

CAUCHY-BINET FOR PSEUDO-DETERMINANTS

OLIVER KNILL

ABSTRACT. The pseudo-determinant $\text{Det}(A)$ of a square matrix A is defined as the product of the nonzero eigenvalues of A . It is a basis-independent number which is up to a sign the first nonzero entry of the characteristic polynomial of A . We prove $\text{Det}(F^T G) = \sum_P \det(F_P) \det(G_P)$ for any two $n \times m$ matrices F, G . The sum to the right runs over all $k \times k$ minors of A , where k is determined by F and G . If $F = G$ is the incidence matrix of a graph this directly implies the Kirchhoff tree theorem as $L = F^T G$ is then the Laplacian and $\det^2(F_P) \in \{0, 1\}$ is equal to 1 if P is a rooted spanning tree. A consequence is the following Pythagorean theorem: for any self-adjoint matrix A of rank k , one has $\text{Det}^2(A) = \sum_P \det^2(A_P)$, where $\det(A_P)$ runs over $k \times k$ minors of A . More generally, we prove the polynomial identity $\det(1 + xF^T G) = \sum_P x^{|P|} \det(F_P) \det(G_P)$ for classical determinants \det , which holds for any two $n \times m$ matrices F, G and where the sum on the right is taken over all minors P , understanding the sum to be 1 if $|P| = 0$. It implies the Pythagorean identity $\det(1 + F^T F) = \sum_P \det^2(F_P)$ which holds for any $n \times m$ matrix F and sums again over all minors F_P . If applied to the incidence matrix F of a finite simple graph, it produces the Chebotarev-Shamis forest theorem telling that $\det(1 + L)$ is the number of rooted spanning forests in the graph with Laplacian L .

1. INTRODUCTION

The **Cauchy-Binet theorem** for two $n \times m$ matrices F, G with $n \geq m$ tells that

$$(1) \quad \det(F^T G) = \sum_P \det(F_P) \det(G_P),$$

where the sum is over all $m \times m$ square sub matrices P and F_P is the matrix F masked by the pattern P . In other words, F_P is an $m \times m$ matrix obtained by deleting $n - m$ rows in F and $\det(F_P)$ is a minor

Date: June 10, 2014.

1991 Mathematics Subject Classification. Primary: 15A15, 15A69, 15A09 .

Key words and phrases. Linear algebra, Cauchy-Binet, Binet-Cauchy, pseudo-determinant, pseudo-inverse, matrix tree theorem, multilinear algebra.

of F . In the special case $m = n$, the formula is the product formula $\det(F^T G) = \det(F^T)\det(G)$ for determinants. Direct proofs can be found in [39, 34, 41]. An elegant multi-linear proof is given in [22], where it is called “almost tautological”. A graph theoretical proof using the Lindström-Gessel-Viennot lemma sees matrix multiplication as a concatenation process of directed graphs and determinants as a sum of weighted path integrals [3]. The classical Cauchy-Binet theorem implies the Pythagorean identity $\det(F^T F) = \sum_P \det^2(F_P)$ for square matrices which is also called Lagrange Identity [4], where P runs over all $m \times m$ sub-matrices of F . This formula is used in multivariable calculus in the form $|\vec{v}|^2|\vec{w}|^2 - (\vec{v} \cdot \vec{w})^2 = |\vec{v} \wedge \vec{w}|^2$ and is useful to count the number of basis choices in matroids [2]. The Cauchy-Binet formula assures that the determinant is compatible with the matrix product. Historically, after Leibniz introduced determinants in 1693 [40], and Vandermonde made it into a theory in 1776 [24], Binet and Cauchy independently found the product formula for the determinant around 1812 [8, 14, 45, 9, 12], even before matrix multiplication had been formalized. [16, 24] noticed that Lagrange had mentioned a similar result before in the three dimensional case. The term “matrix” was used by Sylvester first in 1850 [44, 55, 56]. The book [24] mentions that Binet’s proof was not complete. It was Cayley who looked first at the matrix algebra [15, 32, 43, 42]. It is evident today that the Cauchy-Binet formula played a pivotal role for the development of matrix algebra. Early textbooks like [50] illustrate how much the notation of determinants has changed over time.

In this paper, we extend the Cauchy-Binet formula (1) to matrices with determinant 0. Theorem 2) will follow from Theorem 7) and imply new formulas for classical determinants like Theorem 8). The pseudo-determinant $\text{Det}(A)$ for a square matrix A is defined as the product of the nonzero eigenvalues of A , with the assumption $\text{Det}(0) = 1$ for a matrix with all zero eigenvalues like $A = 0$ or nilpotent matrices. The later assumption renders all formulas also true for zero matrices. Looking at singular matrices with pseudo-determinants opens a new world, which formula (1) buries under the trivial identity “ $0 = 0$ ”. The extension of Cauchy-Binet to pseudo-determinants is fascinating because these determinants are not much explored and because Cauchy-Binet for pseudo-determinants is not a trivial extension of the classical theorem. One reason is that the most commonly used form of Cauchy-Binet is false, even for diagonal square matrices A, B : while

$\text{Det}(AB) = \text{Det}(BA)$ is always true, we have in general:

$$\text{Det}(AB) \neq \text{Det}(A)\text{Det}(B)$$

as the example $A = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ shows.

