can check that (3.21) coincides with (3.19).

Matrices of the form $I + P^i + P^j$, for n prime, and $D(n,i,j)$ small.

The permanent of the matrix $I + P^i + P^j$ depends on n and on $D(n,i,j)$. Note that, by varying i and j, $D(n,i,j)$ takes all possible odd values between 3 and n. In particular, from (3.18), we have that $D(n,1,j) = 1 + 2\cdot(j-1)^{-1}$. Since n is prime, it is easy to check that, for $2 \leq j \leq n-1$, $D(n,1,j)$ goes through all possible values between 3 and n. This means that, if n is prime, the problem of computing $\mathrm{per}(I + P^i + P^j)$ can be reduced to the computation of $\mathrm{per}(I + P + P^k)$, for a suitable k.

Furthermore, note that the function $D_n(i,j) = D(n,i,j)$ is not injective, i.e., there are several pairs (i,j) with the same value of D. In addition, there are other symmetries, because the matrices $I + P^i + P^j$, $I + P^{j-i} + P^{n-i}$ and $I + P^{n-j+i} + P^{n-j}$ have all the same permanent whereas they may have different values of D (see Section 3.7).

We now outline a general approach to derive recurrence formulas for the permanent of matrices with a given value of $D(n,i,j)$.

We operate on the bipartite graph as follows:

- Being n prime, we can represent the graph $G[A_n]$ as a ring, corresponding to $G[B_n]$, with additional chords that connect two nodes of $G[B_n]$ at distance D.
- We consider $D-1$ adjacent nodes, starting from node 1.
- We enumerate all the different ways in which these $D-1$ nodes can be included in a perfect matching.
- The graph on the remaining $n-D+1$ nodes is an *open ring*, i.e. it does not contain a ring any more. This open ring can take only a few different shapes, and these depend on D. This makes it possible to enumerate the perfect matchings according to suitable recurrence formulas.
- Adding up the different terms, one gets a formula for the permanent depending on some simpler functions, each defined by a suitable recurrence.

In figure 3.3 we outline the case when $D = 3$. The nodes considered are 1 and 9 and there are 5 possibilities overall to be analyzed.

In all of the 5 possibilities the ring is opened and we are left with a *necklace of trapezoidal elements*. If we denote by $S(n)$ the number of perfect matchings for a necklace of n trapezoidal elements, it is easy to prove that

$$\begin{cases} S(1) & = & 2 \\ S(2) & = & 3 \\ S(n) & = & S(n-1) + S(n-2) \end{cases}$$

One gets the following formula for the permanent

$$P_{D=3}(n) = 2 + S(n-2) + 2S(n-3).$$

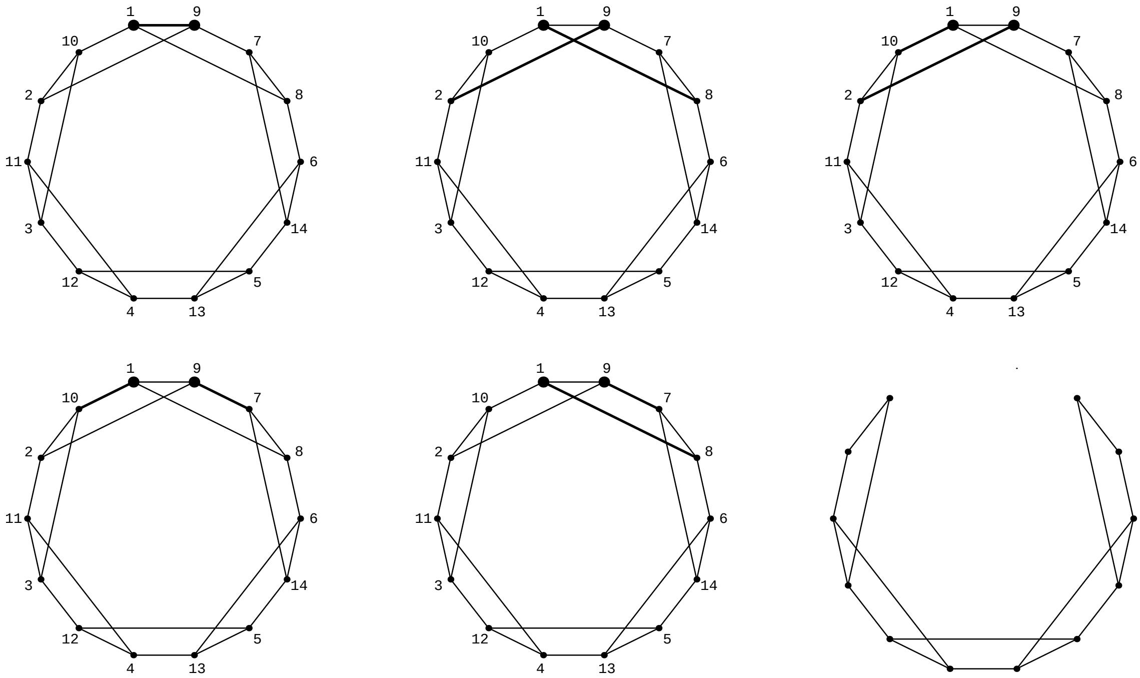


Figure 3.3: The 5 possibilities to include nodes 1 and 9 in a perfect matching, and the typical shape of an open ring.

Note that, although the formula has been obtained under the assumption that n is prime, we experimentally found that it is valid for all values of n up to 25.

The above procedure can be repeated for $D = 5$. In this case, to open the ring we have to consider $D - 1 = 4$ nodes, and the number of possibilities that must be analyzed is 13. (In general, if we define $F(1) = F(2) = 1$, and $F(n) = F(n-1) + F(n-2)$, then the number of possibilities to be analyzed is $F(D+2)$, e.g., if $D = 7$, there are $F(9) = 34$ cases.)

If $D = 5$, the necklaces are of two types, as shown in figure 3.4.

For the number of perfect matchings of the two necklaces, we get the following formulas:

$$\begin{cases} T_1(1) &= 2 \\ T_1(2) &= 3 \\ T_1(3) &= 4 \\ T_1(n) &= T_1(n-1) + T_1(n-3) \end{cases} \qquad \begin{cases} T_2(1) &= 1 \\ T_2(2) &= 2 \\ T_2(3) &= 2 \\ T_2(n) &= T_2(n-2) + T_2(n-3) \end{cases}$$

In this case, the recurrence for the permanent becomes

$$P_{D=5}(n) = 2 + T_1(n-4) + 3T_1(n-5) + T_1(n-6) + 2T_2(n-5) + 3T_2(n-6)\,.$$

(Note that the recurrences obtained here for $D = 3$ and $D = 5$ have some similarities with those obtained by other means in Section 3.2.5 for $I + P + P^2$ and $I + P + P^3$.)

The computational complexity of the above method can be analyzed by recalling Theorem 44. Indeed we have the following corollary, which implies that the permanent of

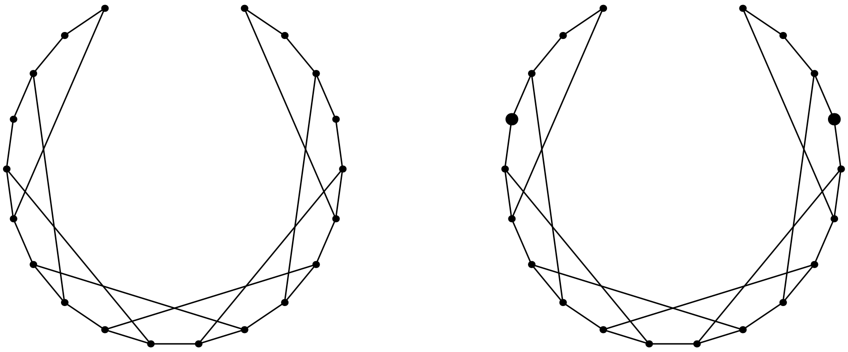


Figure 3.4: The two different necklaces for $D = 5$. Two nodes in the second one are already included in matchings.

$(0, 1)$-circulant matrices of type $(0, d_1, d_2)$ with $D = O(\log n)$ can be computed in polynomial time.

Corollary 47 *Let* $A_n = I_n + {P_n}^i + {P_n}^j$*, with* n *prime, and let* $D = D(n, i, j)$*. Then* $\mathrm{per}(A)$ *can be computed in time* $O(2^{D/2} n^{\gamma})$*,* $\gamma < 3$*.*

Proof. With reference to Theorem 44, the thesis easily follows observing that the matrix $I_n + P_n + {P_n}^{\frac{D+1}{2}}$ has the same permanent as A_n, even if $D(n, 1, (D+1)/2) \neq D$. To see this property, we start drawing $G[A_n]$ in such a way that $G[P^i + P^j]$ is a ring, while I induces chords connecting nodes of $G[P^i + P^j]$ at distance D in $G[P^i + P^j]$. Since the relabeling of the nodes of the bipartite graph does not affect the value of the permanent, we can sort the labels as shown in figure 3.5. According to this new ordering, it is easy to see that the ring now corresponds to $G[I + P]$, and the chords to $G[P^{\frac{D+1}{2}}]$. $\square$

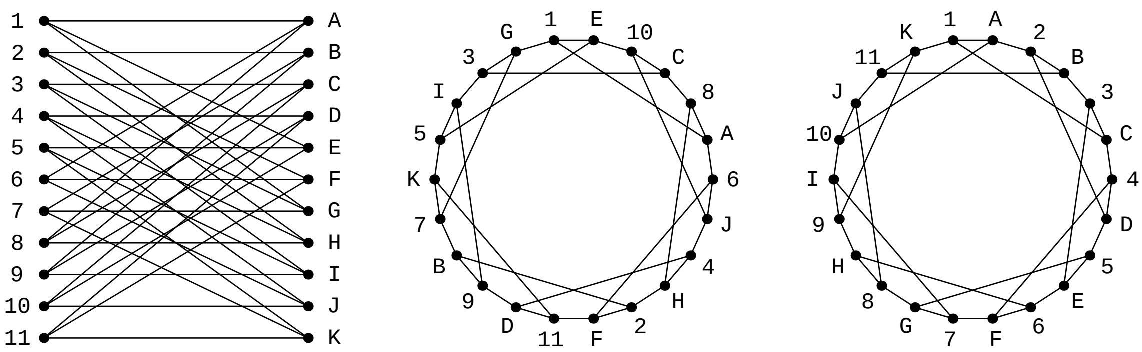


Figure 3.5: From left to right: $G[I+P^4+P^6]$, for $n = 11$; another equivalent description of the same graph which shows that $D = 5$; and the graph with rearranged labels according to the scheme $1, A, 2, B, 3, C \ldots$, which corresponds to the matrix $I + P + P^3$.

Circulants of the form $I+P^i+P^{n-i}$

First of all, note that

$$\mathrm{per}(I+P^i+P^{n-i}) = \mathrm{per}(P^i(I+P^i+P^{n-i})) = \mathrm{per}(I+P^i+(P^i)^2).$$

We are now ready to state the following

Theorem 48 *Let $A_n = I_n + P_n^i + P_n^{n-i}$, and assume that $\frac{n}{\gcd(n,i)} \geq 5$. Then*

$$\mathrm{per}(A_n) = \left[\left(\frac{1+\sqrt{5}}{2}\right)^{\frac{n}{\gcd(n,i)}} + \left(\frac{1-\sqrt{5}}{2}\right)^{\frac{n}{\gcd(n,i)}} + 2\right]^{\gcd(n,i)}.$$

Proof. Let $d=\gcd(i,n)$. Then, by the equivalence of types $(0,i,n-i)$ and $(0,i,2i)$ and from Lemma 45 we have

$$\begin{aligned}
\mathrm{per}(I_n+P_n^i+P_n^{n-i}) &= \mathrm{per}(I_n+P_n^i+(P_n^i)^2)\\
&= \mathrm{per}\left(I_{\frac{n}{d}}+P_{\frac{n}{d}}^{\frac{i}{d}}+P_{\frac{n}{d}}^{2\cdot\frac{i}{d}}\right)^d\\
&= \mathrm{per}\left(I_{\frac{n}{d}}+P_{\frac{n}{d}}+P_{\frac{n}{d}}^2\right)^d.
\end{aligned}$$

By solving Minc's recurrences [Mi85], we can see that, for $k \geq 5$,

$$\mathrm{per}(I_k+P_k+P_k^2) = \left(\frac{1+\sqrt{5}}{2}\right)^k + \left(\frac{1-\sqrt{5}}{2}\right)^k + 2.$$

□

Some further special structures

In this Section, we first find the value of $\mathrm{per}(I+P+P^{\frac{n}{2}+1})$, for n even, and then, with similar techniques, we show that, if $\gcd(k,n)=2$, then $\mathrm{per}(I_n+P_n^k+P_n^{\frac{n}{2}})$ does not depend on k, and, in particular

$$\mathrm{per}(I_n+P_n^k+P_n^{\frac{n}{2}}) = \mathrm{per}(I+P+P^{\frac{n}{2}+1}) = 2\sqrt{2^n}+1.$$

Matrices of the form $I+P+P^{\frac{n}{2}+1}$.

Let n be even. We wish to evaluate the permanent of $I+P+P^{\frac{n}{2}+1}$. In this case, the matrix can be written as a 2×2 block matrix of the form

$$\begin{bmatrix} I_{n/2}+P_{n/2} & P_{n/2} \\ P_{n/2} & I_{n/2}+P_{n/2} \end{bmatrix},$$

i.e., the four blocks of size $\frac{n}{2}\times\frac{n}{2}$ are circulant matrices themselves. The proof of the following theorem takes advantage of the above property.

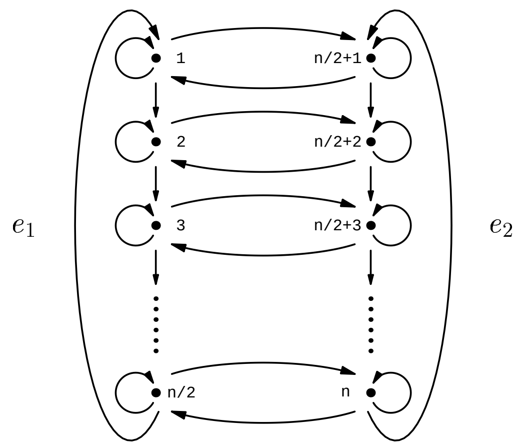


Figure 3.6: The digraph $D(B)$

Theorem 49 *Let n be even. Then*

$$\operatorname{per}(I + P + P^{\frac{n}{2}+1}) = 2\sqrt{2^n} + 1.$$

Proof. We first transform $I+P+P^{\frac{n}{2}+1}$ into a matrix with a more convenient structure. Indeed, we make the following transformation, which does not affect the value of the permanent:

$$B = \begin{bmatrix} P_{n/2}^{n/2-1} & O \\ O & P_{n/2}^{n/2-1} \end{bmatrix} \cdot \begin{bmatrix} I_{n/2} + P_{n/2} & P_{n/2} \\ P_{n/2} & I_{n/2} + P_{n/2} \end{bmatrix} = \begin{bmatrix} I_{n/2} + P_{n/2}^{n/2-1} & I_{n/2} \\ I_{n/2} & I_{n/2} + P_{n/2}^{n/2-1} \end{bmatrix}.$$

We now compute the number of cycle covers of the digraph $D(B)$ (see figure 3.6). We consider the edges $e_1 = (\frac{n}{2}, 1)$ and $e_2 = (n, \frac{n}{2}+1)$, and define

$$\begin{aligned} \mathcal{A} &= \{M \mid M \text{ is a cycle cover whose cycles do not contain neither } e_1 \text{ nor } e_2\}, \\ \mathcal{B}_1 &= \{M \mid M \text{ is a cycle cover whose cycles contain } e_1 \text{ but not } e_2\}, \\ \mathcal{B}_2 &= \{M \mid M \text{ is a cycle cover whose cycles contain } e_2 \text{ but not } e_1\}, \\ \mathcal{C} &= \{M \mid M \text{ is a cycle cover whose cycles contain both } e_1 \text{ and } e_2\}. \end{aligned}$$

We have that

$$\operatorname{per}(B) = |\mathcal{A}| + |\mathcal{B}|_1 + |\mathcal{B}|_2 + |\mathcal{C}|,$$

where $|\mathcal{X}|$ denote the cardinality of the set $\mathcal{X}$.

It is immediate to see that $|\mathcal{C}| = 1$. For the other cardinalities, we have:

Cycle covers in $\mathcal{A}$.

Each cycle cover in $\mathcal{A}$ does not contain edges of type $(i, i+1)$, for $1 \le i \le \frac{n}{2} - 1$ and $\frac{n}{2} + 1 \le i \le n - 1$. So the number of cycle covers is equal to that in the digraph $I_n + P_n^{\frac{n}{2}}$, i.e., $|\mathcal{A}| = 2^{\frac{n}{2}}$.

Cycle covers in $\mathcal{B}_1$ and $\mathcal{B}_2$.

For $\mathcal{B}_1$, each cycle cover corresponds to a simple path between node 1 and node $\frac{n}{2}$. In fact we can cover all the other nodes outside the path with their self-loops. The number of such paths is $2^{\frac{n}{2}-1}$.

Since $|\mathcal{B}|_1 = |\mathcal{B}|_2$ by symmetry, we finally obtain

$$\mathrm{per}(B) = 2^{\frac{n}{2}} + 2 \cdot 2^{\frac{n}{2}-1} + 1 = 2^{\frac{n}{2}+1} + 1.$$

□

Matrices of the form $I + P^k + P^{\frac{n}{2}}$.

Let n be even. We prove that, under certain conditions, the value of k does not affect the permanent of $I + P^k + P^{\frac{n}{2}}$.