What can be generalized? Because eigenvalues of square matrices C and C^T agree, it is true that $\text{Det}(C) = \text{Det}(C^T)$ for square matrices. In particular, $\text{Det}(F^T G) = \text{Det}(G^T F)$ if F, G are $n \times m$ matrices. It is also true - even so it is slightly less obvious - that $\text{Det}(F^T G) = \text{Det}(FG^T)$. The later follows from the fact that $F^T G$ and FG^T are essentially isospectral, meaning that they have the same nonzero eigenvalues. If F, G are not square, then one of the products has zero eigenvalues so that we need the pseudo-determinant for this identity to be interesting. Experiments indicated us that summing over determinants of square matrices on the right hand side works. We will show:

Theorem 1. *If F, G are $n \times m$ matrices, then*

$$(2) \quad \text{Det}(F^T G) = \sum_P \det(F_P) \det(G_P),$$

where the sum is over all $k \times k$ sub matrix masks P of F , and where k is such that $p_k(-x)^{m-k}$ is the smallest order entry in the characteristic polynomial $p_0(-x)^m + p_1(-x)^{m-1} + \dots + p_k(-x)^{m-k}$ of $F^T G$.

The formula holds also in limiting cases like for nilpotent $F^T G$, in which case $k = 0$ and both sides are 1.

This is more general than the classical Cauchy-Binet theorem. If $F^T G$ is not invertible but FG^T is, then one can use that $\text{Det}(F^T G) = \det(FG^T)$ and use the classical Cauchy-Binet result. But Theorem 1) also applies if both $F^T G$ and FG^T are singular. For a self-adjoint square matrix A of rank k in particular, the formula shows that $\text{Det}(A^2)$ is a sum of squares of determinants $\det^2(B)$ of $k \times k$ sub matrices B . This uses that $\text{Det}^2(A) = \text{Det}(A^2)$, which is one of the identities for normal matrices to be discussed in the next section.

The pairing $\langle F, G \rangle = \text{Det}(F^T G) = \langle G, F \rangle$ leads to a functional $\|F\| = \sqrt{\langle F, F \rangle}$ on $n \times m$ matrices. It is far from a norm, as $\|0\| = 1$, and no triangle inequality nor Cauchy-Schwarz inequality $\langle F, G \rangle \leq \|F\| \|G\|$ holds. The sum of determinants $\langle F, G \rangle$ depends in a nonlinear way on both F and G : if $\lambda \neq 0$, then

$$\langle \lambda F, G \rangle = \lambda^k \langle F, G \rangle = \langle F, \lambda G \rangle,$$

where k is the number of nonzero eigenvalues of $F^T G$ and where we again understand that for $k = 0$, all terms are 1.

The “sphere” $X = \|A\| = 1$ in $M(n, \mathbb{R})$ is a union of Cauchy-Binet varieties $X_n = \text{SL}(n, \mathbb{R}), X_{n-1}, \dots, X_1, X_0$. In the case $M(2, \mathbb{R})$ for example, we have $X = X_2 \cup X_1 \cup X_0 = \text{SL}(2, \mathbb{R}) \cup \{ad - bc = 0, a + d = 1\} \cup N$, where $X_0 = N$ is the set of nilpotent matrices. The set $X_1 = \{a(1 - b) - bc = 1\}$ is a two-dimensional quadric. In the case of diagonal 2×2 matrices, we have $X = X_2 \cup X_1 \cup X_0 = \{|ad| = 1\} \cup \{ad = 0, |a + d| = 1\} \cup \{a = d = 0\}$ which is a union of a hyperbola, four points $\{(\pm 1, 0)\}, \{(0, \pm 1)\}$ and the origin $\{(0, 0)\}$. In the case $M(3, \mathbb{R})$ already, the unit sphere X is the 8-dimensional $X_3 = \text{SL}(3, \mathbb{R})$ together with a 7-dimensional X_2 and a 6-dimensional variety X_1 as well as the variety of nilpotent matrices. The classical Cauchy-Binet theorem misses the three later ones. We see that the case $\det(A) = 0$ is an unexpectedly rich place.

One of the main motivations for pseudo-determinants is graph theory, where the Laplacian matrix L always has a kernel. While $\det(L)$ is zero and therefore not interesting, $\text{Det}(L)$ has combinatorial meaning as it allows to count spanning trees in the graph. The number $\text{Det}(L)$ indeed is a measure for the complexity of the graph. This paper grew while looking at questions in graph theory. The classical Kirchhoff matrix tree theorem is an immediate consequence of Theorem 1) because if $F = G$ is the incidence matrix of a graph then $A = F^T G$ is the scalar Laplacian and $\text{Det}(A) = \text{Det}(F^T G) = \sum_P \det(F_P)^2$. The Kirchhoff matrix tree theorem [7] can rely on the classical Cauchy-Binet theorem for invertible matrices. The reason is that for a connected graph, the kernel of the Laplacian is one-dimensional, so that $\text{Det}(A) = n \cdot \det(M)$, where M is a minor of A which is a classical determinant. The proof can then proceed with the classical Cauchy-Binet theorem for M . We are forced to pseudo determinants when looking at geometric interpretations of trees in the simplex graph \mathcal{G} , which is the graph consisting of all complete subgraphs of a graph where two simplices of dimension difference 1 are connected if one is contained in the other. The operator $L = D^2$ on discrete differential forms [27] will then play an important role as the matrix $|D|$ is the adjacency matrix of \mathcal{G} . The matrix L has a large kernel in general with $\dim(\ker(L)) = \sum_i b_i = b$, the sum of the Betti numbers of the graph which by the Hodge theorem is the total dimension of all harmonic forms. Theorem 1) gives a combinatorial interpretation of what $\text{Det}(L)$ means, if L is the form Laplacian of a graph. The matrix L is large in general: if $G = (V, E)$ has v complete

subgraphs, then \mathcal{G} has v vertices and L is a $v \times v$ matrix. For the complete graph K_n for example, $v = 2^n - 1$. We will see below that also the classical Kirchhoff matrix theorem can be understood more naturally using Cauchy-Binet for pseudo determinants.

Acknowledgements: Thanks to **Fuzhen Zhang** for a pointer on generalized matrix functions, to **Nieves Castro González** for references like [5, 6] and **Woong Kook** for information about more sophisticated matrix tree theorems. Thanks to **Shlomo Sternberg** for a historical discussion on Binet in 2007 which sparked my interest in this part of the history of linear algebra. Added in proof is a reference to [28], where it pointed out that the results of the current paper leads to a proof of the Chebotarev-Shamis forest theorem [48, 46]. Some linear algebra in this context has been summarized in the talk [29].

2. THE PSEUDO-DETERMINANT

The **pseudo-determinant** of a square $n \times n$ matrix A is defined as the product of the nonzero eigenvalues of A with the convention that the pseudo-determinant of a nilpotent matrix is 1. The choice for the later makes all formulas hold true unconditionally. As the sum over the empty set is naturally zero, the product over the empty set is naturally assumed to be 1. With this definition, the pseudo determinant is then the smallest coefficient p_k of the characteristic polynomial $p_A(x) = \det(A - x) = p_0(-x)^n + p_1(-x)^{n-1} + \dots + p_k(-x)^{n-k} + \dots + p_n$. This is 1 for nilpotent A , as $\det(A - x) = (-x)^n$ for such matrices. Another extreme is $\text{Det}(A) = p_n = \det(A)$ if the matrix A is invertible.

We start with some basic facts about the pseudo-determinant $\text{Det}(A)$ of a $n \times n$ matrix A . Most points are obvious, but we did not find any references. Some multi-linear treatment of the pseudo-determinant will appear in the proof section of Theorem 7).

We denote with A^* the adjoint of A and with A^+ the Moore-Penrose pseudo-inverse of A defined by $A^+ = VD^+U^*$ if $A = UDV^*$ is the singular value decomposition (SVD) of A and D^+ is the diagonal matrix which has the same zero entries than D and where $D_{ii}^+ = 1/D_{ii}$ for the nonzero diagonal entries of D . Pseudo-inverses appear textbooks like [54] in the context of SVD. While SVD is not unique, the Moore-Penrose pseudo-inverse is. Denote by $p_A(x) = \det(A - x)$ the characteristic polynomial of A . We use a sign choice used from [54, 10]

and computer algebra systems like Mathematica. A matrix is self-adjoint if $A = A^*$, it is normal if $AA^* = A^*A$. We also denote by Q the unit cube in \mathbb{R}^n and by $|Y|_k$ the k -volume of a k -dimensional parallelepiped Y in \mathbb{R}^n . Let $\Lambda^k A$ denote the k 'th exterior power of A . It is a $\binom{n}{k} \times \binom{n}{k}$ matrix which is determined by

$$\Lambda^k A(e_1 \wedge \cdots \wedge e_k) = (Ae_1 \wedge \cdots \wedge Ae_k).$$

If a basis $\{e_i\}$ in \mathbb{R}^n is given, then $e_I = e_{i_1} \wedge \cdots \wedge e_{i_k}$ with $I \subset \{1, \dots, n\}$ of cardinality k defines a basis of $\Lambda^k \mathbb{R}^n$. The matrix entries of $[\Lambda^k A]_{IJ}$ are $(\Lambda^k Ae_I) \cdot e_J$. The Pfaffian of a real skew-symmetric $2n \times 2n$ matrix A is defined as follows: since A is naturally a 2-form in \mathbb{R}^{2n} , one can look at its n 'th exterior power $A \wedge \cdots \wedge A$ which is a $2n$ form $\text{pf}(A)e_1 \wedge \cdots \wedge e_{2n}$ and consequently defines the real number $\text{pf}(A)$. Since the Pfaffian satisfies $\text{pf}^2(A) = \det(A)$, we can define the pseudo Pfaffian $\text{Pf}(A)$ of a real skew-symmetric matrix A as the product of the nonzero entries λ_i , if A is orthogonally conjugated to a block-diagonal matrix containing diagonal blocks $B_j = \begin{bmatrix} 0 & \lambda_j \\ -\lambda_j & 0 \end{bmatrix}$. The block diagonalization follows from the spectral theorem for real skew-symmetric matrices, which in turn follows from the fact that iA is a complex selfadjoint matrix if A is real skew symmetric and that $S = \begin{bmatrix} 1 & -1 \\ i & i \end{bmatrix}$ conjugates the skew diagonal blocks B_j to $S^{-1}B_jS = \begin{bmatrix} i\lambda_j & 0 \\ 0 & -i\lambda_j \end{bmatrix}$. We understand again that the pseudo Pfaffian satisfies $\text{Pf}(A) = 1$ if $A = 0$. While $\text{pf}(A) = 0$ for all skew $(2n+1) \times (2n+1)$ matrices, the pseudo Pfaffian $\text{Pf}(A)$ is never zero.