Theorem 50 *Let $n = 2d$, where d is an odd integer and $A_n = I_n + P_n^k + P_n^{\frac{n}{2}}$. If $\gcd(k, n) = 2$ then*

$$\mathrm{per}(A_n) = 2 \cdot \sqrt{2^n} + 1.$$

Proof. We prove that the digraph $D(A_n)$ is isomorphic to the digraph of the matrix

$$B_n = \begin{bmatrix} I_{n/2} + P_{n/2}^{\frac{n}{2}-1} & I_{n/2} \\ I_{n/2} & I_{n/2} + P_{n/2}^{\frac{n}{2}-1} \end{bmatrix}.$$

Note that, since d is odd and $\gcd(k, n) = 2$, then $D(P_n^k)$ consists of two disjoint cycles of d nodes of the form [2]:

$$(1, 1 + k \bmod n, 1 + 2k \bmod n, \ldots, 1 + (d-1)k \bmod n)$$

and

$$(\frac{n}{2} + 1, \frac{n}{2} + 1 + k \bmod n, \frac{n}{2} + 1 + 2k \bmod n, \ldots, \frac{n}{2} + 1 + (d-1)k \bmod n).$$

Thus we can relabel the nodes of $D(A_n)$ so that it consists of

- two cycles of length d of the form

  $$(1, 1 + k \bmod n, 1 + 2k \bmod n, \ldots, 1 + (d-1)k \bmod n)$$

  and

  $$(2, 2 + k \bmod n, 2 + 2k \bmod n, \ldots, 2 + (d-1)k \bmod n),$$

- the self-loops,
- d cycles of length 2 of the form $(x, x + 1 \bmod n)$.

Consider $D(B_n)$. It consists of the self-loops, the two cycles of length d of the form $(1, 2, \ldots, d)$ and $(d+1, d+2, \ldots, n)$, and the d cycles of length 2 of the form $(x, x + \frac{n}{2} \bmod n)$. Again, we relabel the nodes of $D(B_n)$ as depicted in figure 3.7.

[2] Here we use the cycle notation for permutations.

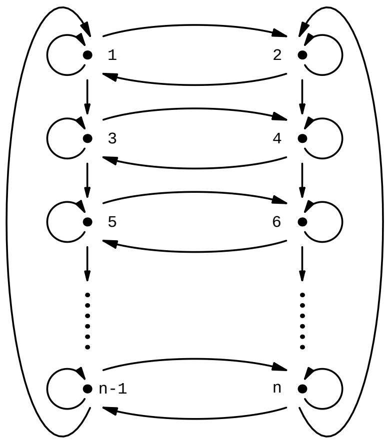


Figure 3.7: The relabeling of the nodes of $D(B)$.

The isomorphism we are going to define maps the two relabeled cycles of $D(B_n)$, $(1, 3, 5, \ldots, n-1)$ and $(2, 4, 6, \ldots, n)$, onto the two relabeled cycles of $D(A_n)$, $(1, 1+k \bmod n, \ldots, 1+(d-1)k \bmod n)$ and $(2, 2+k \bmod n, \ldots, 2+(d-1)k \bmod n)$, respectively.

We define a function $\phi : \mathbf{Z}_n \to \mathbf{Z}_n$ as follows:

$$\phi(x) = \begin{cases} \frac{x-1}{2} \cdot k + 1 \bmod n & \text{if } x \text{ is odd,} \\ \\ (\frac{x}{2} - 1) \cdot k + 2 \bmod n & \text{if } x \text{ is even.} \end{cases}$$

Now we prove that ϕ is an isomorphism between the two relabeled graphs $D(B_n)$ and $D(A_n)$. It is immediate to see that ϕ is surjective and injective. Let $(x, y) \in \mathrm{E}(D(B_n))$, with $x \neq y$. W.l.o.g. suppose that x is odd, so that $\phi(x) = \frac{x-1}{2} \cdot k + 1$. Then $x \in \{1, 3, 5, \ldots, n-1\}$, and y can be either $x+1$ or $x+2 \bmod n$. In the first case we have

$$\begin{aligned} \phi(x+1) &= \left(\frac{x+1}{2} - 1\right) \cdot k + 2 \\ &= \left(\frac{x-1}{2} \cdot k + 1\right) + 1 \\ &= \phi(x) + 1, \end{aligned}$$

which implies that $(\phi(x), \phi(x+1)) \in \mathrm{E}(D(A_n))$. In the second case we have

$$\begin{aligned} \phi(x + 2 \bmod n) &= \frac{x+1}{2} \cdot k + 1 \bmod n \\ &= \left(\frac{x-1}{2} + 1\right) \cdot k + 1 \bmod n \\ &= \left(\frac{x-1}{2} \cdot k + 1\right) + k \bmod n \\ &= \phi(x) + k \bmod n, \end{aligned}$$

which still implies that $(\phi(x), \phi(x + 2 \bmod n)) \in \mathrm{E}(D(A_n))$. The converse follows analogously, thus ϕ is a digraph isomorphism and then by Theorem 49 we obtain $\mathrm{per}(A_n) =$

$\mathrm{per}(B_n) = 2 \cdot \sqrt{2^n} + 1.$ □

Corollary 51 *Let $n = 2p$, where $p \neq 2$ is a prime number, and $A_n = I_n + P_n^k + P_n^{\frac{n}{2}}$. Then*

$$\mathrm{per}(A_n) = 2\sqrt{2^n} + 1.$$

Proof. If k is even, then $\gcd(k, 2p) = 2$, and the result follows from Theorem 50. If k is odd, then we have

$$\mathrm{per}(A_n) = \mathrm{per}\left(P_n^{\frac{n}{2}} A_n\right) = \mathrm{per}\left(I_n + P_n^{\frac{n}{2}} + P_n^{\frac{n}{2}+k}\right),$$

where $\frac{n}{2} + k$ is even. □

3.2.6 Circulants of the form $a \cdot I + b \cdot P^i + c \cdot P^j$

In this Section we aim at generalizing some of the preceding results to circulant matrices whose entries are no longer $(0, 1)$. The easy generalizations are those related to permutations of rows and columns of the given matrices as the following lemma shows.

Lemma 52 *Let $A_n = a \cdot I_n + b \cdot P_n^{d \cdot i} + c \cdot P_n^{d \cdot j}$, for $a, b, c \in \mathbf{R}$.*

1. *If $d \mid n$ then*
$$\mathrm{per}(A_n) = \mathrm{per}\left(a \cdot I_{\frac{n}{d}} + b \cdot P_{\frac{n}{d}}^i + c \cdot P_{\frac{n}{d}}^j\right)^d.$$
2. *If $\gcd(d, n) = 1$ then*
$$\mathrm{per}(A_n) = \mathrm{per}(a \cdot I_n + b \cdot P_n^i + c \cdot P_n^j).$$

where the subscripts indicate the matrix size.

Proof. It follows immediately from the proof of lemma 45 which shows that in case (1) there exists a permutation matrix Q such that $Q^{-1} \cdot A_n \cdot Q$ is the direct product of d identical matrices of the form $a \cdot I_{\frac{n}{d}} + b \cdot P_{\frac{n}{d}}^i + c \cdot P_{\frac{n}{d}}^j$ and in case (2) there exists a permutation matrix K such that $K^{-1} \cdot A_n \cdot K = a \cdot I_n + b \cdot P_n^i + c \cdot P_n^j$. □

As we will see in Section 3.6 we let $T_n[a, ib, jc]$ denote the matrix $a \cdot I_n + b \cdot Q^i + c \cdot (Q^T)^j$.

Lemma 53 *Let $a, b, c \in \mathbf{R}$.*

$$\mathrm{per}(a \cdot I_n + b \cdot P_n + c \cdot P_n^2) = b \cdot \mathrm{per}(T_{n-1}[b, 1a, 1c]) + 2ac \cdot \mathrm{per}(T_{n-2}[b, 1a, 1c]) + a^n + c^n.$$

Proof. We operate exactly as for $\mathrm{per}(I + P + P^2)$ using Laplace expansion on the same rows and columns (see Fig. 3.1). □

Lemma 54 *Let $a, b, c \in \mathbf{R}$ and $G[a,b,c](n) = \mathrm{per}(a \cdot I_n + b \cdot P_n + c \cdot P_n^2)$.*

$$\mathrm{per}(a \cdot I_n + b \cdot P_n^i + c \cdot P_n^{n-i}) = G[c,a,b]\left(\frac{n}{\gcd(n,i)}\right)^{\gcd(n,i)}.$$

Proof. First of all, since the permanent is invariant under arbitrary permutations of rows and columns it holds that

$$\mathrm{per}(a \cdot I_n + b \cdot P_n^i + c \cdot P_n^{n-i}) = \mathrm{per}(P_n^i(a \cdot I_n + b \cdot P_n^i + c \cdot P_n^{n-i})) = \mathrm{per}(c \cdot I_n + a \cdot P_n^i + b \cdot (P_n^i)^2).$$

Now we apply the generalized reduction lemma 52 as in theorem 48. □

Theorem 55 *Let n be even, $a, b, c \in \mathbf{R}$ and $A = a \cdot I + b \cdot P + c \cdot P^{\frac{n}{2}+1}$. Then*

$$\mathrm{per}(A) = (b^2 + c^2)^{\frac{n}{2}} + a^n + 2a^{\frac{n}{2}} \cdot \sum_{k=0}^{\frac{n}{2}-1} \binom{\frac{n}{2}-1}{k} c^k \cdot b^{\frac{n}{2}-k-1} \cdot b^{k \bmod 2} \cdot c^{k+1 \bmod 2}.$$

Proof. We proceed as in the proof of theorem 49. Let $A = a \cdot I + b \cdot P + c \cdot P^{\frac{n}{2}+1}$ and

$$C = \begin{bmatrix} P_{\frac{n}{2}}^{\frac{n}{2}-1} & O \\ O & P_{\frac{n}{2}}^{\frac{n}{2}-1} \end{bmatrix}.$$

Then let $B = C \cdot A$ as in 49. $D(B)$ is now a weighted digraph with the same structure as in figure 3.6 but with labels on its edges. W.l.o.g.[3] those are a on the vertical edges, on e_1 and on e_2. The self-loops get all b except for the last pair of nodes: $\frac{n}{2}$ and n, which get c. Finally the horizontal edges get all c except for $(\frac{n}{2}, n)$ and $(n, \frac{n}{2})$ which get both b. If we define the weight of a cycle cover as the product of the labels of the edges it is composed of then the permanent of B is thus given by the sum of the weights of all the cycle covers. We consider again the four disjoint sets of cycle covers:

$$\begin{aligned}
\mathcal{A} &= \{M \mid M \text{ is a cycle cover whose cycles do not contain neither } e_1 \text{ nor } e_2\},\\
\mathcal{B}_1 &= \{M \mid M \text{ is a cycle cover whose cycles contain } e_1 \text{ but not } e_2\},\\
\mathcal{B}_2 &= \{M \mid M \text{ is a cycle cover whose cycles contain } e_2 \text{ but not } e_1\},\\
\mathcal{C} &= \{M \mid M \text{ is a cycle cover whose cycles contain both } e_1 \text{ and } e_2\}.
\end{aligned}$$

We have that

$$\mathrm{per}(B) = W(\mathcal{A}) + W(\mathcal{B}_1) + W(\mathcal{B}_2) + W(\mathcal{C}),$$

where $W(\mathcal{X})$ denotes the sum of the weights of the elements in $\mathcal{X}$.

Clearly $W(\mathcal{C}) = a^n$ because the unique cycle cover which uses both e_1 and e_2 has weight a^n. For the other weights we have

Estimate of $W(\mathcal{A})$.
Again, vertical edges $(i, i+1)$ are never used by the cycle covers in $\mathcal{A}$ so $W(\mathcal{A}) = (b^2+c^2)^{\frac{n}{2}}$. To see this first notice that each cycle cover of this class covers the generic pair of nodes

[3] We have actually renumbered the nodes of $D(B)$ obtaining an isomorphic digraph.

$i, \frac{n}{2} + i$, for $i = 1, 2, \ldots, \frac{n}{2}$, either with the two self-loops $(i,i), (i+\frac{n}{2}, i+\frac{n}{2})$ or with the cycle $(i, i+\frac{n}{2})$ of length 2. Now, for $1 \le i \le \frac{n}{2} - 1$, the weight of the two self-loops is b^2 whereas the weight of the cycle of length 2 is c^2, instead, for $i = \frac{n}{2}$, it is the other way around. Thus the total weight will be $(b^2 + c^2)^{\frac{n}{2}}$.

Estimate of $W(\mathcal{B}_1)$.
Each cycle cover in $\mathcal{B}_1$ corresponds to a cycle passing through node 1. In fact the nodes outside the cycle can be covered by their self-loop and we cannot have two disjoint cycles. Each of these cycles must contain $\frac{n}{2}$ edges labeled with a, plus a variable number, say k, of edges labeled with c, and at most one edge labeled with b. Clearly k ranges between 0 and $\frac{n}{2} - 1$. Furthermore if k is even then the cycle cannot contain the edge $(n, \frac{n}{2})$ which is labeled with b. Thus we have the following cases:

- **Case** 1**:** k **is even.** The cycle has weight
$$a^{\frac{n}{2}} \cdot c^k,$$
and so the corresponding cycle cover has weight
$$a^{\frac{n}{2}} \cdot c^{k+1} \cdot b^{\frac{n}{2}-1-k}.$$

- **Case** 2**:** k **is odd.** The cycle has weight
$$a^{\frac{n}{2}} \cdot c^k \cdot b,$$
and so the corresponding cycle cover has weight
$$a^{\frac{n}{2}} \cdot c^k \cdot b^{\frac{n}{2}-k}.$$

Given a certain k there are $\binom{\frac{n}{2}}{k}$ cycles with the same weight then summing up for $k = 0, 1, \ldots, \frac{n}{2} - 1$ we obtain

$$W(\mathcal{B}_1) = a^{\frac{n}{2}} \cdot \sum_{k=0}^{\frac{n}{2}-1} \binom{\frac{n}{2}-1}{k} c^k \cdot b^{\frac{n}{2}-1-k} \cdot c^{k+1 \bmod 2} \cdot b^{k \bmod 2}.$$

Finally, it is immediate to see that $W(\mathcal{B}_1) = W(\mathcal{B}_2)$. This concludes the proof. □

Theorem 56 *Let* $n = 2d$, *where* d *is an odd integer and* $A_n = a \cdot I_n + b \cdot P_n^k + c \cdot P_n^{\frac{n}{2}}$. *If* $\gcd(k, n) = 2$ *then*

$$\mathrm{per}(A_n) = (a^2 + c^2)^{\frac{n}{2}} + b^n + 2b^{\frac{n}{2}} \cdot \sum_{k=0}^{\frac{n}{2}-1} \binom{\frac{n}{2}-1}{k} c^k \cdot a^{\frac{n}{2}-k-1} \cdot a^{k+1 \bmod 2} \cdot c^{k \bmod 2}.$$

Proof. The proof is the same as in the previous theorem. We have to consider the same digraph (see figure 3.6) but with a slightly different labeling. In particular all the horizontal edges get c, the self-loops get a and the vertical edges together with e_1 and e_2 get b. Then we apply the technique used in the preceding proof *mutatis mutandis.* □

3.3 Approximation issues

In Section 1.5.2 we described some approximation algorithms for the computation of the function permanent. In particular we saw that the Broder algorithm is an fpras for a class C of matrices provided that the ratio $r =\mid M_{n-1} \mid / \mid M_n \mid$ is bounded from above by a polynomial, say n^k. In fact the algorithm takes as input a $2n$-bipartite graph, a real number $0 \leq \epsilon \leq 1$, an integer parameter $t > r$ and returns the required approximation after a polynomial time of steps in n,$1/\epsilon$ and t. Note that even if the input graph has a polynomial ratio r, we need to know it in advance (or any upper bound of it) before running the algorithm.

It is natural to wonder whether or not the bipartite graphs associated with circulant matrices having three nonzeros per row have a "good ratio".

Now, consider the n square circulant $(0,1)$-matrix $A = I_n + P_n + P_n^d$ and its associated $2n$-bipartite graph $G[A]$. It is immediate to see that

$$\mid M_{n-1} \mid \leq n^2 \cdot \mathrm{per}(A(1|j^*)),$$

where $\mathrm{per}(A(1|j^*)) = \max\limits_{j:j\neq 1,2,d+1} \mathrm{per}(A(1|j))$. Moreover, for $j \neq 1,2,d+1$, Laplace expansion formula implies that

$$\mathrm{per}(A(1|j)) \leq \mathrm{per}(A + E_{1,j}),$$

where $E_{1,j}$ is the $(0,1)$-matrix whose entries are all zeros, except for $(1,j)$. Thus

$$\frac{\mid M_{n-1} \mid}{\mid M_n \mid} \leq n^2 \cdot \frac{\mathrm{per}(A(1|j^*))}{\mathrm{per}(A)} \leq n^2 \cdot \frac{\mathrm{per}(A + E_{1,j^*})}{\mathrm{per}(A)}.$$

We could formally prove that $\mathrm{per}(A(1|j^*))$ is not too large with respect to $\mathrm{per}(A)$ only for small values of d applying basically the same technique used to state the closed formulas for the exact permanent of the matrices of type $(0,1,2)$ and $(0,1,3)$. Unfortunately, we could not extend the approach to all the circulant matrices of type $(0,1,d)$. Nevertheless, for generic d's, and with n up to 17, we got experimental evidence that

$$\frac{\mathrm{per}(A + E_{1,j^*})}{\mathrm{per}(A)} \leq 2.$$

We have reported below some experimental results obtained using Ryser method. The following two tables contain the values of $\mathrm{per}(A+E_{1,j})$, for $A = I_n + P_n + P_n^d$, with $n = 13$ and $n = 14$. In particular column l of table 3.8 contains, in the second position, the value of $\mathrm{per}(I_n+P_n+P_n^{l+1})$, and, in the subsequent positions, the values of $\mathrm{per}(I_n+P_n+P_n^{l+1}+E_{1,j})$, for all j: $3 \leq j \leq n$, $j \neq l+2$, with $n = 13$. Table 3.9 is obtained repeating the same calculations but with $n = 14$. The entries of the first row, i.e., the couples of the form $(1,d)$, identify the matrix $I_n + P_n + P_n^d$.