Proposition 2. *Assume $A, B \in M(n, \mathbb{C})$.*

- 1) *If A is similar to B then $\text{Det}(A) = \text{Det}(B)$.*
- 2) *If A is invertible, then $\text{Det}(A) = \det(A)$.*
- 3) *If P is an orthogonal projection onto a subspace then $\text{Det}(P) = 1$.*
- 4) *If A is real and normal, then $|\text{Det}(A)| = |A(Q)|_k$ if $k = \text{ran}(A)$.*
- 5) *$\text{Det}(A^T) = \text{Det}(A)$ and $\text{Det}(A^*) = \overline{\text{Det}(A)}$.*
- 6) *Det is discontinuous on $M(n, \mathbb{R})$ or $M(n, \mathbb{C})$ for $n \geq 2$.*
- 7) *The pseudo-inverse of a normal matrix satisfies $\text{Det}(A^+) = 1/\text{Det}(A)$.*
- 8) *$\text{Det}(A) = p_k$, where $p_A(x) = \det(A - x) = p_0(-x)^n + \cdots + p_k(-x)^k$.*
- 9) *$\text{Det}(A) = \text{tr}(\Lambda^k A)$, if A has exactly k nonzero eigenvalues.*
- 10) *For any A and $m \in \mathbb{N}$, one has $\text{Det}(A^m) = \text{Det}^m(A)$.*
- 11) *For $A, B \in M(n, \mathbb{R})$ then $\text{Det}(A^T B) = \text{Det}(A B^T) = \text{Det}(B A^T)$.*
- 12) *$\text{Det}(A)$ is never zero and if $A = A^*$, then $\text{Det}(A)$ is real.*

- 13)** For block diagonal $A = \text{Diag}(A_1, \dots, A_k)$, $\text{Det}(A) = \prod_i \text{Det}(A_i)$.
14) For real skew A , the pseudo Pfaffian satisfies $\text{Pf}^2(A) = \text{Det}(A)$.

Proof. **1)** The eigenvalues of similar matrices are the same.

2) We use the definition and the fact that the classical determinant is the product of the eigenvalues.

3) If we diagonalize P , we get a matrix with only 1 or 0 in the diagonal. In the case of $P = 0$ the result holds also by definition.

4) $\text{Det}(A)$ is basis independent and self-adjoint matrices can be diagonalized. For an orthogonal projection in particular, the pseudo-determinant is 1.

5) The eigenvalues are the same for A and A^T and A^* has the complex conjugate eigenvalues of A .

6) The functional is already discontinuous for $n = 2$, where

$$\text{Det}\left(\begin{bmatrix} a & 0 \\ 0 & 2 \end{bmatrix}\right) = 2a \text{ for } a \neq 0 \text{ and } \text{Det}\left(\begin{bmatrix} a & 0 \\ 0 & 2 \end{bmatrix}\right) = 2 \text{ for } a = 0.$$

7) Use that the pseudo inverse has the nonzero eigenvalues λ_j^{-1} if λ_j are the eigenvalues of A . For non normal matrices this can be false:

the matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ has the eigenvalues 0, 1. Its pseudo inverse

$$A^+ = \begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \end{bmatrix} \text{ has eigenvalues } 0, 1/2.$$

8) Write $p_A(x) = x^k \prod_j (\lambda_j - x)$, where λ_j runs over the set of nonzero eigenvalues.

9) This follows from the previous step and the fact that $\text{tr}(\Lambda^k A) = p_k$, a fact which can be deduced from $\det(1 + A) = \sum_{j=0}^n \text{tr}(\Lambda^j A)$ (see i.e. [49] p.322), an identity which allows to define the determinant $\det(1 + A)$ in some infinite dimensional setups. The result holds also in the nilpotent case $k = 0$ as $\Lambda^0(A)$ is the 1×1 matrix 1.

10) Normal matrices A can be diagonalized over the field of complex numbers: there exists a unitary U such that U^*AU is diagonal. In general, we can bring the matrix in Jordan normal form. The result is also true for $n = 0$, where both sides are 1.

11) The matrices $A^T B$ and AB^T have the same nonzero eigenvalues because their characteristic polynomials differ by a factor λ^k only. For the last identity, use 5).

12) The first statement is by definition and can also be seen from 8) as the characteristic polynomial is never zero. For $A = A^*$, then the eigenvalues of A are real by the spectral theorem.

13) Group the eigenvalues of A as eigenvalues of A_i . Also here, the formula works if some of the A_i are nilpotent.

14) A is orthogonally conjugated to a block diagonal matrix B . Now square this and use part 1). \square

Remarks.

1) We can compute pseudo-determinants almost as fast as determinants because we only need to know the characteristic polynomial. Some computer algebra code can be found in the Appendix. We can find $\text{Det}(A)$ also by row reduction if we do safe row reduction steps. As mentioned below, we have to make sure that we do not count any sign changes when swapping two parallel rows and do scalings of dependent rows, nor subtract a row from a parallel row. When doing safe row reductions, we end up with a matrix which looks like a row reduced echelon matrix but where parallel rows can appear. For such a reduced matrix, the eigenvalues can be computed fast.

2) The least square solution formula $Ax = A(A^T A)^{-1} A^T y = Py$ is pivotal in statistics and features a projection matrix P with pseudo-determinant 1. We mention this because the least square inverse is often also called pseudo inverse even so it is not Moore-Penrose pseudo inverse in general. The former deals with overdetermined systems $Ax = y$, the later is defined for square matrices A only.

3) Define $\log^+ |x| = \log |x|$ for $x \neq 0$ and $\log^+ |0| = 0$. If A is normal, we can define $\log^+ |A|$ by diagonalizing A and get

$$\text{tr}(\log^+ |A|) = \log |\text{Det}(A)| ,$$

Also this is consistent with nilpotent A and an other reason why the definition $\text{Det}(A) = 1$ for nilpotent A is natural.

4) For finite simple graphs, the Laplace operator L always has a kernel. It is one-dimensional if the graph is connected. The pseudo-determinant is considered a measure for complexity because it allows to count the number of maximal spanning trees in the graph. The Laplace-Beltrami operator on forms has a large kernel in general. Its dimension is the sum of the Betti numbers of the graph. Studying this matrix associated to a graph was the main motivation for us to look at Cauchy-Binet in the singular case.