Our feeling is that the class of circulant $(0,1)$-matrices with three nonzeros per row has ratios bounded from above by polynomials with small degree and then can be well approximated by the Broder algorithm. We hope to find a formal proof for this last conjecture.

1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10	1,11	1,12
523	185	159	185	185	523	185	185	159	185	523
613	238	199	227	220	580	232	221	199	232	613
580	225	202	220	220	549	227	224	199	235	580
560	220	202	225	224	544	222	232	202	227	560
549	221	196	235	235	560	221	238	199	224	549
544	220	199	222	238	613	225	220	199	222	544
544	222	199	220	225	613	238	222	199	220	544
549	224	199	238	221	560	235	235	196	221	549
560	227	202	232	222	544	224	225	202	220	560
580	235	199	224	227	549	220	220	202	225	580
613	232	199	221	232	580	220	227	199	238	613

Figure 3.8: Table of the values of $\mathrm{per}(I_{13} + P_{13} + P_{13}^d + E_{1,j})$. Each column starts with the couple $(1, d)$ and contains, in the second position, the value of $\mathrm{per}(I + P + P^d)$ and, in the subsequent positions, the values of $\mathrm{per}(I + P + P^d + E_{1,j})$, for $3 \leq j \leq n$, $j \neq d+1$.

3.4 Application of reductions

Karpinski and Dahlhaus' construction (see Section 2.1) appears interesting in the context of computing the permanent of generic circulants since we can prove the following.

Theorem 57 *Let* $A = \sum_{i=0}^{n-1} \alpha_i \cdot P^i$, $\alpha_i \in \{0,1\}, \alpha_0 = 1$, *a* $n \times n$ *circulant matrix. It is possible to construct a block circulant* $(0,1)$ *matrix* B *of order* $(2nk-n) \times (2nk-n)$, *being* k *the number of nonzeros per row of* A, *with at most three nonzeros per row and column such that*

$$\mathrm{per}(A) = \mathrm{per}(B).$$

Proof. Let $1 = i_0 < i_1 < i_2 < \ldots < i_{k-1}$ be the column indices of the nonzero elements of the first row of A. We define the function $r : \{i \mid \alpha_i = 1\} \to \mathbf{N}$ as $r(i_j) = j$. The idea of the proof is applying the Karpinski and Dahlhaus' construction (see the transformation shown in 2.1 and consider only the submatrix B) to A for a suitable choice of the permutation matrix P among the possible ones.

Since A is circulant then it has k nonzeros per row and column. Now, we consider the n-sequence of pairs $S = \{(x_i, y_i) \mid i = 1, 2, \ldots, n\}$ with $x_i = y_i = k, i = 1, 2, \ldots, n$ and the permutation matrix

$$P = \begin{bmatrix} P_0 & P_1 & \ldots & P_{n-1} \\ P_{n-1} & P_0 & \ldots & P_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ P_1 & P_2 & \ldots & P_0 \end{bmatrix},$$

where for $i = 0, 1, \ldots, n-1$, $P_i[u, v] = 1$ if and only if $\alpha_i = 1$ and $u = k - r(i), v = r(i) + 1$.

1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10	1,11	1,12	1,13
845	264	229	264	229	257	257	229	264	229	264	845
990	340	282	312	281	322	337	284	324	282	332	990
936	320	288	316	281	323	329	282	312	281	336	936
903	312	287	336	279	325	325	282	312	294	324	903
884	312	279	320	288	329	323	282	340	282	320	884
874	312	284	312	294	337	322	287	332	281	316	874
871	312	285	320	285	321	321	285	320	285	312	871
874	316	281	332	287	322	337	294	312	284	312	874
884	320	282	340	282	323	329	288	320	279	312	884
903	324	294	312	282	325	325	279	336	287	312	903
936	336	281	312	282	329	323	281	316	288	320	936
990	332	282	324	284	337	322	281	312	282	340	990

Figure 3.9: Table of the values of $\mathrm{per}(I_{14} + P_{14} + P_{14}^d + E_{1,j})$.

Then the Karpinski and Dahlhaus' construction applied to (S, P) gives $B = (B_{i,j}), 1 \leq i, j \leq n$, where each $B_{i,j}$ is of order $(2k-1) \times (2k-1)$ and B has at most three nonzeros per row and column.

We finally show that B is block circulant. But this follows from the fact that P is also block circulant. Furthermore we have that

$$B = \begin{bmatrix} B_0 & B_1 & \dots & B_{n-1} \\ B_{n-1} & B_0 & \dots & B_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ B_1 & B_2 & \dots & B_0 \end{bmatrix},$$

where each block B_i is equal to $B_{1,i+1}$ and

$$B_0 = \begin{bmatrix} 1 & & & & \\ 1 & 1 & & & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \end{bmatrix},$$

whereas for $i > 0$,

1. if $\alpha_i = 0$ then $B_i = O$ and
2. if $\alpha_i = 1$ then $B_i[u, v] = 1$ if and only if $u = k - r(i)$ and $v = k + r(i)$.

□

Observation. The above theorem states that the computation of the permanent of a $(0,1)$-circulant matrix is not harder than the computation of the permanent of a $(0,1)$ block circulant matrix with at most three nonzeros per row and column. □

3.5 An easy subclass of matrices with at most three nonzeros per row

In this Section we consider symmetric $(0,1)$ matrices with at most three nonzeros per row and column and with all ones along the main diagonal.

The computation of the permanent of these matrices turns out to be easy.

Lemma 58 *Let $A = I + H$ be a $(0,1)$ symmetric matrix with at most three nonzeros per row and column. Then* $\mathrm{per}(A)$ *can be computed in time $O(n)$.*

Proof. Let C_k^* denote the digraph which is obtained from an (undirected) cycle of length k by replacing each of its edges $\{x, y\}$ with two oppositely directed arcs (x, y) and (y, x). Furthermore let $D_k = D(T_k)$ denote the k-chain (see Section 3.6).

Let B be the matrix obtained from A by zeroing the lower triangular part including the main diagonal. Then each row of B has at most two nonzeros. This implies that the outdegree of each node x in the digraph $D(B)$ ranges between 0 and 2. Thus, by symmetry, $D(A)$ must be composed of a finite number of edge-disjoint strongly connected components, and each of those components can be only either of type C_k^* or of type D_k for some k.

In general, if $D(A)$ is composed of the edge-disjoint strong components

$$D_{k_1}, D_{k_2}, \ldots, D_{k_u}, C^*_{h_1}, C^*_{h_2}, \ldots, C^*_{h_v},$$

then it holds that

$$\mathrm{per}(A) = \prod_{i=1}^{u} \mathrm{per}(T_{k_i}) \prod_{i=1}^{v} \mathrm{per}(U_{h_i}), \tag{3.22}$$

where $U_{h_i} = I_{h_i} + P_{h_i} + P_{h_i}^{h_i - 1}$. Thus, $\mathrm{per}(A)$ can be computed according to the following stages:

i) find the strong components of $D(A)$;

ii) for each strong component, compute its permanent;

iii) compute $\mathrm{per}(A)$ according to (3.22).

Stage (i) takes time $O(|V| + |E|)$, i.e., $O(n)$. Stages (ii) and (iii) take time $O(n)$ overall. In fact both the computation of $\mathrm{per}(U_m)$ and $\mathrm{per}(T_m)$ take time $O(\log m)$ (see Section 3.6 and Section 3.2). Suppose now that the strong components to be considered are $W_{l_1}, W_{l_2}, \ldots, W_{l_k}$, the subscripts indicating their number of nodes. Then the number of operations is at most

$$\begin{aligned}
k - 1 + \sum_{i=1}^{k} \log(l_i) &= k - 1 + \log\left(\prod_{i=1}^{k} l_i\right) \\
&\leq k - 1 + \log\left(\frac{n}{k}\right)^k \\
&\leq k - 1 + k \cdot \log\frac{n}{k} \\
&\leq \left\lceil \frac{2n}{e} \cdot \log\left(\frac{e}{2}\right) + \frac{2n}{e} - 1 \right\rceil \\
&= O(n).
\end{aligned}$$

Thus, the overall computation takes time $O(n)$.

□

3.6 Toeplitz Matrices

3.6.1 Preliminaries

In this Subsection we deal with certain Toeplitz matrices. A Toeplitz matrix is characterized by the property that the elements along each diagonal are the same, i.e., its (i,j)-th entry depends only on $i-j$. So a Toeplitz matrix of order $n \times n$ is fully determined by the $2n-1$ elements of its first row and column. We denote the element $a_{i,j}$ of a Toeplitz matrix as a_{j-i}, so that, for $n=5$, we have

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 \\ a_{-1} & a_0 & a_1 & a_2 & a_3 \\ a_{-2} & a_{-1} & a_0 & a_1 & a_2 \\ a_{-3} & a_{-2} & a_{-1} & a_0 & a_1 \\ a_{-4} & a_{-3} & a_{-2} & a_{-1} & a_0 \end{bmatrix}.$$

Note that circulant matrices are a special case of Toeplitz matrices. A $(0,1)$ Toeplitz matrix A can also be represented as

$$A = \sum_{i=0}^{n-1} a_i \cdot Q^i + \sum_{i=1}^{n-1} b_i \cdot (Q^T)^i,$$

with $a_i, b_i \in \{0,1\}$ and

$$Q = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

3.6.2 Toeplitz Matrices of the form $I + Q^p + (Q^T)^p$

We begin with a very simple Toeplitz matrix. For $1 \le p \le n-1$, let $T_n[p] = I_n + Q_n^p + (Q_n^T)^p$, i.e.,

$$a_k = \begin{cases} 1 & \text{if } k=0 \text{ or } k=p \text{ or } k=-p, \\ 0 & \text{otherwise.} \end{cases}$$

Let $F(n)$ be the n-th Fibonacci number, i.e., $F(n) = F(n-1) + F(n-2)$ with $F(0)=0$, $F(1)=1$. The two following results give us an efficient formula to compute the permanent of $T_n[p]$ for any p. We first deal with the case $p=1$.

Lemma 59 *Let $T_n = T_n[1]$. Then*

$$\mathrm{per}(T_n) = F(n+1).$$

Proof. The proof is by transfinite induction on the number n.

Base. $\mathrm{per}(T_1) = 1 = F(2)$ and $\mathrm{per}(T_2) = 2 = F(3)$.

Induction. Suppose that the theorem holds for n, we shall prove it for $n+1$. We consider the digraph $D(T_{n+1})$. For its special shape we will refer to it as the *chain of* $n+1$ *nodes.* Now, let us count the number of its cycle covers. The chain $D(T_{n+1})$ can be thought of as constituted by the chain $D(T_n)$ connected to an additional node, labeled $n+1$, through two edges $(n, n+1)$ and $(n+1, n)$. Node $n+1$ has its self-loop, $(n+1, n+1)$, as all the other nodes of the chain.

Let M_1 be the set of cycle covers that involve that self-loop $(n+1, n+1)$ and M_2 be the set of those which don't. Then, clearly, $\mathrm{per}(T_{n+1}) = |M_1| + |M_2|$. It is evident that $|M_1| = \mathrm{per}(T_n)$. Furthermore, it holds that $|M_2| = \mathrm{per}(T_{n-1})$ because to cover node $n+1$ all the cycle covers in M_2 must use the two edges $(n, n+1)$ and $(n+1, n)$. So, they must contain the cycle $\{(n, n+1), (n+1, n)\}$ and thus they coincide with those which involve only nodes $1, 2, \ldots, n-1$.

Thus, by applying the induction hypothesis, we have that

$$\begin{aligned}
\mathrm{per}(T_{n+1}) &= \mathrm{per}(T_n) + \mathrm{per}(T_{n-1}) \\
&= F(n+1) + F(n) \\
&= F(n+2)
\end{aligned}$$

$\square$

We now generalize the previous formula to all the Toeplitz $T_n[p]$, for any $p \geq 1$.

Theorem 60 *Let $T_n[p]$ be the $(0,1)$ Toeplitz matrix defined above. Then*

$$\mathrm{per}(T_n[p]) = \hat{F}\left(\left\lfloor \frac{n}{p} \right\rfloor\right)^p \cdot \left(\frac{\hat{F}(\lceil \frac{n}{p} \rceil)}{\hat{F}(\lfloor \frac{n}{p} \rfloor)}\right)^{n - \lfloor \frac{n}{p} \rfloor \cdot p} \tag{3.23}$$

where $\hat{F}(n) = F(n+1)$, for $n \geq 1$.

Proof. Let $n = q \cdot p + r$, with $0 \leq r < p$. Clearly

$$\left\lfloor \frac{n}{p} \right\rfloor = q, \left\lceil \frac{n}{p} \right\rceil = q+1, n - \left\lfloor \frac{n}{p} \right\rfloor \cdot p = r.$$

Thus we can restate (3.23) as

$$\mathrm{per}(T_n[p]) = \hat{F}(q)^{p-r} \cdot \hat{F}(q+1)^r.$$

Note that the digraph $D(T_n[p])$ consists of the disjoint union of $d - r$ chains of q nodes and r chains of $q+1$ nodes for a total amount of p chains. To show this, let x, $1 \leq x \leq p$, be one of the first p nodes of $D(T_n[p])$. Then x belongs to the chain

$$x \leftrightarrow x+p \leftrightarrow x+2p \leftrightarrow \ldots \leftrightarrow x+k \cdot p,$$

where $x + k \cdot p \leq n$ and $x + (k+1) \cdot p > n$. The chain contains $k+1$ nodes. We can now estimate the value of k, which is the greatest integer such that $x + k \cdot p \leq n$. Thus $k = \lfloor \frac{n-x}{p} \rfloor$. Substituting $n = q \cdot p + r$, we get $k = \lfloor q + \frac{r}{p} - \frac{x}{p} \rfloor$, from which we finally obtain

$$k = \begin{cases} q & \text{for } x = 1, 2, \ldots, r, \\ q-1 & \text{for } x = r+1, r+2, \ldots, p. \end{cases}$$

We can conclude that there are r disjoint chains of $k+1 = q+1$ nodes and $p-r$ chains of $k+1 = q$ nodes. Furthermore, since these p chains involve $(p-r) \cdot q + r \cdot (q+1) = n$ nodes overall and considering the degree of each node, then those chains are exactly the whole digraph $D(T_n[p])$. The statement then follows by the application of lemma 59 to compute the permanent of the chains. $\square$

It is worth noting that theorem 60 implies that there exists a permutation matrix Q_n such that

$$Q_n^{-1} \cdot T_n[p] \cdot Q_n = B^{(1)} \oplus B^{(2)} \oplus \ldots \oplus B^{(r)} \oplus C^{(1)} \oplus C^{(2)} \oplus \ldots \oplus C^{(p-r)},$$

where $B^{(i)} = T_{q+1}[1]$, for $i = 1, 2, \ldots, r$, and $C^{(j)} = T_q[1]$, for $j = 1, 2, \ldots, p-r$.

3.6.3 Toeplitz Matrices of the form $I + Q^i + (Q^T)^j$

Let $T_n[i,j] = I_n + Q_n^i + (Q_n^T)^j$, for $i \neq j$. We start with the following theorem which doesn't give us explicit formulas for the permanent of this class of Toeplitz matrices but provides an immediate algorithm to compute the permanent reducing it to the determinant.

Theorem 61 *Let us consider the $n \times n$ matrix $I + Q^i + (Q^T)^j$. Then*

$$\operatorname{per}(I + Q^i + (Q^T)^j) = \begin{cases} \det(I + Q^i + (Q^T)^j) & \textit{if } \ (i+j)/\gcd(i,j) \ \textit{ is odd,} \\ \\ \det(I - Q^i + (Q^T)^j) & \textit{otherwise.} \end{cases}$$

Proof. W.l.o.g. we assume that $i < j$. Let $A = I + Q^i + (Q^T)^j$, and consider $D(A)$. Without loss of generality we can restrict ourselves to the case for which there are cycles in $D(A)$. Indeed the absence of cycles implies that the nonzero entries off the main diagonal do not contribute to the value of the permanent.

We now analyze the length of the cycles in $D(A)$. Let x be a node belonging to a cycle γ, and let $l = \text{length}(\gamma)$. Starting from x we reach the next node along γ either by adding i or by subtracting j as long as $x + i \leq n$ (resp. $x - j \geq 1$). Thus, γ can be represented by a sequence of integers

$$d_1, d_2, \ldots, d_l,$$

where d_k is equal to either i or $-j$, for $k = 1, 2, \ldots, l$ and the following properties hold:

1. $\sum_{k=1}^{l} d_k = 0,$

2. $\sum_{k=u}^{v} d_k \neq 0, \ \forall \ 1 \leq u < v \leq l.$

Let $m = \mathrm{lcm}(i, j)$ and h, k be the two positive integers such that

$$m = h \cdot i = k \cdot j. \tag{3.24}$$

By 3.24 we must have

$$x + c \cdot h \cdot i - c \cdot k \cdot j = x,$$

for any positive integer c. Then length$(\gamma) = c \cdot (h + k)$, for some $c \geq 1$. In other words, 3.24 implies that the lengths of the cycles must be multiples of $h + k$. Before proceeding we show that for each cycle γ, length(γ) must be exactly $h + k$ (i.e., $c = 1$).