5) The fact that $F^T G$ and $F G^T$ have the same nonzero eigenvalues is a consequence of $p_A(x) = (-x)^k p_B(x)$ if $A = F^T G$ and $B = F G^T$. For $x = -1$, this leads to “the most important identity in mathematics (Deift)” [57] $\det(1 + FG) = \det(1 + GF)$ for $n \times m$ and $m \times n$ matrices F, G . [17] illustrates how rich such identities can be. Switching operators is useful to construct the spectrum of the quantum harmonic oscillator, for the numerical QR algorithm $A = QR \rightarrow RQ$ used to diagonalize a matrix, or to construct new solutions to nonlinear PDE's

with so called Bäcklund transformations.

Lets look at some pitfalls:

1) One could think that if B is invertible, then $\text{Det}(AB) = \text{Det}(A)\det(B)$.

This is false: a counter example is $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. It

is even false for unitary B like $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, where

$AB = A$ and $\det(B) = -1$. This example is the row swapping pitfall using the elementary swapping matrix B . We mention this pitfall because we actually tried to prove this first by repeating the textbook proof of $\det(AB) = \det(A)\det(B)$ in which one makes row reduction on the augmented matrix $[A|B]$ until the matrix B is the identity matrix. As we see below, pseudo-determinants need safe row reduction operations.

2) Even basic multi-linearity fails, even if we apply it to nonzero rows.

An example like $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ shows that if we scale a row by λ then

the pseudo-determinant gets scaled by $(1 + \lambda)/2$ and not by λ . This is just an example and not a general rule. It is difficult to say in general how scaling a linearly dependent row affects the pseudo-determinant.

3) The case of block diagonal matrices $A = \text{Diag}(A_1, A_2)$ with square matrices A_1, A_2 again illustrates that it is useful to define $\text{Det}(A) = 1$ for nilpotent matrices. This renders $\text{Det}(\text{Diag}(A_1, A_2)) = \text{Det}(A_1) \cdot \text{Det}(A_2)$ true. Without the assumption like assuming $\text{Det}(0) = 0$, this block diagonal result would be wrong.

4) While $F^T G$ and FG^T are essentially isospectral and therefore have the same number k of nonzero eigenvalues, it is not true in general that $\text{rank}(F^T G)$ and $\text{rank}(FG^T)$ are the same. It is possible that k is smaller than both: let F be the identity matrix I_3 and let $G =$

$\begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \\ -2 & 2 & -1 \end{bmatrix}$, then $p_G(x) = -x^2 - x^3$ so that we have to take $k = 1$

in Theorem 1) and $\text{Det}(F^T G) = -1$. But $\text{rank}(F^T G) = \text{rank}(FG^T) = 2$.

3. EXAMPLES

1) An extreme example of Theorem 1) is obtained when F, G are two column vectors in \mathbb{R}^n and where the left hand side is the “dot product”

$F^T G$. It agrees with

$$\text{Det} \begin{bmatrix} F_1 G_1 & F_1 G_2 & \cdots & F_1 G_n \\ F_2 G_1 & F_2 G_2 & \cdots & F_2 G_n \\ \cdots & \cdots & \cdots & \cdots \\ F_n G_1 & F_n G_2 & \cdots & F_n G_n \end{bmatrix}$$

which can be seen as the pseudo determinant analogue of a **Gram determinant** [19] and which has the only non-zero eigenvalue $F^T G$. While the characteristic polynomial of the 1×1 matrix $A = F^T G$ is $p_A(x) = x - \lambda$, the characteristic polynomial of the $n \times n$ matrix FG^T is $(-x)^{n-1}(x - \lambda)$.

2) Assume F is an $n \times m$ matrix for which every row v_j of F is a multiple a_j of v and that G is an $n \times m$ matrix for which every row w_j of G is a multiple b_j of a vector w . Then $\text{Det}(F^T G) = (v \cdot w)(a \cdot b)$. This is the same than the right hand side of Cauchy-Binet. Since there is only nonzero eigenvalue, it has to be the trace of $F^T G$. The later matrix has the entries $(v \cdot w)a_i b_j$. The left hand side of Cauchy-Binet is $(\sum_i v_i w_i)(\sum_j a_j w_j)$. The right hand side of Cauchy-Binet is $\sum_{i,j} (a_j v_i)(b_j w_i)$. These two sums are the same. The same works also if F is rank 1 and G is arbitrary. Since $F^T G$ has rank 1, we have $\text{Det}(F^T G) = \text{tr}(F^T G) = \sum_{i,j} F_{ij} G_{ij}$.

3) The two 3×2 matrices

$$F = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}, G = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$$

give

$$F^T G = \begin{bmatrix} 6 & 2 \\ 15 & 5 \end{bmatrix}, FG^T = \begin{bmatrix} 1 & 5 & 1 \\ 2 & 7 & 2 \\ 3 & 9 & 3 \end{bmatrix}.$$

Both are singular matrices with the same determinant $\text{Det}(F^T G) = \text{Det}(FG^T) = 11$. If F, G are $n \times 2$ matrices with column vectors a, b and c, d , then the classical Cauchy-Binet identity is

$$(a \cdot c)(b \cdot d) - (a \cdot d)(b \cdot c) = - \sum_{1 \leq i, j \leq n} (a_i b_j - a_j b_i)(c_i d_j - c_j d_i)$$

which has the special case $|a|^2|c|^2 - (a \cdot c)^2 = |a \times c|^2$ in three dimensions for the cross product which expresses the Pythagoras theorem $1 - \cos^2(\alpha) = \sin^2(\alpha)$. Assume now that one of the matrices has rank 1 like in the case $a = c$. The classical identity is then 0 on both sides. The

new Cauchy-Binet identity is not interesting in this case: since $F^T G$ has the eigenvalues $0, a(\sum_i b_i + \sum_i c_i)$, we have $\text{Det}(F^T G) = a(\sum_i b_i + \sum_i c_i)$. The right hand side is $\sum_{|P|=1} \det(F_P) \det(G_P) = \sum_i ab_i + ac_i$.

4) For

$$F = \begin{bmatrix} 3 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}, G = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 3 & 2 \end{bmatrix}$$

we have

$$F^T G = \begin{bmatrix} 0 & 6 & 10 \\ 0 & 6 & 6 \\ 0 & 6 & 6 \end{bmatrix}, FG^T = \begin{bmatrix} 2 & 5 \\ 4 & 10 \end{bmatrix},$$

which both have the pseudo-determinant 12. Now, the 2×2 square sub-matrices of F and G are

$$F_1 = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}, F_2 = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}, F_3 = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

$$G_1 = \begin{bmatrix} 0 & 0 \\ 0 & 3 \end{bmatrix}, G_2 = \begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix}, G_3 = \begin{bmatrix} 0 & 2 \\ 3 & 2 \end{bmatrix}.$$

But all their products

$$F_1^T \cdot G_1 = \begin{bmatrix} 0 & 6 \\ 0 & 6 \end{bmatrix}, F_2^T \cdot G_2 = \begin{bmatrix} 0 & 10 \\ 0 & 6 \end{bmatrix}, F_3^T \cdot G_3 = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}$$

have determinant 0. The reason is that while $\text{rank}(F) = \text{rank}(G) = \text{rank}(F^T G) = 2$, we have $\text{rank}(FG^T) = 1$. We see that we have to take the sum over the products $\det(F_P^T G_P)$ where P runs over all 1×1 matrices. And indeed, now the sum is 12 too. This example shows that even so $F^T G$ and FG^T can have different rank, their pseudo-determinants are still the same. Of course, this follows in general from the fact that all the nonzero eigenvalues are the same.

5) For non-invertible non-nilpotent 2×2 matrices A , we have $\text{Det}(A) = \text{tr}(A)$. For example,

$$\text{Det}\left(\begin{bmatrix} 5 & 6 \\ 10 & 12 \end{bmatrix}\right) = 17.$$

The same is true for matrices for which all columns are parallel like

$$\text{Det}\left(\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{bmatrix}\right) = \text{tr}(A) = 10.$$

6) The two matrices

$$F = \begin{bmatrix} 2 & 5 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, G = \begin{bmatrix} 3 & 7 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

have different rank. The determinant of $F^T G = \begin{bmatrix} 6 & 14 \\ 15 & 35 \end{bmatrix}$ is equal to its trace 41, as seen in the previous example. This agrees with the sum $\sum_{|P|=1} \det(F_P) \det(G_P) = 2 \cdot 3 + 5 \cdot 7 = 41$.

7) If A is a square Vandermonde matrix defined by n numbers a_i , then the determinant of A is $\prod_{i < j} (a_i - a_j)$. If the numbers are not different, like for

$$A = \begin{bmatrix} 1 & 1 & 1 \\ a & b & a \\ a^2 & b^2 & a^2 \end{bmatrix},$$

we have $\text{Det}(A) = (b - a)(b + a + ab)$ if $a \neq b$ and $\text{Det}(A) = \text{tr}(A) = 1 + a + a^2$ if $a = b$.

8) As defined, the result also holds for cases like the nilpotent $F = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$ and $G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ in which case $\text{Det}(F^T G) = \text{Det}(F) = 1$ and $\sum_{|P|=0} \det(F_P) \det(G_P) = 1$.

4. CONSEQUENCES

Theorem 1) implies a symmetry for the nonlinear pairing $\langle F, G \rangle = \text{Det}(F^T G)$:

Corollary 3 (Duality). *If F, G are matrices of the same shape, then $\text{Det}(F^T G) = \text{Det}(FG^T)$.*

Proof. The characteristic polynomials of $F^T G$ and FG^T satisfy $p_{F^T G}(x) = \pm x^l p_{FG^T}(x)$ for some integer l . \square

Remarks:

1) The matrices $F^T G$ and FG^T have in general different shapes. The result is also true in the square matrix case with the usual determinant because both sides are then $\det(F) \det(G)$.

2) We know that $M(n, m)$ with inner product $\text{tr}(F^* G)$ is a Hilbert space with the Hilbert-Schmidt product $\sum_{i,j} \overline{F_{ij}} G_{ij}$ which is sometimes called Frobenius product. While the pairing $\langle A, B \rangle = \text{Det}(A^T B)$ is not an inner product, it inspires to define $\|A\|^2 := \text{Det}(A^T A) > 0$. It

satisfies $\langle A, B \rangle = \langle B, A \rangle$, $d(A, B) = \|A - B\|$ is not a metric as the triangle inequality does not hold and $\|0\|^2 = 1$.