Basically we want to prove that any sequence of $c \cdot h$ elements i and $c \cdot k$ elements $-j$ satisfies condition 2 only if $c = 1$. Let $s = h + k$ and $c > 1$, suppose we are given the sequence

$$d_1, d_2, \ldots, d_{cs},$$

satisfying condition 1, i.e., $\sum_{k=1}^{cs} d_k = 0$. We know by 3.24 that the sequence contains exactly $c \cdot h$ elements equal to i and $c \cdot k$ elements equal to $-j$. Let $S_u = \sum_{k=u}^{u+s-1} d_k$. We introduce variables h_u and k_u, for $u = 1, 2, \ldots, cs - s + 1$, such that

$$S_u = h_u \cdot i - k_u \cdot j.$$

Let

$$\begin{aligned} \Delta^{(t)} h &= h_t - h, \\ \Delta^{(t)} k &= k_t - k. \end{aligned}$$

We now consider the sequence

$$S_1, S_{s+1}, S_{2s+1}, \ldots, S_{(c-1)s+1}.$$

Since there are $c \cdot h$ elements equal to i, then

$$\sum_{t=0}^{c-1} \Delta^{(ts+1)} h = 0.$$

W.l.o.g. suppose that, for some t', $\Delta^{(t's+1)} h \neq 0$ and $\Delta^{(t's+1)} h < 0$. Then there must be another t'' such that $\Delta^{(t''s+1)} h > 0$. Suppose now that $t' < t''$, and consider the sequence

$$S_{t's+1}, S_{t's+2}, S_{t's+3}, \ldots, S_{t''s+1}.$$

Since for each u, $\Delta^{(u+1)} h$ is either $\Delta^{(u)} h$ or $\Delta^{(u)} h \pm 1$, then $\Delta^{(u)} h$ must be zero for some u, $t's + 1 \leq u \leq t''s + 1$, and thus condition 2 is violated.

Now, note that $(i+j)/\gcd(i,j)$ is odd if and only if $h+k$ is odd, so following the hypothesis, we consider two cases:

Case 1*: $h + k$ is odd.*

In this case all the cycles have odd length and the theorem follows by a result of Bassett, Maybee and Quirk (1968).

Case 2: $h+k$ is even.

In this case both h and k must be odd by the property of lcm. Indeed if they were both even we would have

$$\frac{h}{2}\cdot i = \frac{k}{2}\cdot j < \operatorname{lcm}(i,j),$$

which contradicts the property of lcm. Thus, after changing the sign to the entries of the main diagonal elements and to the entries of one of the two other diagonals, each cycle gets weight -1 (recall that we measure the weight of a path by multiplying altogether the weights of the edges belonging to it). Then we can apply the Bassett, Maybee and Quirk theorem to guarantee that $\operatorname{per}(A) = |\det(-I + Q^i - (Q^T)^j)|$. Finally, since the conversion matrix is unique up to multiplication of rows or columns by -1, we obtain $\operatorname{per}(A) = \det(I - Q^i + (Q^T)^j)$ as claimed.

□

Let $F_j(n)$ be the n-th number of the sequence $F_j(k) = F_j(k-1) + F_j(k-j-1)$, with $F_j(j) = 2$, and $F_j(h) = 1$, for $h < j$. In particular $F_1(n) = F(n+1)$, i.e., the $(n+1)$-th Fibonacci number.

Theorem 62

$$\begin{array}{lrcl} (1) & \operatorname{per}(T_n[1,j]) & = & F_j(n). \\ (2) & \operatorname{per}(T_n[i,ki]) & = & \operatorname{per}\left(T_{\lceil \frac{n}{i} \rceil}[1,k]\right)^{n-\lfloor \frac{n}{i} \rfloor i} \operatorname{per}\left(T_{\lfloor \frac{n}{i} \rfloor}[1,k]\right)^{i-n+\lfloor \frac{n}{i} \rfloor i}. \end{array}$$

Proof.
Case 1. The proof follows by applying Laplace expansion.

Case 2. The proof is easily seen by reducing the matrix to block Hessenberg form, by means of a symmetric row and column permutation defined as

$$(1, i+1, 2i+1, \ldots, \left(\left\lfloor \frac{n}{i} \right\rfloor - 1\right) i+1, 2, i+2, 2i+2, \ldots, \left(\left\lfloor \frac{n}{i} \right\rfloor - 2\right) i+2, 3, \ldots).$$

□

Expression (2) of Theorem 62 for $\operatorname{per}(T_n[i,ki])$ can be generalized. Indeed, if we let $g = \gcd(i,j)$, then

$$\operatorname{per}(T_n[i,j]) = \operatorname{per}\left(T_{\lceil \frac{n}{g} \rceil}\left[\frac{i}{g},\frac{j}{g}\right]\right)^{n-\lfloor \frac{n}{g} \rfloor g} \operatorname{per}\left(T_{\lfloor \frac{n}{g} \rfloor}\left[\frac{i}{g},\frac{j}{g}\right]\right)^{g-n+\lfloor \frac{n}{g} \rfloor g}.$$

This equality is obtained by applying to the matrix $T_n[i,j]$ the row and column permutation defined as

$$(1, g+1, 2g+1, \ldots, \left(\left\lfloor \frac{n}{g} \right\rfloor - 1\right) g+1, 2, g+2, 2g+2, \ldots, \left(\left\lfloor \frac{n}{g} \right\rfloor - 2\right) g+2, 3, \ldots).$$

This leaves still open the problem of finding an explicit expression for the permanent of $T_n[i,j]$, when i and j are relatively prime.

3.6.4 Toeplitz Matrices of the form $Q^i + Q^j + (Q^T)^k$

A more general type of $(0,1)$ Toeplitz matrix with at most three nonzeros per row takes the form $Q_n^i + Q_n^j + (Q_n^T)^k$. For these matrices we can obtain results similar to those for matrices of the form $I_n + Q_n^h + (Q_n^T)^k$. In fact the matrix $A_n = Q_n^i + Q_n^j + (Q_n^T)^k$

1. is a submatrix of $I_{n+i} + Q_{n+i}^{j-i} + (Q_{n+i}^T)^{k+i}$,
2. contains as a submatrix $I_{n-i} + Q_{n-i}^{j-i} + (Q_{n-i}^T)^{k+i}$.

From (1) we can deduce that A_n is convertible, because the bipartite graph associated with $I_{n+i} + Q_{n+i}^{j-i} + (Q_{n+i}^T)^{k+i}$ is planar, which implies that the bipartite graph associated with A_n is also planar.
From (2), and by observing that the remaining submatrices of A_n have exactly a one per row or column, one can see that the actual conversion can be efficiently computed by adapting Brualdi and Shader's algorithm (see [BS95]) to this special case.

3.6.5 Toeplitz Matrices of the form $a \cdot I + b \cdot Q^i + c \cdot (Q^T)^j$

We start with the main brick of this Section: a formula for the permanent of $a \cdot I_n + b \cdot Q_n + c \cdot Q_n^T$. In the following $T_n[a, ib, jc]$ will denote the Toeplitz matrix $a \cdot I_n + b \cdot Q_n^i + c \cdot (Q_n^T)^j$.

Lemma 63 *Let $a, b, c \in \mathbf{R}$ and $F[a,b,c](n)$ (F for short) be the n-th element of the sequence recursively defined as $F(n) = a \cdot F(n-1) + bc \cdot F(n-2)$, $F(0) = 0, F(1) = a, F(2) = a^2 + bc$. Then*

$$\operatorname{per}(T_n[a, 1b, 1c]) = F(n).$$

Proof. Immediate, using Laplace expansion. □

Theorem 64 *Let $F(n)$ be the n-th element of the above parameterized sequence.*

$$\operatorname{per}(T_n[a, ib, ic]) = F\left(\left\lfloor \frac{n}{i} \right\rfloor\right)^i \cdot \left(\frac{F(\lceil \frac{n}{i} \rceil)}{F(\lfloor \frac{n}{i} \rfloor)}\right)^{n - \lfloor \frac{n}{i} \rfloor \cdot i}. \tag{3.25}$$

Proof. Let $n = i \cdot q + r$, with $0 \leq r < i$. The proof of theorem 60 allows us to find a permutation matrix Q_n such that $Q_n \cdot T_n[a, ib, ic] \cdot Q_n^{-1}$ consists of the direct sum of r matrices of the form $T_{q+1}[a, 1b, 1c]$ and $i - r$ matrices of the form $T_q[a, 1b, 1c]$. Thus the result follows.

□

Now, we solve explicitly the above recurrence equation defining the parameterized sequence $\{F(n)\}$. To that purpose we first consider another sequence defined by the following equations: $u_n = a \cdot u_{n-1} + b \cdot c \cdot u_{n-2}, u_0 = 0, u_1 = a$. If $U(z)$ is the generating function of the sequence $\{u_n\}$ and $G(z)$ is the generating function of the task sequence $\{F(n)\}$ then with easy manipulations we can show that $G(z) = U(z) + (bc/a)z \cdot U(z)$, which would imply that $F(n) = u_n + (bc/a) \cdot u_{n-1}$, for all n. This reasoning allows us to focus on the sequence $\{u_n\}$.

Lemma 65 *Let $\{u_n\}$ be the sequence defined as above and let $\rho_1 = (a+\sqrt{a^2+4bc})/2$ and $\rho_2 = (a-\sqrt{a^2+4bc})/2$. Then*

$$u_n = \begin{cases} \frac{a\sqrt{a^2+4bc}}{a^2+4bc} \cdot (\rho_1)^n - \frac{a\sqrt{a^2+4bc}}{a^2+4bc} \cdot (\rho_2)^n & \text{if } a^2+4bc > 0, \\ \left[-\frac{a^2}{2bc}(n-1)+2\right] \cdot (\frac{a}{2})^n & \text{if } a^2+4bc = 0, \\ \frac{-2|a|\sqrt{-a^2-4bc}}{a^2+4bc}(\sqrt{-bc})^n \cdot \cos(n \cdot \theta - \mathrm{sign}(a) \cdot \frac{\pi}{2}) & \text{if } a^2+4bc < 0, \end{cases}$$

where θ is the amplitude of the complex number ρ_1, in case its discriminant is negative.

Proof. We apply the standard technique in [GKP95] to solve recurrences. Thus we first derive the generating function

$$U(z) = \frac{a \cdot z}{1 - a \cdot z - bc \cdot z^2}$$

of the sequence $\{u_n\}$. It is immediate to verify that ρ_1 and ρ_2 are the roots of the reciprocal polynomial of the denominator of $U(z) = P(z)/Q(z)$. We focus on their discriminant and consider the three possible cases.

Case 1*:* $a^2 + 4bc > 0$.
The two roots are real and distinct. The result follows from a direct application of the Rational Expansion Theorem for Distinct Roots in [GKP95]. In particular we get

$$u_n = c_1 \cdot (\rho_1)^n + c_2 \cdot (\rho_2)^n,$$

where

$$\begin{aligned} c_1 &= -\rho_1 \cdot \frac{P(1/\rho_1)}{Q'(1/\rho_1)} &= \frac{a\sqrt{a^2+4bc}}{a^2+4bc}, \\ c_2 &= -\rho_2 \cdot \frac{P(1/\rho_2)}{Q'(1/\rho_2)} &= -\frac{a\sqrt{a^2+4bc}}{a^2+4bc}. \end{aligned}$$

Case 2*:* $a^2 + 4bc = 0$.
The two roots are real and coincident. The result follows from a direct application of the General Expansion Theorem for Rational Generating Functions in [GKP95]. In particular we get

$$u_n = f(n) \cdot \rho^n,$$

where $\rho = a/2$ and $f(n) = r \cdot n + s$ and

$$r = \frac{(-\rho)^2 \cdot P(1/\rho) \cdot 2}{Q^{(2)}(1/\rho)} = -\frac{a^2}{2bc}.$$

To get s it is sufficient to force the sequence to assume the value a for $n = 1$. This leads to the equation

$$\left(-\frac{a^2}{2bc} + s\right) \cdot \frac{a}{2} = a,$$

from which

$$s = \frac{a^2}{2bc} + 2.$$

Case 3*:* $a^2 + 4bc < 0$
The two roots are complex and conjugated. We proceed as in case 1. We derive the two coefficients c_1 and c_2 which are complex numbers. To get a sequence whose general term is a real we need to manipulate the complex numbers to get rid of the imaginary parts.

$$\begin{aligned}
c_1 &= \frac{a\sqrt{-a^2-4bc}}{a^2+4bc} \cdot i &= \frac{|a|\sqrt{-a^2-4bc}}{|a^2+4bc|} \cdot e^{-\mathrm{sign}(a)\frac{\pi}{2}\cdot i}, \\
c_2 &= \frac{a\sqrt{-a^2-4bc}}{a^2+4bc} \cdot i &= \frac{|a|\sqrt{-a^2-4bc}}{|a^2+4bc|} \cdot e^{\mathrm{sign}(a)\frac{\pi}{2}\cdot i}, \\
\rho_1 &= \frac{a+i\cdot\sqrt{-a^2-4bc}}{2} &= \sqrt{-bc}\cdot e^{i\cdot\theta}, \\
\rho_2 &= \frac{a-i\cdot\sqrt{-a^2-4bc}}{2} &= \sqrt{-bc}\cdot e^{-i\cdot\theta}.
\end{aligned}$$

Thus

$$u_n = \frac{|\,a\,|\,\sqrt{-a^2-4bc}}{|\,a^2+4bc\,|} \cdot (\sqrt{-bc})^n \cdot \left[e^{i\cdot(n\theta - \mathrm{sign}(a)(\pi/2))} + e^{-i\cdot(n\theta-\mathrm{sign}(a)(\pi/2))}\right].$$

Using Euler formulas it holds that $e^{ix} + e^{-ix} = 2\cos(x)$ and then

$$u_n = \frac{2\,|\,a\,|\,\sqrt{-a^2-4bc}}{|\,a^2+4bc\,|} \cdot (\sqrt{-bc})^n \cdot \cos(n\theta - \mathrm{sign}(a)(\pi/2)).$$

□

Corollary 66

$$\mathrm{per}(I - Q + Q^T) = \frac{2}{3}\sqrt{3}\left[\cos\left(n\cdot\frac{\pi}{3}-\frac{\pi}{2}\right) - \cos\left((n-1)\cdot\frac{\pi}{3}-\frac{\pi}{2}\right)\right]$$

for all $n \geq 1$.

Proof. Here $a = 1, b = -1$ and $c = 1$. Since $a^2 + 4bc = -3 < 0$ we fall in case 3 of theorem 65. The amplitude θ is equal to $\pi/3$ and thus the result follows. □

Theorem 67 *Let* $a, b, c \in \mathbf{R}$.

$$\mathrm{per}(a\cdot I + b\cdot Q^i + c\cdot(Q^T)^j) = \begin{cases} \det(a\cdot I + b\cdot Q^i + c\cdot(Q^T)^j) & \text{if } \frac{i+j}{\gcd(i,j)} \text{ is odd} \\ \det(a\cdot I - b\cdot Q^i + c\cdot(Q^T)^j) & \text{otherwise} \end{cases}$$

Proof. Theorem 61 gives us the conversion matrix of $I + Q^i + (Q^T)^j$. Then we apply theorem 27. $\square$

Observation. Actually, it holds an even stronger result since we can consider also arbitrary matrices with three nonzero diagonals and treat them the same way simply by changing (according to the parity of $(i+j)/\gcd(i,j)$) the sign to the entries of the diagonal corresponding to Q^i and then computing the determinant. The properties of convertible matrices guarantee the correctness of the operation. $\square$

Let $F_j[a,b,c](n)$ ($F_j(n)$ for short) the n-th number of the sequence $F_j(k) = a \cdot F_j(k-1) + c \cdot b^j \cdot F_j(n-j-1)$, $F_j(0) = 0, F_j(j+1) = a^{j+1} + c \cdot b^j, F_j(h) = a^j$, for $h \leq j$.

Theorem 68

$$\begin{array}{lll} (1)\ \ \mathrm{per}(T_n[a,1b,jc]) & = & F_j[a,b,c](n). \\ (2)\ \ \mathrm{per}(T_n[a,ib,(k \cdot i)c]) & = & \mathrm{per}\left(T_{\lceil \frac{n}{i} \rceil}[a,1b,kc]\right)^{n - \lfloor \frac{n}{i} \rfloor i} \mathrm{per}\left(T_{\lfloor \frac{n}{i} \rfloor}[a,1b,kc]\right)^{i-n+\lfloor \frac{n}{i} \rfloor i}. \end{array}$$

Proof.
Case 1. The proof follows by applying Laplace expansion.
Case 2. We use the same reduction as in theorem 62 to get a matrix composed of $n - \lfloor \frac{n}{i} \rfloor i$ matrices of the form $T_{\lceil \frac{n}{i} \rceil}[a,1b,kc]$ and $i - n + \lfloor \frac{n}{i} \rfloor i$ matrices of the form $T_{\lfloor \frac{n}{i} \rfloor}[a,1b,kc]$. $\square$

Theorem 69

$$\mathrm{per}(T_n[a,ib,(n-i)c]) = \left(a^{\frac{n}{\gcd(n,i)}} + b^{\frac{n-i}{\gcd(n,i)}} \cdot c^{\frac{i}{\gcd(n,i)}}\right)^{\gcd(n,i)}.$$

Proof. Let $D = D(T_n[a,ib,(n-i)c])$. We consider two cases.
Case 1*:* $\gcd(n,i) = 1$.
In this case the digraph D is composed of a single cycle containing exactly $n - i$ edges labeled with b and the remaining i edges labeled with c. Also the self-loops of the nodes are labeled with a, thus the total weight of the two only cycle covers is $a^n + b^{n-i} \cdot c^i$.

Case 2*:* $\gcd(n,i) > 1$.
Now D is composed of the disjoint union of $d = \gcd(n,i)$ weighted digraphs D_k, for $k = 1, 2, \ldots, \gcd(n,i)$. Moreover those digraphs are all equal. They all have $\frac{n}{d}$ nodes and are composed of a single cycle of length $\frac{n}{d}$. Also the nodes have self-loops labeled with a and the single cycle has exactly $\frac{n}{d} - \frac{i}{d}$ edges labeled with b and the remaining edges $\frac{i}{d}$ labeled with c. Thus the total weight of the two only cycle covers of D_k is equal to $a^{\frac{n}{d}} + b^{\frac{n}{d} - \frac{i}{d}} \cdot c^{\frac{i}{d}}$, for all k. $\square$

3.7 Conjectures and Open questions

As a support to our investigations on $(0,1)$-circulant matrices of type $(0,i,j)$, we have performed a number of experiments on matrices of size up to 31, from which we have derived several *structural* questions on their permanents.

Some of the unresolved questions led us to state the following conjecture and observation.

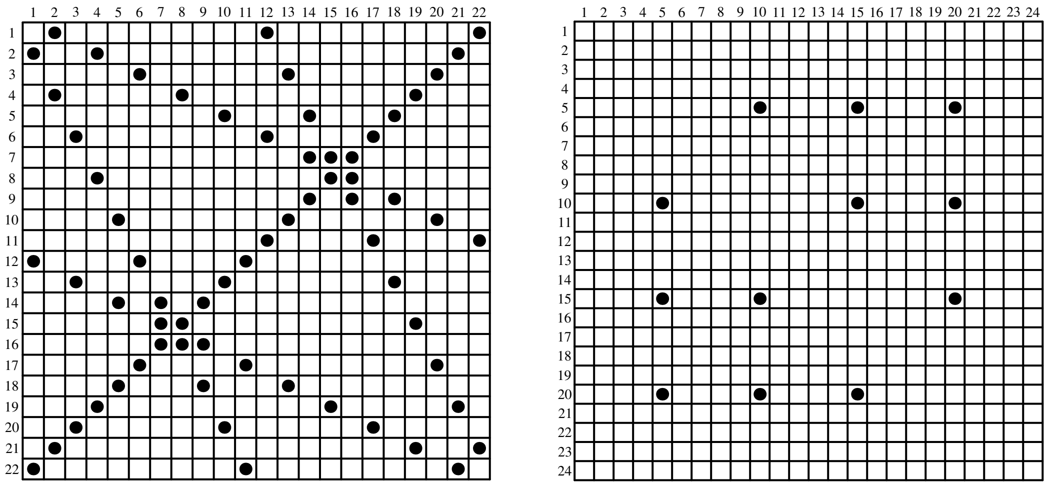


Figure 3.10: Pattern of maximal values of $\mathrm{per}(I + P^i + P^j)$, for $n = 23$ and $n = 25$. The entry (h,k) contains a filled circle if $\mathrm{per}(I + P^h + P^k) \geq \mathrm{per}(I + P^i + P^j)$, for all pairs (i,j).

Conjecture. Because of symmetry, we assume, w.l.o.g., that $i < j$. The set of pairs of indices that maximize $\mathrm{per}(I_n + P_n^i + P_n^j)$ is given by

$$\begin{cases} \{\frac{n}{3}, \frac{2n}{3}\} & \text{if 3 divides } n \\ \{\frac{n}{4}, \frac{n}{2}\}, \{\frac{n}{4}, \frac{3n}{4}\}, \{\frac{n}{2}, \frac{3n}{4}\} & \text{if 4 divides } n \text{ and 3 does not divide } n \\ \{\frac{n}{p}i, \frac{n}{p}j\} \text{ for } i+j=n, i=2j, i=2j-n & \text{otherwise,} \end{cases}$$

where p is the smallest prime factor of n greater than or equal to 5 (see figure 3.10). As a consequence, we have that the cardinality of the set of pairs that maximize $\mathrm{per}(I+P^i+P^j)$ is equal to m, where

$$m = \begin{cases} 1 & \text{if 3 divides } n \\ 3 & \text{if 4 divides } n \text{ and 3 does not divide } n \\ \frac{3}{2}(p-1) & \text{otherwise.} \end{cases}$$

□

Observation. We have seen in Section 3.2.5 some examples of different matrices of the type $I + P^i + P^j$ with the same permanent. If n is prime, there are only a few different values for $\mathrm{per}(I+P^i+P^j)$, by varying i and j (see figure 3.11). By analyzing the bipartite

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1		A	B	C	C	C	C	B	C	C	B	A	B	C	C	B	C	C	C	C	B	A
2	A		B	A	C	B	C	C	B	C	C	C	C	C	C	B	C	C	B	C	A	B
3	B	B		C	C	A	C	C	B	B	C	C	A	C	C	B	B	C	C	A	C	C
4	C	A	C		C	B	C	A	B	C	C	B	C	C	B	C	C	B	A	C	B	C
5	C	C	C	C		C	C	C	B	A	B	C	B	A	B	C	B	A	B	C	C	C
6	C	B	A	B	C		C	C	B	C	B	A	C	C	C	C	A	B	C	B	C	C
7	C	C	C	C	C	C		B	B	B	C	C	C	A	A	A	C	C	C	B	B	B
8	B	C	C	A	C	C	B		C	C	C	B	B	C	A	A	C	B	B	C	C	C
9	C	B	B	B	B	B	B	C		C	C	C	C	A	C	A	C	A	C	C	C	C
10	C	C	B	C	A	C	B	C	C		B	C	A	C	B	C	C	B	C	A	C	B
11	B	C	C	C	B	B	C	C	C	B		A	C	C	B	C	A	C	B	C	C	A
12	A	C	C	B	C	A	C	B	C	C	A		B	C	C	C	B	B	C	C	C	B
13	B	C	A	C	B	C	C	B	C	A	C	B		C	C	B	C	A	C	B	C	C
14	C	C	C	C	A	C	A	C	A	C	C	C	C		C	B	B	B	B	B	B	C
15	C	C	C	B	B	C	A	A	C	B	B	C	C	C		B	C	C	A	C	C	B
16	B	B	B	C	C	C	A	A	A	C	C	C	B	B	B		C	C	C	C	C	C
17	C	C	B	C	B	A	C	C	C	C	A	B	C	B	C	C		C	B	A	B	C
18	C	C	C	B	A	B	C	B	A	B	C	B	A	B	C	C	C		C	C	C	C
19	C	B	C	A	B	C	C	B	C	C	B	C	C	B	A	C	B	C		C	A	C
20	C	C	A	C	C	B	B	C	C	A	C	C	B	B	C	C	A	C	C		B	B
21	B	A	C	B	C	C	B	C	C	C	C	C	C	B	C	C	B	C	A	B		A
22	A	B	C	C	C	C	B	C	C	B	A	B	C	C	B	C	C	C	C	B	A	

Figure 3.11: Permanents of $I + P^i + P^j$, when $n = 23$. The entry (h, k) contains the value of the permanent of $I + P^h + P^k$. These permanents take only 3 different values, namely $A = 64081$, $B = 7225$ and $C = 4097$.

graph $G[I + P^i + P^j]$, we have tried to understand this phenomenon. For instance, if $n = 31$, $D(n, i, j)$ can take all the 15 odd values $3, 5, 7, \ldots, 31$. By multiplying $I + P^i + P^j$ by P^{n-i} and P^{n-j} we have obtained the following "classes" of different values of D for which the permanent is the same:

$$\{11\}, \{3, 31\}, \{5, 21, 29\}, \{7, 15, 19\}, \{9, 13, 17\}, \{23, 25, 27\}.$$

Experimentally we have checked that the actual *partition* is:

$$\{11\}, \{3, 31\}, \{5, 21, 29\}, \{7, 15, 19, 23, 25, 27\}, \{9, 13, 17\},$$

which shows that the problem hides further symmetries.

The knowledge of all these symmetries can be computationally very useful. In fact one could try to reduce the computation of $\mathrm{per}(I_n + P_n^i + P_n^j)$ to that of the permanent of a matrix of type $(0, i', j')$, where i' and j' satisfy the requirements for which Theorem 44 provides, e.g., a polynomial time algorithm (see Table 3.1).

n	#	$i' + j'$	n	#	$i' + j'$	n	#	$i' + j'$
5	1	2	31	6	6	67	12	9
7	2	3	37	7	7	71	12	9
11	2	3	41	7	7	73	13	9
13	3	4	43	8	7	79	14	10
17	3	4	47	8	7	83	14	9
19	4	5	53	9	7	89	15	10
23	4	5	59	10	8	97	17	11
29	5	5	61	9	11	101	17	11

Table 3.1: For n prime, $3 \leq n \leq 101$, we report, in the second column, the number of different values taken by $\mathrm{per}(I_n + P_n^i + P_n^j)$. In the third column we indicate the maximum value of $i' + j'$ (with reference to the notation of Theorem 44) that one has to consider in order to compute any of the possible different values taken by the permanent. Note that the values reported in both columns are actually upper bounds, since they are not outcomes of experimental results, but are obtained from known relationships between the values of $\mathrm{per}(I + P^i + P^j)$, as i and j vary.

$\square$

Summarizing, we would like to raise the following questions:

- Which are the properties of the pairs (i, j) that maximize the value of the permanent of a $(0, 1)$-circulant matrix of type $(0, d_1, d_2)$? What is the number of such pairs?
- How many different values can the permanents of $(0, 1)$-circulant matrices of type $(0, d_1, d_2)$ take?
- Is there an algorithm for the computation of the permanent of a $(0, 1)$-circulant matrix of type $(0, 1, t)$ that runs in polynomial time for all the values of t? Is this the case at least if n is prime?

Chapter 4

Computing the Permanent via Groebner Bases computation

4.1 Introduction

Certain computational problems can be rewritten in terms of systems of equations, so that their associated decision problem consists of asking for the existence of a solution to the system, whereas the corresponding counting problem consists of asking for the number of solutions. In this chapter we build upon the above idea to develop a method for solving hard counting problems which consists of determining the number of solutions to a system of equations derived from the definition of the original problem. The exact number of solutions is computed as follows. We first develop an algorithm for the computation of a *Groebner basis* in the Boolean setting, and then, from such a basis, we actually compute the number of solutions, which results to be equal to the number of monomials that are not divisible by the *leading terms* of the polynomials in the basis (see Section 4.3 for the appropriate definitions). Informally, the intuition behind this *modus operandi* is that we first transform a system of polynomial equations into an equivalent one (the Groebner basis) with special properties, and then, by taking advantage of these properties, we more easily compute the number of solutions. This process can be viewed as a sort of generalization of triangularization techniques for linear systems, where the triangular form gives immediate information, e.g., on the rank and the determinant of the coefficient matrix, and thus on properties of the solutions.

This chapter has three main goals:

- To present an optimized code for the computation of Groebner bases, when each variable x_i is restricted to the Boolean setting by the equation $x_i^2 - x_i = 0$. This code can be used in a variety of applications. As an example, [CEI96] suggests using *Groebner proofs* as an alternative to resolution for the construction of proof systems. The efficiency of such Groebner proofs crucially relies on the possibility of taking advantage of the special structure induced by the presence of the equation $x_i^2 - x_i = 0$, for each of the variables x_i.

- To show that the above outlined approach can sometimes be a viable alternative to existing methods for finding the exact number of solutions to counting problems.

For example we show that our algorithm computes the permanent of some classes of matrices much faster than the best known algorithm, which is due to Ryser [R63]. While we do not claim that we can achieve good running times in general, we view our effort as a first step towards providing a unified and highly adaptive computational environment for counting problems. In fact, one good feature of our algorithms is that, unlike Ryser's method, they are intrinsically very sensitive to the structure of the specific input, and this makes it possible to solve efficiently several different classes of easy instances, without tailoring the computation to handle them in a specific way.

- To study properties of special permanents for whose computation it is still unclear whether or not there exist efficient algorithms. For example the best known algorithm to compute the permanent of an $n \times n$ *circulant matrix* with three ones per row takes $O(n2^{\frac{n}{2}})$ time [CCR96], roughly the square root of the time of Ryser method. The analysis of the rich structure of the Groebner bases associated with these matrices (see Section 4.6) might lead to better algorithms.

As a first case study to test the efficiency of our approach, we indeed focus our attention on the computation of the permanent which represents the hardest problems within the class #P.

From a computational viewpoint, this study is an attempt of employing *computational algebraic geometry* techniques in the Boolean setting. The use of algebraic geometry tools in the Boolean domain has a recent history. For example, Smolensky has shown that the Hilbert function is responsible for certain lower bounds [Sm93]. As already mentioned, Clegg et al. have suggested that Groebner basis techniques might be possible alternatives to resolution in the construction of proof systems [CEI96].

The chapter is organized as follows.

In Section 4.2 we describe our approach to the solution of counting problems with a special attention to the permanent computation. In Section 4.4 we give a high level description of the two main stages of our algorithms, i.e., the computation of a minimal Groebner basis and the computation of the number of solutions starting from the initial monomials of the Groebner basis. In Section 4.5 we describe some key implementation issues, focusing on the data structures used and on the peculiarities related to the Boolean setting. We also analyze the computational cost of the algorithms. In Section 4.6 we present the experimental results obtained and we compare the performance of our algorithms with that of Ryser method and of the Macaulay package [BS82]. In Section 4.3 we give some background on the permanent computation, we briefly review some basic notions from algebraic geometry, and formally show the correctness of our algorithms. Section 4.7 contains figures and tables.

4.2 The approach

For simplicity, we first describe our method in the case of the permanent computation. At the end of the Section, we show how it can be applied to other counting and decision problems.

Let $A = (a_{ij})$ be a $(0,1)$-matrix. We define a matrix $X = (x_{i,j})$ depending on variables $t_{i,j}$, in the sense that its (i,j)-th entry $x_{i,j}$ is equal to $t_{i,j}a_{i,j}$. We then consider the following system of $n^2 + 2n$ equations:

$$\begin{cases} \sum_{j=1}^{n} x_{i,j} = 1 & \text{for } i = 1, 2, \dots, n \\ \sum_{i=1}^{n} x_{i,j} = 1 & \text{for } j = 1, 2, \dots, n \\ x_{i,j}(1 - x_{i,j}) = 0 & \text{for } i, j = 1, 2, \dots, n. \end{cases} \quad (1)$$

Fact 70 *Let A be a $(0,1)$ $n \times n$ matrix and let X be the matrix constructed as described above. Then* $\mathrm{per}(A)$ *is equal to the number of solutions of the system of equations (1).*

Proof. The permanent of A is the number of its *nonzero permutations*, where a permutation σ is nonzero if $a_{k,\sigma(k)} = 1$, for $k = 1, 2, \dots, n$. It is easy to see that each nonzero permutation σ uniquely corresponds to a solution of (1). We simply let $x_{i,j} = 1$ if $j = \sigma(i)$, and $x_{i,j} = 0$ otherwise. The converse is also true since the third equation restricts the range of the solutions to 0 and 1, and the other two equations select exactly one nonzero entry in each row and column of A. □

This shows that it is #P-complete to determine the number of $\{0,1\}$ solutions to a linear system.

Fact 70 allows us to bring the computation of the permanent into the framework of algebraic geometry. Indeed it hints at computing the permanent via, e.g., Groebner basis and Hilbert function computation (see Section 4.3 for the appropriate definitions). More precisely, we proceed as follows. We first compute a Groebner basis for the ideal of polyonomials $\sum_{j=1}^{n} x_{i,j} - 1$, $\sum_{i=1}^{n} x_{i,j} - 1$, and $x_{i,j}(1 - x_{i,j})$. Then we compute the number of monomials that are not divisible by any of the leading terms of the Groebner basis, which, from Lemma 74 of Section 4.3 (also Lemma 1 in [Sm93]), is equal to the number of solutions of (1). The actual description of the algorithms will be provided in Section 4.4 and a more detailed proof of their correctness at the end of Section 4.3.

Note that this approach can be extended to many other problems. For instance, to describe $3SAT$, in addition to the equations for the restriction of the variables to the Boolean domain, one needs equations of the type

$$x + y + z - xy - xz - yz + xyz = 1\,.$$

In the case of graph isomorphism, the goal is to find permutation matrices P such that $PA = BP$, where A and B are the adjacency matrices of two graphs. Then one can write equations in terms of a matrix X of variables for which one requires that $XA = BX$, in addition to the restrictions that guarantee that the row and column sums are both equal to one, and to the restrictions of the variables to the Boolean domain.

4.3 Algebraic Geometry background

We recall some basic notions from algebraic geometry, focusing on *monomial ideals*, *Groebner bases*, and *Hilbert function*, which are central to our investigation. For more details, a good reference is the textbook [CLO92].

Let $S \subset A^n$ be a subset of the n dimensional affine space over a field k. Let $I(S) \subseteq k[x_1, x_2, \ldots, x_n]$, denote the ideal of polynomials that vanish on S. For any nonnegative integer m, let $k[x_1, x_2, \ldots, x_n]_{\leq m}$ denote the k-module of polynomials of total degree $\leq m$ in $k[x_1, x_2, \ldots, x_n]$. Similarly, let $I(S)_{\leq m}$ denote the k-module of polynomials of total degree $\leq m$ vanishing on S. As customary, we will use the compact notation x^α, $\alpha \in \mathcal{Z}^n_{\geq 0}$, to represent a monomial of the form $\prod_i {x_i}^{\alpha_i}$.

Definition 5 *A* monomial ordering *on $k[x_1, x_2, \ldots, x_n]$ is any relation $>$ on the set of monomials x^α, $\alpha \in \mathcal{Z}^n_{\geq 0}$, satisfying:*

- *$>$ is a total ordering on $\mathcal{Z}^n_{\geq 0}$.*
- *If $\alpha > \beta$ and $\gamma \in \mathcal{Z}^n_{\geq 0}$, then $\alpha + \gamma > \beta + \gamma$.*
- *$>$ is a well-ordering on $\mathcal{Z}^n_{\geq 0}$. This means that every nonempty subset of $\mathcal{Z}^n_{\geq 0}$ has a smallest element under $>$.*

Note that the usual lexicographic ordering on $\mathcal{Z}^n_{\geq 0}$ is a monomial ordering.