Corollary 4 (Kirchhoff). *For a connected finite simple graph with Laplacian L and m vertices, $\text{Det}(L)/m$ is the number of spanning trees in the graph.*

Proof. If the graph (V, E) has $|V| = m$ vertices and $|E| = n$ edges, then the incidence matrix F is a $n \times m$ matrix. This matrix depends on a given orientation of the edges and is a discrete gradient which assigns to a function on vertices a function on edges with $Ff((a, b)) = f(b) - f(a)$. Its adjoint F^T is a discrete divergence. The combination $L = F^T F = \text{div}(\text{grad}(f))$ is the scalar Laplacian and agrees with the matrix $B - A$, where B is the diagonal vertex degree matrix and A is the adjacency matrix. Unlike F , the matrix L does not depend on the chosen edge orientation. A connected graph L has only one eigenvalue 0 with the constant eigenvector 1. The matrix L therefore has rank $k = n - 1$. Choosing a pattern P means to choose a vertex q as a root and to select $m - 1$ edges. If F_P has determinant 1 or -1 , it represents a connected tree rooted at q because we have $m - 1$ edges and m vertices. Theorem 1) equals $\text{Det}(L)$ and counts the rooted spanning trees and consequently, $\text{Det}(L)/m$ is the number of spanning trees in the graph. \square

Here is an obvious corollary of Theorem 1):

Corollary 5. *If A is any matrix of rank k , then $\text{Det}(A^T A) = \text{Det}(A A^T) = \sum_P \det^2(A_P)$, where A_P runs over all $k \times k$ minors.*

Proof. This is a special case of Theorem 1), where $F = G = A$ using the fact that the two matrices $A^T A$ and $A A^T$ have the same rank k , if A has rank k . \square

Especially, we have a Pythagoras theorem for pseudo-determinants:

Corollary 6 (Pythagoras). *For a selfadjoint matrix A of rank k , then*

$$\text{Det}^2(A) = \text{Det}(A^2) = \sum_P \det^2(A_P),$$

where A_P runs over all $k \times k$ minors of A .

Proof. Use that $\text{Det}^2(A) = \text{Det}(A^2) = \text{Det}(A A^T)$ if A is selfadjoint, as a consequence of 10) in Proposition 2). \square

Remarks.

1) If A is invertible, [13] uses a special case of this identity that if A is a $(n - 1) \times n$ matrix, then $\det(A A^T) = nS$ where S is a perfect square.

2) As in the classical case, this result can be interpreted geometrically: the square of the k -volume of the parallelepiped spanned by the columns of A is related to the squares of the volumes of projections of AQ_I onto planes spanned by Q_J where for $I = \{i_1, \dots, i_k\}$ the set Q_I is the parallelepiped spanned by e_{i_1}, \dots, e_{i_k} and the subsets I, J of $\{1, \dots, n\}$ encode a $k \times k$ sub mask P of A . (See [41]).

3) Theorem 6) is obvious for diagonal matrices. An alternative proof could be to show that the right hand side is basis independent. But this exactly needs to go through the multi-linear approach which sees the right hand side as a Hilbert-Schmidt norm of a k -Fermion matrix.

Examples.

1) If $A = \begin{bmatrix} a & b \\ a & b \end{bmatrix}$, then $\text{Det}(A^T A) = 2a^2 + 2b^2$ and agrees with the right hand side of Corollary 6).

2) If A is invertible, the Pythagoras formula is trivial and tells $\det(A^2) = \det^2(A)$. If A has everywhere the entry a then the left hand side is $(na)^2$ and the right hand side adds up n^2 determinant squares a^2 of 1×1 matrices.

3) Corollary 6) can fail for non-symmetric matrices. The simplest case is $A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$, for which $\text{Det}(A) = 1$ but $\text{Det}(A^T A) = \sum_{|P|=1} \det^2(A_P) =$

5. The correct formula for non-selfadjoint matrices is $\text{Det}(A^T A) = \sum_P \det^2(A_P)$ as in Corollary 5). This example illustrates again that $\text{Det}(A^T A) \neq \text{Det}^2(A)$ in general.

4) The matrix

$$A = \begin{bmatrix} 0 & 4 & 4 \\ 4 & 0 & 3 \\ 4 & 3 & 6 \end{bmatrix}$$

has the characteristic polynomial $41x + 6x^2 - x^3$ and has the eigenvalues $0, (3 \pm 5\sqrt{2})$. The pseudo-determinant is -41 so that $\text{Det}^2(A) = 1681$. Now let's look at the determinants of all the 2×2 sub-matrices of A .

$$\begin{aligned}
A_1 &= \begin{bmatrix} 0 & 4 \\ 4 & 0 \end{bmatrix} & A_2 &= \begin{bmatrix} 0 & 4 \\ 4 & 3 \end{bmatrix} & A_3 &= \begin{bmatrix} 4 & 4 \\ 0 & 3 \end{bmatrix} \\
A_4 &= \begin{bmatrix} 0 & 4 \\ 4 & 3 \end{bmatrix} & A_5 &= \begin{bmatrix} 0 & 4 \\ 4 & 6 \end{bmatrix} & A_6 &= \begin{bmatrix} 4 & 4 \\ 3 & 6 \end{bmatrix} \\
A_7 &= \begin{bmatrix} 4 & 0 \\ 4 & 3 \end{bmatrix} & A_8 &= \begin{bmatrix} 4 & 3 \\ 4 & 6 \end{bmatrix} & A_9 &= \begin{bmatrix} 0 & 3 \\ 3 & 6 \end{bmatrix}
\end{aligned}$$

The determinant squares 256, 256, 144, 256, 256, 144, 144, 144, 81 add up to 1681.

5) The matrix

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

has $\text{Det}(A) = 4$. From the $15^2 = 225$ sub-matrices of A of size 4×4 , there are 16 which have nonzero determinant. Each of them either has determinant 1 or -1 . The sum of the determinants squared is 16.