Definition 6 *An ideal $I \subseteq k[x_1, x_2, \ldots, x_n]$ is a* monomial ideal *if there is a subset $A \subseteq \mathcal{Z}^n_{\geq 0}$ (possibly infinite) such that I consists of all polynomials which are finite sums of the form $\sum_{\alpha \in A} h_\alpha x^\alpha$, where $h_\alpha \in k[x_1, x_2, \ldots, x_n]$.*

The value of the *Hilbert function* computed at m, $H_{I(S)}(m)$, is the dimension as a vector space of the set of polynomials of total degree at most m restricted to S. The space of polynomials restricted to S is the quotient space of the space of all polynomials over k, over the ideal $I(S)$. Indeed two polynomials are distinct in the quotient space if and only if they differ at some point of S. The quotient of the corresponding k-modules is therefore the required set (and a k-module itself). More formally, we have the following definition.

Definition 7 *The* Hilbert function *of $I(S)$ is the function on the non-negative integers m defined as the vector-space dimension of the quotient k-module*

$$H_{I(S)}(m) = \dim\, k[x_1, x_2, \ldots, x_n]_{\leq m} / I(S)_{\leq m} \ .$$

(For simplicity we often refer to the Hilbert function of the ideal of polynomials vanishing on a set S as the Hilbert function of the set S.) The computation of the Hilbert function of a polynomial ideal I can be reduced to the computation of the Hilbert function of a monomial ideal. First of all, given a monomial ordering λ it becomes possible to define, for any $p \in k[x_1, x_2, \ldots, x_n]$, its *leading term*, i.e. its largest monomial with respect to λ. Then, for any ideal I, we can define its *ideal of leading terms* as follows.

Definition 8 *Let $I \subseteq k[x_1, x_2, \ldots, x_n]$ be an ideal, and let $I \neq 0$.*

- *$LT(I)$ denotes the set of leading terms of elements of I.*
- *$\langle LT(I) \rangle$ denotes the ideal generated by the elements of $LT(I)$.*

$\langle LT(I) \rangle$ is a monomial ideal.

Theorem 71 (see [BW93]) *For any ideal $I \subseteq k[x_1, x_2, \ldots, x_n]$, and for any monomial ordering, the monomial ideal $\langle LT(I)\rangle$ has the same Hilbert function as I.*

If we are given a finite generating set for I, say $I = \langle f_1, f_2, \ldots, f_s\rangle$, $f_i \in k[x_1, x_2, \ldots, x_n]$, then $\langle LT(f_1), LT(f_2), \ldots, LT(f_s)\rangle$ and $\langle LT(I)\rangle$ may be *different* ideals. Indeed

$$\langle LT(f_1), LT(f_2), \ldots, LT(f_s)\rangle \subseteq \langle LT(I)\rangle,$$

but, in general, equality needs not to occur.

A generating set of I with the special property that its leading terms generate the whole ideal $\langle LT(I)\rangle$ is called *Groebner basis.* More precisely, for any fixed monomial ordering, we have the following definition.

Definition 9 *A finite generating set $G = \{g_1, g_2, \ldots, g_t\}$ for an ideal I is said to be a* Groebner basis *if*

$$\langle LT(g_1), LT(g_2), \ldots, LT(g_t)\rangle = \langle LT(I)\rangle \ .$$

Equivalently, a set $\{g_1, \ldots, g_t\} \subseteq I$ is a Groebner basis of I if and only if the leading term of any element of I is divisible by one of the $LT(g_i)$'s. Note that, although a given ideal may have many Groebner bases, it is possible to single out a special Groebner basis, which turns out to be unique.

Definition 10 *A* reduced Groebner basis *for an ideal I is a Groebner basis G for I such that:*

- *Any $p \in G$ is monic, i.e., the coefficient of its leading term is equal to 1.*
- *For all $p \in G$, no monomial of p is divisible by any of the monomials in $\langle LT(G\backslash\{p\})\rangle$.*

Proposition 72 (see [CLO92]) *Let $I \neq \{0\}$ be a polynomial ideal. Then, for a given monomial ordering, I has a unique reduced Groebner basis.*

Theorem 71 shows that the leading terms of the polynomials in I play an important role in the computation of the Hilbert function.

Proposition 73 (see [CLO92]) *Let $I \subseteq k[x_1, x_2, \ldots, x_n]$ be a monomial ideal, generated by a set Q of monomials. Then, the value of the Hilbert function of I, computed at m, is equal to the number of the monomials of degree at most m that are not divisible by any of the monomials in Q.*

Lemma 74 (see [Sm93]) *Let $S \subseteq A^n$ be a finite set. A set $G = \{g_1, g_2, \ldots, g_t\}$ of polynomials, $G \subseteq I(S)$, is a Groebner basis for $I(S)$ if and only if the number of monomials that are not divisible by any of the leading terms of polynomials in G equals $|S|$.*

Theorem 75 ([BW93]) *Let K be a finite field. If $dim(I) = 0$ and L is an algebraically closed extension field of K, then the number of zeros of I in L^n equals $dim_K(K[\underline{X}]/\sqrt{I})$.*

Lemma 76 ([BW93]) *Let I be a zero-dimensional ideal of $K[\underline{X}]$, and assume that for $1 \leq i \leq n$, I contains a polynomial $f_i \in K[x_i]$ with $\gcd(f_i, f_i') = 1$. Then I is a radical ideal.*

Correctness of our approach. We aim at computing the number of zeros of the ideal $I \subseteq K[\underline{X}]$ generated by the system of equations (1). Since the number of solutions is finite, then I is zero-dimensional. Furthermore I satisfies theorem (76) since $\gcd(x^2-x, 2x-1) = 1$ and so it is radical, i.e., $I = \sqrt{I}$. Finally, our work field is $\mathbf{Z}_p$, with p prime, i.e., $K = \mathbf{Z}_p$. This field is obviously finite.

Let L be any algebraically closed extension field of $K = \mathbf{Z}_p$ and G be a Groebner basis for I. Theorem 75 and $I = \sqrt{I}$ imply that the Hilbert function associated with I, HF_I, eventually stabilizes onto a constant value which is precisely $\dim_K(K[\underline{X}]/I)$. Let s_0 be an integer satisfying $HF_I(s) = \dim_K(K[\underline{X}]/I)$, for all $s \geq s_0$. The the following equalities hold:

$$\begin{aligned}
|V_{L^n}(I)| &= \#\{\sigma \mid \sigma \in L^n \text{ is a zero of } I\} \\
&= \dim_K(K[\underline{X}]/\sqrt{I}) \\
&= \dim_K(K[\underline{X}]/I) \\
&= HF_I(s_0) \\
&= HF_{\langle LT(I)\rangle}(s_0) \text{ (see Lemma 71)} \\
&= \#\{x^\alpha \in K[\underline{X}] \mid x^\alpha \notin \langle LT(I)\rangle\}
\end{aligned}$$

By the properties of monomial ideals we thus have that a monomial m is not in $\langle LT(I)\rangle$ if and only if it is not divisible by any of the $LT(g_i)$'s for $g_i \in G$. We now obtain

$$|V_{L^n}(I)| = \#\{x^\alpha \in K[\underline{X}] \mid x^\alpha \text{ is not divisible by any of the } LT(g_i)\text{'s for } g_i \in G\}.$$

Furthermore, since the zeroes of the ideal I belong to K^n, it turns out that $|V_{L^n}(I)| = |V_{K^n}(I)|$ from which

$$|V_{K^n}(I)| = \#\{x^\alpha \in K[\underline{X}] \mid x^\alpha \text{ is not divisible by any of the } LT(g_i)\text{'s for } g_i \in G\},$$

which proves the correctness of our algorithms.

4.4 Algorithms

Our algorithms are divided into two main stages. We first compute, from the set of equations that describe the original problem, an equivalent set of equations which corresponds to a *minimal* Groebner basis, i.e. a basis with the minimum number of polynomials. Then we use the leading terms of the basis (see Section 4.3) in order to compute the number of solutions of the original set of equations.

A Groebner basis algorithm for the Boolean setting. We have implemented four algorithms for the computation of a Groebner basis, all based on Buchberger method and its subsequent refinements [CLO92, GM88]. For their description, we need the following definitions. Given two polynomials, f and g, the *S-polynomial* of f and g is the polynomial $S(f,g) = \frac{h}{LT(f)} \cdot f - \frac{h}{LT(g)} \cdot g$, where h is the least common multiple of $LT(f)$ and $LT(g)$, i.e., the leading monomials of f and g. Let $F = (f_1, f_2, \ldots, f_t)$ be a t-tuple of polynomials in $k[x_1, \ldots, x_n]$, ordered according to a fixed monomial ordering. Then every $g \in k[x_1, \ldots, x_n]$

can be expressed as $g = r + \sum_{i=1}^{t} a_i f_i$, where $a_i, r \in k[x_1, \ldots, x_n]$, and either $r = 0$ or none of the monomials of r is divisible by any of the monomials $LT(f_1), \ldots, LT(f_t)$. We call r a *remainder* of g on division by F, and we denote it with $r = \overline{g}^F$.

Buchberger algorithm incrementally computes a Groebner basis G of a set of polynomials $F = \{f_1, f_2, \ldots, f_k\}$. Initially we set $G = F$. Subsequently, we compute, for each pair of polynomials $\{f_p, f_q\} \in G$, $f_p \neq f_q$, the remainder $R = \overline{S(f_p, f_q)}^G$. If $R \neq 0$, then R is added to the basis G. This procedure is iterated until $\overline{S(f_p, f_q)}^G = 0$, for each pair of polynomials in G. The resulting set G is a (non minimal) Groebner basis for F. This basic version of the algorithm is very inefficient, because it takes into account a very large number of pairs of polynomials. Building upon the suggestions reported in [CLO92], we have thus implemented the following version of the algorithm, which is the first of our four algorithms (Algorithm 1).

`Algorithm 1`
`Input` : A set of polynomials $F = \{f_1, f_2, \ldots, f_k\}$.
`Output` : A minimal Groebner basis G for F.

---

1. $B \leftarrow \{(i,j) \mid 1 \leq i < j \leq t\}$
2. $G \leftarrow F$
3. `while` $B \neq \emptyset$ `do`
4.   `select` $(i,j) \in B$
5. $B \leftarrow B - \{(i,j)\}$
6.   `if Good`(i,j) `then`
7. $R = \overline{S(f_i, f_j)}^G$
8.   `if` $R \neq 0$ `then`
9. $t \leftarrow t + 1$
10. $f_t \leftarrow R$;
11. $G \leftarrow G \cup \{f_t\}$;
12. $B \leftarrow B \cup \{(i,t) \mid 1 \leq i < t\}$
13.   `endif`
14.   `endif`
15. `endwhile`
16. `Minimize`(G)

The criterium `Good`(i,j) adopted at line 6 of Algorithm 1 allows us to avoid the computation of $R = \overline{S(f_i, f_j)}^G$ for a non negligible number of pairs (i,j), and is defined as follows:

$$\texttt{Good}(i,j) = \begin{cases} \texttt{False} & \text{if } \mathrm{lcm}(LT(f_i), LT(f_j)) = LT(f_i)\,LT(f_j) \\ \texttt{False} & \text{if } \exists k \notin \{i,j\} : (i,k) \notin B\ (j,k) \notin B \text{ and} \\ & \quad LT(f_k) \text{ does not divide } \mathrm{lcm}(LT(f_i), LT(f_j)) \\ \texttt{True} & \text{elsewhere} \end{cases}$$

After a Groebner basis G has been computed, we *minimize* it (line 16), simply by deleting all the polynomials whose leading term is a multiple of the leading term of another polynomial in the basis. We thus obtain a minimal basis, which in general is not unique. Note that the minimality of the basis is sufficient for our purposes, since in the subsequent computation of the number of solutions we just need the leading terms of G. Algorithm 1 selects a pair from B (line 4) according to a simple LIFO strategy.

We have developed another algorithm (Algorithm 2), which implements Buchberger suggestion to select from B the pair (i, j) for which the degree of $\text{lcm}(LT(f_i), LT(f_j))$ is minimum.

Algorithms 1 and 2 often introduce (unnecessarily) large sets B. This is due to the fact that both algorithms perform insertion in the set B without any check. Gebauer and Möller [GM88] describe a different, more complicated, criterion that can be applied at the time of insertion, and which allows us to avoid all the checks after selection. We have implementated this criterion in Algorithm 3. We have also implemented another version of Buchberger method (Algorithm 4). This consists of the following minor modification of Algorithm 3: via preliminary reductions and sorting of the input polynomials F, we maintain in a minimal form the incrementally built basis. For Algorithm 4 the final step of minimization is thus not necessary any more.

Computation of the number of solutions. Let $M = \{m_1, \ldots, m_k\}$ be the set of leading monomials of a minimal Groebner basis G, defined in terms of the variables $V = \{v_1, \ldots, v_n\}$, and let M' be the set of the multilinear monomials of M. According to Lemma 74, the value of the permanent is equal to the cardinality of the set $\overline{M}$ of the monomials that are not divisible by any $m_i \in M$. We claim that this number is equal to the number of $\{0, 1\}$ solutions of the system obtained equating to zero all the monomials in M'. In fact, by construction, M contains only multilinear monomials and monomials of the form v_i^2. Since $\overline{M}$ is finite, for every $v_i \in V$, either v_i or v_i^2 belong to M, and thus all the monomials in $\overline{M}$ are multilinear. Hence $\overline{M}$ is also equal to the set of the multilinear monomials that are not divisible by any of the monomials of M'. Given an assigment $S = \{v_i \leftarrow \alpha_i\}_{i=1,\ldots,n}$, with $\alpha_i \in \{0, 1\}$, we claim that S is a solution of the system $M' = 0$ if and only if the monomial $m_S = \prod_{i=1}^n v_i^{\alpha_i}$ belongs to $\overline{M}$. In fact, if S is a solution, then any given $m \in M'$ contains at least a variable v which takes the value zero in S, and thus, by definition, does not appear in m_S, which thus cannot be divisible by m. Since this holds for any $m \in M'$, we obtain that $m_S \in \overline{M}$. The reverse implication follows in a similar way.

To evaluate the number of solutions of M' we have thus adopted the recursive algorithm described below. We denote by $|m|$ the degree of a monomial and by $M[v = \alpha]$ the set of monomials obtained from M by setting $v = \alpha$, for $\alpha \in \{0, 1\}$.

NumSol(M,V)

Input : A set of multilinear monomials $M = \{m_1, m_2, \ldots, m_k\}$
The set of variables $V = \{v_1, v_2, \ldots, v_n\}$.
Output : The number of solutions in $\mathbf{Z}_2^n$.

1. if $k = 0$ return 2^n
2. elseif $k = 1$ return $2^{n-|m_1|}(2^{|m_1|} - 1)$
3. elseif $\exists j : |m_j| = 1$ return NumSol($M[v_j = 0], V - \{v_j\}$)
4. else
5. select $v_j \in V$
6. return NumSol($M[v_j = 0], V - \{v_j\}$) + NumSol($M[v_j = 1], V - \{v_j\}$)
7. endif

In the current version of the algorithm, at line 5 we select the variable which appears more frequently in the monomials of M with smallest degree. Other (more refined) strategies could be used for very large size problems.

4.5 Implementation Issues

In this Section we describe the most important features of our implementations, expecially focusing on the specialization of computations to the Boolean domain. For reasons of space, our implementation of Ryser algorithm is in Section 4.3. The algorithms have been implemented in the C language, the code compiled with the GNU GCC compiler, and the experiments carried out on a SUN Superspark 20 workstation.

Computation of Groebner Bases. The polynomials that occur in the execution of algorithms 1, 2, 3, and 4 are either of the form $x_i^2 - x_i$, or multilinear. In fact, if a term of the form x_i^k, $k > 1$, appears in a polynomial (different from $x_i^2 - x_i$), then it can be immediately simplified to x_i, since the constraint $x_i^2 - x_i = 0$ implies that $x_i^k = x_i$ for every $k > 1$. This observation led us to choose (in the implementation of all the algorithms) the following representation for monomials and polynomials. A polynomial is represented by a record containing the number of its monomials (the 'length' of the polynomial) and a pointer to an array of monomials. The polynomials of the form $x_i^2 - x_i$ are encoded as polynomials of length -1, and with only a monomial representing x_i (see Figure 4.1). If at most n variables are used, a monomial is identified with a record containing the monomial coefficient and a vector of n bits that represents the exponents of the variables contained in the monomial. Given the nature of the problems at hand, we can restrict the range of the coefficients to $\mathcal{Z}_p$, where p is a prime number greater than of equal to the maximum number of variables that might appear in one equation (for the permanent computation, this is the maximum number of ones appearing in a row or column of the input matrix). We take advantage of the fact that monomials are square-free in order to use only one bit per variable (see Figure 4.1). Although for a large number of variables this representation might waste some space, it has the desirable feature to allow the implementation of all the needed monomial computations as fast bit-wise logical operations. For example, monomial multiplication (as well as the computation of lcm), can be realized with a bit-wise OR, while division can be realized with a bit-wise XOR (exclusive or), once divisibility has been tested. To speed up the execution of these operations, we have arranged bits in larger sets that can be treated as single quantities by the compiler on the target machine. In our implementation we perform logical operations between `unsigned long integer` quantities, that are represented with 32 bits. Thus the multiplication of two monomials needs $\lceil n/32 \rceil$ OR operations on 32-bit integers, independently of the actual number of variables in the monomials, plus a fixed amount of work for the multiplication of the coefficients. Note that, according to this representation, we have, e.g., $xy \cdot yz = xyz$. This is correct since all the variables are restricted to the Boolean domain. In addition, with this representation we achieve a very fast implementation of comparison between monomials. (These comparisons must be performed in order to sort monomials according to, e.g., the lexicographic monomial ordering.) Indeed at most $\lceil n/32 \rceil$ comparisons between 32-bit unsigned integers are sufficient to decide the ordering relation between two monomials.

The choice of adopting a special representation for the quadratic polynomials of the form $x_i^2 - x_i$, does not cause any problem, because these polynomials just belong to the initial sets of polynomials, and can not be generated during the computation of G. Accordingly, our monomial multiplication algorithm outputs only multilinear polynomials. Quadratic polynomials can mix with multilinear ones only in the computation of the S-polynomial $S(f_i, f_j)$, when either f_i or f_j is of the form $x^2 - x$. However the S-polynomial has the property of annihilating the leading terms of f_i and f_j, and thus it can be computed without actually multiplying the quadratic term. The case when both f_i and f_j are quadratic, and thus do not share any variable, is avoided *a priori* by appropriate criteria used by the algorithms. In the subsequent division of $S(f_i, f_j)$ by G, the way in which we implement the operations between monomials always leads to square-free partial results, and thus the quadratic polynomials contained in G have not to be taken into account as possible divisors. The 1-bit representation is also compatible with the criterion (`Good()`) adopted by Algorithms 1 and 2. The criteria used in Algorithms 3 and 4 are more complicated and need a temporary representation of monomials with 2 bits per variable. The set of pairs B is implemented in different ways in the four algorithms. In Algorithm 1, B behaves like a *stack* and it consists of a simple array of pairs of numbers, each related to an element of G. The selection of a pair (line 4) is a Pop operation, while the insertion of new pairs (line 12) consists of several Push operations. The sequence of Pop and Push operations maintains the stack sorted with respect to an ordering of pairs, and this is exploited in the implementation of the criterion `Good()`, where we test for the existence of a certain pair in B in logarithmic time using a binary search subroutine.

In Algorithm 2, B is implemented at the same time as a priority Heap and as a Red-Black Tree. The Heap allows us to select the pair (i, j) that minimizes the degree of $\mathrm{lcm}(LT(f_i), LT(f_j))$, while the Tree allows us to perform a fast existence test. For both structures, insertion and deletion (as well as search, for the Tree) can be done in logarithmic time.

Algorithms 3 and 4, which operate according to different criteria, do not need to test if a certain pair (i, j) belongs to B. For this reason, B is simply implemented as a stack. It is worth reporting that, while in our typical runs of Algorithms 1 and 2 the size of B can exceed, say, 30,000, in Algorithms 3 and 4 it is almost always less than 100.

Computation of the number of solutions. The implementation of algorithm `NumSol(`M`,`V`)` does not need to represent polynomials, and uses a different, simpler than above, representation of monomials, which are always multilinear, and whose coefficients are always equal to 1, and thus can be disregarded. In the following we assume that there are at most 255 variables, numbered as $v_1, v_2, \ldots$. A monomial m can thus be represented by the string of the subscripts of its variables. This string, actually an array of bytes, is terminated with a null byte, according to the conventions of the C language. A set (M,V) is implemented as a record containing a string which represents the set of variables V, an integer containing the number of monomials in M, and an array of strings representing M. The set $M[v_i = 0]$ is obtained from M deleting all the monomials (i.e., strings) that contain v_i, while the set $M[v_i = 1]$ is obtained from M deleting the characters corresponding to v_i from its strings. Since `NumSol(`M`,`V`)` can reach very deep levels of recursion, it has been implemented in a non-recursive fashion, using a stack.

Cost Analysis. All the monomial operations have a cost proportional to the number of variables n_v defining the problem. This follows from the fact that each monomial is represented by a 16-bit integer for the coefficient and $\lceil \frac{n_v}{32} \rceil$ 32-bit integers for the exponents of the variables (see above). The cost of the simple polynomial operations that are employed by the four algorithms, namely addition, subtraction, and multiplication by a monomial, have a cost roughly proportional to $n_v l_p \log l_p$, where l_p is the maximum number of monomials in the polynomials that occur during the computation. The logarithmic factor $\log l_p$ arises from the execution of a sorting stage (needed for simplification purposes) after each polynomial computation. The most expensive operation performed by the algorithms is the computation of the remainder R of the division of an S-polynomial $S(f,g)$ by the polynomials already in the basis G. Let n_G and n_B be the cardinalities of the sets G and B, respectively. At each step of the division, the algorithm scans the n_G elements of G looking for a polynomial whose leading term divides the leading term $LT(R')$ of the current remainder R' (which initially coincides with $S(f,g)$). If the search is unsuccessful, then $LT(R')$ is deleted from R', and appended to R. If there exists $h \in G$ such that $LT(h)$ divides $LT(R')$, then R' is updated according to the rule $R' \leftarrow R' - \frac{LT(R')}{LT(h)} h$. The process terminates when $R' = 0$. The overall time needed to carry out one step of the division algorithm is thus proportional to $n_v n_G + n_v l_p \log l_p$.

Our four algorithms differ for the criteria adopted to execute the following two steps: (i) update the set B (to be performed each time a new element of the basis is found); (ii) select from B a pair to form a new S-polynomial.

For Algorithms 1 and 2, the selection from B and the check of the `Good()` criterion, have a cost proportional to $n_v n_B \log n_B$. The cost of insertion is proportional to n_G for Algorithm 1, and to $n_G n_v \log n_B$ for Algorithm 2, since in the latter case the n_G new pairs are inserted into two dynamic structures (a heap and a tree). For Algorithm 3 and 4, the selection from B has unitary cost since all the checks are performed when B is updated. The criteria used to decide the insertion of new pairs have a cost roughly proportional to $n_v(n_G^2 + n_B)$. The Minimization step at the end of Algorithms 1, 2 and 3, and the Reduction step at the beginning of Algorithm 4, have generally a negligible cost with respect to the rest of the computation. The four algorithms have a different behaviour for what concerns the growth of the quantities n_G, n_B, and l_p, during their execution. Experimentally we observed that Algorithm 1 produces larger bases G (i.e., there are many extra elements, to be deleted during the Minimization step) and executes more divisions than the other algorithms. However it has the best rate of growth for l_p. Algorithm 2 often requires a number of divisions which is 10 times less than the other algorithms, but it also builds sets B 10 times larger than those of Algorithm 1. Furthermore it generates very long polynomials. Algorithms 3 and 4 maintain a much smaller set B, but the cost to manage it is greater than for Algorithms 1 and 2. Algorithm 4 generates small sets G, since it maintains them in minimal form, but, as Algorithm 2, produces a fast growth of l_p. As we will see in Section 4.6, it turns out that Algorithms 2 and 4 exploit better than the other two algorithms certain structural properties of the matrices. In particular, in the case of circulant matrices of the form $I+P^i+P^j$ (see Section 1.5.1 for the appropriate definitions), Algorithms 2 and 4 are significantly faster either when the matrix is symmetric or when i, j, and the size of the matrix have a large common factor.

4.6 Experimental Results

Given the fact that the permanent of a $(0,1)$ matrix with at most three ones per row and columns keeps all the difficulty of the general problem [DLMV88], we have decided to tailor our investigation to this kind of matrices. We report the outcomes of experiments on the following classes of matrices, all with at most three ones per row: circulant matrices of the type $I + P + P^j$, random symmetric matrices with ones on the main diagonal, random matrices with ones on the main diagonal, random Hessenberg matrices, and block circulant matrices. The selection of these case studies has been made in order to test our approach against a wide range of *difficulties*. In fact the above classes include examples of convertible matrices, of very structured matrices that nevertheless are not known to have "easy permanents", and of matrices whose permanents are as hard to compute as in the general case.

Circulant Matrices. Despite the fact that the permanent of circulant matrices have been widely investigated, and several of its algebraic properties discovered, there are no clear indications of whether or not it can be computed efficiently, even for matrices with only three ones per row. We have gathered evidence that the methodology employed in this chapter can be a valuable tool for investigations on these classes of matrices. In fact, the results of Figure 4.2 closely reflect known algebraic differences between matrices of the type $I + P + P^j$, as j varies. Roughly speaking, the permanent turns out to be particularly easy, e.g., either for j very small or for j close to $\frac{n}{2}$ (see [Mi87, CCR96]). In particular, the permanent of the matrix $I + P + P^2$ has a very simple expression, and two of our algorithms, as well as the package Macaulay [BS82], clearly detect it, by running in (observed) linear time. In particular, one of our algorithms is faster than Ryser algorithm, for any $n \geq 20$ (see Table 4.1).

Other test matrices. Also for the other test cases, the time performance of the algorithms (expecially algorithms 2 and 4) is widely influenced by the structure of the problems. In particular, it is nice to see that algorithms 2 and 4 exploit the structure of symmetric matrices with ones on the main diagonal, which are known to be easy (see Table 4.2).

Tables and Figures in Section 4.7 show also that our algorithms provide a substantial improvement over the package Macaulay.

4.7 Figures and Tables

In this Section we provide tables and figures related to implementation issues and experiments.

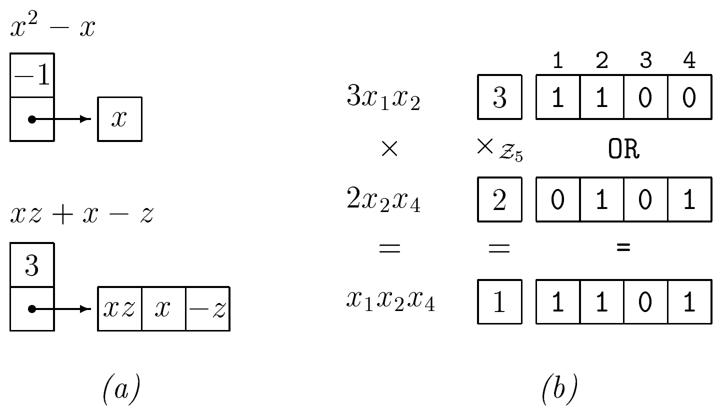


Figure 4.1: *(a)* An example of representation for polynomials. *(b)* An example of representation and multiplication of monomials.

j	Macaulay	H. f.	Alg. 1	Alg. 3	NumSol
2	7.0	2.06	0.3	0.5	0.02
3	129.4	2.36	29.9	30.7	0.04
4	400.6	3.24	229.0	188.3	0.09
5	205.3	3.25	237.8	171.1	0.09
6	575.7	3.49	417.5	276.3	0.10
7	240.1	3.40	279.0	260.9	0.10
8	3.5	2.17	0.3	0.5	0.03
9	152.0	3.40	341.9	161.9	0.09
10	536.1	3.58	550.7	268.4	0.11
11	174.2	3.26	361.6	172.5	0.08
12	292.9	3.34	360.7	238.1	0.09
13	155.1	2.36	124.9	148.9	0.06
14	12.6	2.43	0.4	0.7	0.02

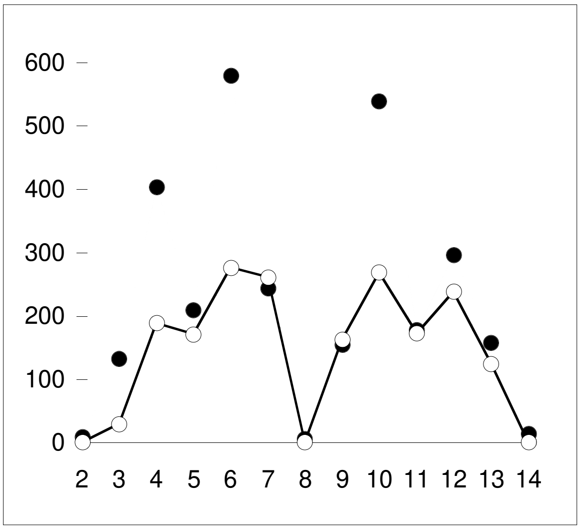


Figure 4.2: Computation of the permanent of $I+P+P^j$ for $n=15$, as j varies. The Table on the left reports the time performance (in seconds) for Macaulay package (Groebner basis and Hilbert function) and for our Algorithms 1 and 3. The last column gives the time performance of the computation of the number of solutions, after Algorithm 1 and/or 3 provided a Groebner basis. The graph on the right shows the time performance as a function of j, and visualizes the improvement achieved by our algorithms (white circle) over Macaulay (black circle).

n	Ryser	Macaulay	H. f.	Alg. 1	Alg. 2	Alg. 3	Alg. 4	NumSol
10	0.01	1.6	0.38	0.1	0.2	0.1	0.1	0.01
11	0.01	2.3	0.49	0.1	0.3	0.2	0.1	0.01
12	0.02	3.1	0.68	0.2	1.1	0.3	0.2	0.01
13	0.02	4.2	0.94	0.2	1.2	0.4	0.2	0.01
14	0.05	5.3	1.43	0.3	1.0	0.4	0.3	0.02
15	0.09	6.9	2.05	0.3	1.4	0.5	0.4	0.02
16	0.18	8.7	3.36	0.4	1.6	0.7	0.5	0.03
17	0.35	11.2	5.14	0.5	5.0	0.8	0.6	0.04
18	0.72	13.7	*	0.6	6.1	1.0	0.7	0.04
19	1.46	17.1	*	0.7	6.7	1.2	1.0	0.05
20	2.96	20.4	*	0.8	4.6	1.3	1.1	0.07
21	6.00	25.7	*	1.0	6.7	1.6	1.4	0.09
22	12.23	29.8	*	1.1	6.5	2.2	1.7	0.13
23	24.98	35.8	*	1.3	36.1	2.6	2.3	0.16
24	50.98	43.1	*	1.5	44.6	3.0	2.5	0.21
25	126.12	50.5	*	1.7	41.5	4.2	3.3	0.28
26	255.66	57.8	*	1.9	43.2	3.9	3.6	0.35
27	8 min.	68.4	*	2.2	48.3	4.6	4.5	0.46
28	17 min.	78.9	*	2.4	45.8	6.0	5.0	0.62
29	35 min.	116.6	*	2.7	37.3	5.9	6.3	0.82
30	71 min.	104.2	*	3.1	42.0	6.6	6.7	1.08
31	145 min.	121.7	*	3.4	52.8	8.6	8.5	1.42
32	5 hours	136.9	*	3.8	48.5	8.5	9.1	1.89
40	58 days	353.2	*	8.8	314.1	21.8	*	17.63

Table 4.1: Computation of the permanent of $I + P + P^2$ for different values of n. The entries contain running times (in seconds). We filled an entry with the symbol * when the computation time exceeded a given bound. The running time of Ryser algorithm for large instances have been estimated.

	Non Symmetric + I			Symmetric + I			Hessenberg			Block Circulant		
n	8	12	16	10	20	40	30	40	50	4x5=20	6x6=36	5x9=45
Ryser	0.005	0.02	0.16	0.01	2.75	53 d.	1088	12 d.	36 y.	2.51	82 d.	4 y.
Alg. 1	0.18	43.57	-	3.79	-	-	0.26	0.54	0.91	59.69	-	-
Alg. 2	0.15	13.30	-	0.30	1.96	5.36	0.84	1.86	3.08	0.55	2.86	4.91
Alg. 3	0.16	16.34	-	0.91	-	-	0.58	1.14	1.35	115.18	-	-
Alg. 4	0.06	5.77	250	0.07	0.37	3.25	0.11	0.24	0.64	0.17	1.06	1.46
NumSol	0.01	0.02	0.10	0.01	0.07	10.12	0.01	0.01	0.01	0.06	0.47	5.43

Table 4.2: Computation of the permanent for different kinds of test matrices, for different values of n. The entries contain running times (in seconds). The entries with the symbol "-" correspond to instances for which the given algorithm exceeded a certain time bound. The running time of Ryser algorithm for large instances have been estimated. The letter "d" stands for "days", and the letter "y" for "years".

Chapter 5

Conclusions

In this thesis we have provided a contribution to the investigation on the permanent of some sparse and structured $(0,1)$ matrices. We found that the permanents of the matrices analyzed here present several strong properties which sometimes make their computation tractable.

In particular, by building upon a few basic and simple facts, i.e., that

- the permanent of $(0,1)$ symmetric Toeplitz matrices with three diagonals grows as a generalized Fibonacci sequence with the matrix size,
- the permanent of $(0,1)$-circulant matrices with 2 nonzeros per row grows exponentially with the gcd between the matrix size and the index that identifies its nonzero off-diagonal,
- the permanent of certain circulant matrices and of symmetric Toeplitz matrices is a power of the permanent of a smaller matrix of the same type,
- the permanent of certain non convertible circulant matrices can be expressed as a sum of a few determinants of Toeplitz matrices,

we have been able to devise more general formulas and efficient algorithms for the computation of permanents of $(0,1)$ nonsymmetric Toeplitz matrices with three diagonals and $(0,1)$-circulant matrices with three nonzeros per row (see [CCR96] and [CCR96b]).

We believe that the partial results obtained in this thesis deserve further investigation. The goal is to achieve a deeper understanding of the structural and computational properties of the permanents of certain Toeplitz matrices.

Bibliography

[ACGS88] W. Alexi, B. Chor, O. Goldreich, C.P. Schnorr. RSA Rabin functions: Certain parts are as hard as the whole. *SIAM J. on Computing* 17:194-209 (1988).

[Al38] A.D. Alexandrov. *Mat. Sbornik (n.s.)* 3:227-251 (1938).

[An80] D. Angluin. On counting problems and the polynomial time hyerarchy. *TCS* 12:161-173 (1980).

[BG77] R.A. Brualdi and P.M. Gibson. Convex polyhedra of doubly stochastic matrices I. Applications of the permanent function. *J. Combin. Theory, Ser. A* 22:194-230 (1977).

[BGM88] R.A. Brualdi, J.L. Goldwasser, and T.S. Michael. Maximum permanents of matrices of zeros and ones. *J. Combin. Theory, Ser. A* 47:207-245 (1988).

[Bi812] J.P.M. Binet. Mémoire sur un système de formules analytiques, et leur application à des considérations géometriques. *J. École Polytechnique 9* (1812); *Cah.* 16,280-302.

[BMQ68] L. Bassett, J. Maybee, and J. Quirk. Qualitative Economics and the scope of the correspondence principle. *Econometrica* 36:544-563 (1968).

[BN65] R. Brualdi and M. Newman. Inequalities for permanents and permanental minors. *Proc. Cambridge Philos. Soc.* 61:741-746 (1965).

[BN70] R. Brualdi and M. Newman. An enumeration problem for a congruence equation. *Journal of Research, (U.S.) National Bureau of Standards* 74B:37-40 (1970).

[Bo855] C.W. Borchardt. Bestimmung der symmetrischen Verbindungen vermittelst ihrer erzeugenden Funktion. *Monatsb. Akad. Wiss. Berlin* 165-171 (1855); or *Crelle's J.* 53 (1855); or *Gesammelte Werke*, 97-105.

[Bo67] P. Botta. Linear transformations that preserve the permanent. *Proc. Amer. Math. Soc.* 18:566-569 (1967).

[Bo68] P. Botta. On the conversion of the determinant into the permanent. *Canad. Math. Bull.* 11:31-34 (1968).

[Br73] L.M. Brégman. Certain properties of nonnegative matrices and their permanents. *Dokl. Akad. Nauk SSSR* 211:27-30 (1973).

[Br86] A.Z. Broder. How hard is it to marry at random ? (On the approximation of the permanent). *Proc. of the 18th ACM Symp. on Theory of Computing* 50-58 (1986). Erratum in *Proc. of the 20th ACM Symp. on Theory of Computing* 551 (1988).

[BR91] R.A. Brualdi, and H.J. Ryser. Combinatorial matrix theory. *Cambridge University Press* (1991).

[BS82] D. Bayer and M. Stillman. Macaulay: A system for computation in algebraic geometry and commutative algebra (1982-1994). Available via ftp at math.harvard.edu.

[BS87] L. Babai and A. Seress. Private Communication, March 1987.

[BS91] R.A. Brualdi and B.L. Shader. On converting the permanent into the determinant and sign-singular matrices. *Applied Geometry and Discrete Mathematics*, (P. Gritzman and B. Sturmfels, eds) Amer. Math. Soc. Providence, R.I.

[BS95] R.A. Brualdi and B.L. Shader. Matrices of sign-solvable linear systems. *Cambridge University Press* (1995).

[BW79] J.L. Brenner and E.T.H. Wang. Permanental pairs of doubly stochastic matrices II. *Linear Algebra and its Applications* 28:39-41 (1979).

[BW93] T. Becker and V. Weispfenning. Groebner Bases. *Springer-Verlag* (1993).

[Ca812] A.L. Cauchy. Mémoire sur les fonctions qui ne peuvent obtenir que deux valeurs égales et de signes contraires par suite des transpositions oprèrèes entre les variables qu'elles renferment. *J. École Polytechnique 10* (1812); *Cah.* 17,29–112, Ouvres (2)i.

[Ca859] A. Cayley. Note sur les normales d'une conique. *Crelle's J.* 56 (1859).

[Ca59] E.R. Caianiello. Regularization and renormalization, I. *Nuovo Cimento (10)* 13:637-661 (1959).

[Ca59b] E.R. Caianiello. Theory of coupled fields. *Nuovo Cimento Suppl.* 14:177-191 (1959).

[CBS55] E. Cohen, J. de Boer and Z. Salsburg. A cell-cluster theory for the liquid state II. *Physica* XXI:137-147 (1955).

[CCR96] B. Codenotti, V. Crespi, and G. Resta. On the Permanent of Certain $(0,1)$ Toeplitz Matrices. *To appear in Linear Algebra and its Applications* (1996).

[CCR96b] B. Codenotti, V. Crespi, and G. Resta. Groebner bases in the boolean setting with applications to hard counting. *Submitted for publication* (1996).

[CDS79] D.M. Cvetkovic, M. Doob, and H. Sachs. Spectra of Graphs. *Academic Press* (1979).

[CEI96] M. Clegg, J. Edmonds, and R. Impagliazzo. Using the Groebner Basis Algorithm to Find Proofs of Unsatisfiabilit. *Proc. 28th ACM Symp. on the Theory of Comput.* (1996).

[CL60] L. Carlitz and J. Levine. An identity of Cayley. *Amer. Math. Monthly* 67:571–573 (1960).

[CLO92] D. Cox, J. Little, D. O'Shea. Ideals, Varieties, and Algorithms. *Springer-Verlag, New York* (1992).

[Co87] C.J. Colbourn. The combinatorics of network reliability. *Oxford University Press, New York* (1987).

[CW77] L. Cummings, J. Wallis. An Algorithm for the Permanent of Circulant Matrices. *Canad. Math. Bull.* 20:67-70 (1977).

[DK92] E. Dahlaus, M. Karpinski. Perfect Matchings for Regular Graphs is AC^0-Hard for the General Matching Problem. *Journal of Computer and System Sciences* 44:94-102 (1992).

[DL92] P. Dagum and M. Luby. Approximating the permanent of graphs with large factors. *Theoretical Computer Science, Part A* 102:283-305 (1992).

[DLMV88] P. Dagum, M. Luby, M. Mihail, and U. Vazirani. Polytopes, Permanents, and Graphs with Large Factors. *Proc. 27th IEEE Symp. on Found. of Comput. Sc.* (1988).

[Do67] D.Z. Doković. Some permanental inequalities. *Publ. Inst. Math. (Beograd) (N.S.)(21)* 7:191-195 (1967).

[Eg80] G.P. Egoryĉev. A solution of the van der Waerden's permanent problem. *Kirebski institute of Physics, Academy of Sciences SSSR, Preprint IFSO-13M, Kransnoyarsk (Russian)* (1980).

[FJ95] A. Frieze and M. Jerrum. An analysis of a Monte Carlo algorithm for estimating the permanent. *Combinatorica* 15 (1):67-83 (1995).

[FL92] U. Feige, and C. Lund. On the Hardness of Computing the Permanent of Random Matrices. *Proc. 24th ACM Symp. on the Theory of Comput.* 643-654 (1992).

[Fo75] T.H. Foregger. An upper bound for the permanent of a fully indecomposable matrix. *Proc. Amer. Math. Soc.* 49:319-324 (1975).

[Fr82] S. Friedland. A proof of a generalized van den Waerden conjecture on permanents. *Linear and Multilinear Algebra* 6:227-231 (1982).

[Ga87] J. von zur Gathen. Permanent and determinant. *Linear Algebra and its Applications* 96:87-100 (1987).

[Ga87b] J. von zur Gathen. Feasible arithmetic computations: Valiant's hypothesis. *J. Symbolic Comp.* 4:137-172 (1987).

[GB74] S. Gal and Y. Breitbart. A method for obtaining all the solutions of a perfect matching problem. *IBM Israel Scientific Center Tech. Rep.* 016 (1974).

[GG81] C.D. Godsil and I. Gutman. On the matching polynomial of a graph. *Algebraic Methods in Graph Theory*, Lovász and Sós, editors, Colloq. Math. Soc. János Bolyai, 25, North-Holland, Amsterdam, 241-249 (1981).

[Gi71] Peter Gibson. Conversion of the Permanent into the Determinant. *Proc. Amer. Math. Society* 27:471-476 (1971).

[Gi80] Peter Gibson. Permanental Polytopes of Doubly Stochastic Matrices. *Linear Algebra and its Appl.* 32:87-111 (1980).

[GJ79] M. Garey and D. Johnson. Computers and Intractability. *W.H. Freeman and Company, New York* 1979.

[GKP95] R. Graham, D. Knuth, O. Patashnik. Concrete Mathematics. *Addison-Wesley* (1995).

[GM88] R. Gebauer, and H.M. Möller. On an Installation of Buchberger's Algorithm. *J. Symbolic Computation* 6:275-286 (1988).

[Gu52] E.A. Guggenheim. Mixtures. *Clarendon Press, Oxford* (1952).

[Gy73] Béla Gyires. Discrete distributions and permanents. *Publ. Math. Debrecen* 20:93-106 (1973).

[HL72] O.J. Heilman and E.H. Lieb. Theory of monomer-dimer systems. Communications in Mathematical Physics 115:553-569 (1972).

[HLP34] G.H. Hardy, J.E. Littlewood and G. Pólya. Inequalities. *London* (1934).

[Ho64] F. Holens. Two aspects of doubly stochastic matrices: permutation matrices and the minimum of the permanent function (Thesis abstract). *Canad. Math. Bull.* 7:509-510 (1964).

[HP73] F. Harary and E.M. Palmer. Graphical enumeration. *Academic Press, New York* (1973).

[JR67] W.B. Jurkat and H.J. Ryser. Term ranks and permanents of nonnegative matrices. *J. Algebra* 5:342-357 (1967).

[JS88] M. Jerrum and A. Sinclair. Conductance and the rapid mixing property for Markov chains: the approximation of the permanent resolved (extended abstract). *Proc. of the Twentieth Ann. Symp. on Theory of Computing* 235-243 (1988).

[JS89] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM Journal on Computing* 18:1149-1178 (1989).

[JS93] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal of Computing* 22:1087-1016 (1993).

[JV92] M. Jerrum and U. Vazirani. A mildly exponential approximation algorithm for the permanent. *FOCS* 33:320-326 (1992).

[JVV86] M. Jerrum, L. Valiant, and V.V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical Computer Science* 43:169-188 (1986).

[KKLLL93] N. Karmarkar, R. Karp, R. Lipton, L. Lovász, and M. Luby. A Monte-Carlo algorithm for estimating the permanent. *SIAM Journal on Computing* 22:284-293 (1993).

[KL83] R. Karp and M. Luby. Monte Carlo algorithms for enumeration and reliability problems. *Proc. of the Fifteenth Ann. Symp. on Theory of Computing* (1983).

[KL85] R. Karp and M. Luby. Monte Carlo algorithms for the planar multiterminal reliability problem. *Journal of Complexity* 1:45-64 (1985).

[KLM84] V. Klee, R. Ladner and R. Manber. Signsolvability revisited. *Linear Algebra and its Applications* 59:131-157 (1984).

[KP69] B.W. King, and F.D. Parker. A Fibonacci Matrix and the permanent function. *Fibonacci Quart.* 7:539-544 (1969).

[KRS93] C. Kenyon, D. Randall, and A. Sinclair. Matchings in lattice graphs. *ICSI Technical Report* N. TR-93-019 (1993).

[KS82] P. Knopp and R. Sinkhorn. Minimum permanents of doubly stochastic matrices with at least one zero entry. *Linear and Multilinear Algebra* 11:351-355 (1982).

[Le59] J. Levine. Note on an identity of Cayley. *Amer. Math. Monthly* 66:290-292 (1959).

[Li75] C.H.C. Little. A Characterization of Convertible $(0,1)$-Matrices. *Journal of Combinatorial Theory (B)* 18:187-208 (1975).

[LP84] A.S. LaPaugh and C. Papadimitriou. The Even-Path Problem for Graphs and Digraphs. *Networks* 14:507-513 (1984).

[Lu86] M. Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM J. Computing* 15 (4):1036-1053 (1986).

[Ma64] M. Marcus. The Hadamard theorems for permanents. *Proc. Amer. Math. Soc.* 15:967-973 (1964).

[Me87] R. Meshulam. Private Communication, January 1987.

[Mi63] H. Minc. Upper bounds for permanents of $(0,1)$-matrices. *Bull. Amer. Math. Soc.* 69:789-791 (1963).

[Mi69] H. Minc. On lower bounds for permanents of $(0,1)$-matrices. *Proc. Amer. Math. Soc.* 22:117-123 (1969).

[Mi76] H. Minc. The invariance of elementary symmetric functions. *Linear and Multilinear Algebra* 4:209-215 (1976).

[Mi78] H. Minc. Permanents. *Encyclopedia of Mathematics and its Appl. Vol.* 6 (1978).

[Mi83] H. Minc. Theory of Permanents 1978-1981. *Linear and Multilinear Algebra* 1:227-263 (1983).

[Mi85] H. Minc. Recurrence Formulas for Permanents of $(0, 1)$ Circulants. *Linear Algebra and its Appl.* 71:241-265 (1985).

[Mi87] H. Minc. Permanental Compounds and Permanents of $(0, 1)$ Circulants. *Linear Algebra and its Appl.* 86:11-42 (1987).

[Mi89] M. Mihail. On coupling and the approximation of the permanent. *Information Processing Letters* 30:91-95 (1989).

[MM61] M. Marcus and H. Minc. On the relation between the determinant and the permanent. *Illinois J. Math.* 5:376-381 (1961).

[MM62] M. Marcus and F.C. May. The permanent function. *Canad. J. Math.* 14:177-189 (1962).

[MM64] M. Marcus and H. Minc. Inequalities for general matrix functions. *Bull. Amer. Math. Soc.* 70:308-313 (1964).

[MM67] M. Marcus and H. Minc. On a conjecture of B.L. van der Waerden. *Proc. Cambridge Philos. Soc.* 63:305-309 (1967).

[MN59] M. Marcus and M. Newman. On the minimum of the permanent of doubly stochastic matrices. *Duke Math. J.* 26:61-72 (1959).

[MN62] M. Marcus and M. Newman. Inequalities for the permanent function. *Ann. Math.* 675:47-62 (1962).

[MSS69] N. Metropolis, M.L. Stein, and P.R. Stein. Permanents of cyclic $(0, 1)$ matrices. *J.Combinatorial Theory B* 7:291-321 (1969).

[Mu03] R.F. Muirhead. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proc. Edinburgh Math. Soc.* 21:144-157 (1903).

[Mu882] T. Muir. On a class of permanent symmetric functions. *Proc. Roy. Soc. Edinburgh* 11:409-418 (1882).

[Mu] T. Muir. The Theory of Determinants in the Historical Order of Development. London Vol. I (1906), II (1911), III (1920), IV (1923).

[Mu28] T. Muir. A Treatise on the Theory of Determinants. *London* (1928).

[Mu30] T. Muir. Contributions to the History of determinants. *London* (1930).

[Os70] P.A. Ostrand. Systems of distinct representatives II. *J. Math. Anal. Applics.* 32:1-4 (1970).

[Pa94] Christos H. Papadimitriou. Computational complexity. *Addison-Wesley Publishing Company* (1994).

[Po13] G. Pólya. Aufgabe 424. *Arch. Math. Phys.* (3) 20:271 (1913).

[R63] H.J. Ryser. Combinatorial Mathematics. *Carus Mathematical Monograph No. 14 (1963).*

[Ri45] J. Riordan. Three-line latin rectangles-II. *Amer. Math. Monthly* 53:18-20 (1946).

[Ro35] J.K. Roberts. Some properties of adsorbed films of oxygen and tungsten. *Proceedings of the Royal Society of London A* 152:464-480 (1935).

[Sc18] I. Schur. Uber endliche Gruppen und Hermitesche Formen. *Math. Z.* 1:184-207 (1918).

[Sc78] A. Schrijver. A short proof of Minc's conjecture. *J. Combin. Theory Ser. A* 25:80-83 (1978).

[Si88] A. Sinclair. Randomised algorithms for counting and generating combinatorial structures. *Ph.D thesis, University of Edinburgh, Scotland, june* (1988).

[Sm93] R. Smolensky. On Representations by Low-Degree Polynomials. Proc. of the 34^{th} IEEE Symposium on the Foundations of Computer Science, pp. 130-138, 1993.

[ST87] P. Seymour and C. Thomassen. Characterization of Even Directed Graphs. *J. Combinatorial Theory B* 42:36-45 (1987).

[Th86] C. Thomassen. Sign-nonsingular matrices and even cycles in directed graphs. *Linear Algebra and its Applications* 75:27-41 (1986).

[Ti60] M. Tinsley. Permanents of Cyclic Matrices. Pacific Journal Math. 10:1067-1082 (1960).

[To34] J. Touchard. Sur un problème de permutations. *C.R. Acad. Sci. Paris* 198:631-633 (1934).

[To89] S. Toda. On the computational power of PP and $\oplus$P. *Proc. 30th Ann. IEEE Symp. on Foundations of Computer Science* 514-519 (1989).

[Tv63] H. Tverberg. On the permanent of bistochastic matrices. *Math. Scand.* 12:25-35 (1963).

[V179] L.G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science 8:189-201 (1979).*

[V279] L.G. Valiant. Completeness classes in algebra. *ACM Symp. on the Theory of Comput. 249-261 (1979).*

[Vo79] M. Voorhoeve. A lower bound for the permanents of certain $(0, 1)$-matrices. *Indag. Math.* 41:83-86 (1979).

[VV89] V.V. Vazirani. NC algorithms for computing the number of perfect matchings in $K_{3,3}$-free graphs and related problems. *Information and Computation* 80:152-164 (1989).

[Wa26] B.L. van der Waerden. Aufgabe 45. *Jber. Deutsch. Math. Verein.* 35:117 (1926).

[Wa78] E.T.H. Wang. Permanental pairs of doubly stochastic matrices. *Amer. Math. Monthly* 85:188-189 (1978).

[Wa79] E.T.H. Wang. When is the permanent function convex on the set of doubly stochastic matrices. *Amer. Math. Monthly* 86:119-121 (1979).

[Wi68] H.S. Wilf. A mechanical counting method and combinatorial applications. *Journal of Combinatorial Theory* 4:246-258 (1968).