6) $F = [1 \ 2 \ 3 \ 4]$ then

$$A = F^T F = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \\ 4 & 8 & 12 & 16 \end{bmatrix}$$

is selfadjoint with $(\text{Det}(A))^2 = 30^2 = 900$. Since $k = 1$, the right hand side of Pythagoras is the sum of the squares of the matrix entries. This is also 900. This can be generalized to any row vector F for which the identity reduces to the obvious identity $(\sum_{i=1}^n F_i^2)^2 = (\sum_{i,j=1}^n (F_i F_j)^2)$ which tells that the Euclidean norm of F is the Hilbert-Schmidt norm of $A = F^T F$.

5. A GENERALIZATION

We have the Hilbert-Schmidt identity

$$p_1 = \text{tr}(F^T G) = \sum_{|P|=1} \det(F_P) \det(G_P) = \sum_{i,j} F_{ij} G_{ij} ,$$

where F_P is the sub-matrix matched by the pattern P and $|P| = k$ means that P is a $k \times k$ matrix and where

$$(3) \quad p(x) = p_0(-x)^m + p_1(-x)^{m-1} + \cdots + p_k(-x)^{m-k} + \cdots + p_m$$

is the characteristic polynomial of the $m \times m$ matrix $F^T G$ with $p_0 = 1$. Having seen an identity holds for $k = 1$ (and see that it is the classical Cauchy-Binet for $k = m$) we can try for general $0 \leq k \leq m$. Experiments prompted to generalize the result. Theorem 1) is the special case of the following, when k is the minimal k for which p_k is not zero:

Theorem 7 (Generalized Cauchy-Binet). *If F, G are arbitrary $n \times m$ matrices and $0 \leq k$, then*

$$p_k = \sum_{|P|=k} \det(F_P) \det(G_P),$$

where the sum is over k -minors and where p_k are the coefficients of the characteristic polynomial (3) of $F^T G$. This implies the polynomial identity

$$\det(1 + zF^T G) = \sum_P z^{|P|} \det(F_P) \det(G_P)$$

in which the sum is over all minors A_P including the empty one $|P| = 0$ for which $\det(F_P) \det(G_P) = 1$.

We will prove this in the next section. The second statement follows then from the fact that for $z = (-x)^{-1}$ and any $m \times m$ matrix $F^T G$, then

$$\det(F^T G - x) = \sum_P (-x)^{m-|P|} \det(F_P) \det(G_P).$$

Note again that the polynomial identity in the theorem holds also for $z = 0$, where both sides are 1 as $|P| = 0$ contributes 1 to the sum.

In the next corollary, we again sum over all minors, including the empty set P for which $\det(F_P) = 1$:

Corollary 8. *If F, G are two general $n \times m$ matrices, then*

$$\det(1 + F^T G) = \sum_P \det(F_P) \det(G_P),$$

where the right hand side sums over all minors.

Proof. Just evaluate the polynomial identity in Theorem 7) at $z = 1$. □

This again implies the Sylvester determinant theorem $\det(1 + F^T G) = \det(1 + G F^T)$. Especially,

Corollary 9. *If F is an arbitrary $n \times m$ matrix, then*

$$\det(1 + F^T F) = \sum_P \det^2(F_P)$$

where the right hand side sums over all minors.

Examples:

1) If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $\det(1 + A^T A) = a^2 d^2 + a^2 - 2abcd + b^2 c^2 + b^2 + c^2 + d^2 + 1$ which is equal to $1 + a^2 + b^2 + c^2 + d^2 + (ad - bc)^2$.

2) If $v = [v_1, \dots, v_n]$ is a $n \times 1$ matrix, then $1 + v^T \cdot v = 1 + \sum_k v_k^2$.

3) If $v = [a, b]^T$ is a 1×2 matrix we have $\det\left(\begin{bmatrix} 1 + a^2 & ab \\ ab & 1 + b^2 \end{bmatrix}\right) = 1 + a^2 + b^2$.

4) If A is a symmetric $n \times n$ matrix then $\det(1 + A^2) = \sum_P \det^2(A_P)$. This is useful if $A = D = (d + d^*)$ is the Dirac matrix of a graph and $A^2 = L$ is the form Laplacian. Since A_P can have several nonzero permutation patterns, the interpretation using rooted spanning trees is a bit more evolved, using double covers of the simplex graph.

5) If F is the incidence matrix for a graph so that $F^T F$ is the scalar Laplacian then $\det(F_P)^2 \in \{1, 0\}$ and choosing the same number of edges and vertices in such a way that every edge connects with exactly one vertex and so that we do not form loops. These are rooted forests, collections of rooted trees. Trees with one single vertex are seeds which when included, lead to rooted spanning forests.

Corollary 10 (Chebotarev-Shamis). *$\det(1 + L)$ is the number of rooted spanning forests in a connected finite simple graph G with scalar Laplacian L .*

Proof. Again, we assume that the graph (V, E) has $|V| = m$ vertices and $|E| = n$ edges so that the incidence matrix F is a $n \times m$ matrix. The condition $m \leq n$ follows from the connectedness. The Laplacian L is of the form $L = F^T F$ where F is gradient defined by Poincaré and F^T the divergence. Corollary 9) equates $\det(1 + L)$ with the sum over all possible determinants $\det^2(F_P) \in \{0, 1\}$. Now if $|P| = m$, then $\det(F_P) = 0$ as L has a kernel. If $|P| = m - 1$ we count all rooted spanning trees. For $|P| = n - 2$ we count all cycle free subgraphs with $n - 2$ edges. These are rooted spanning forests as individual trees are no more necessarily connected and even can be seeds which are graphs without edges. Since a forest with $n - k$ edges has Euler characteristic

$|V| - |E| = n - (n - k) = k = b_0 - b_1 = b_0$ by Euler Poincaré, where b_0 is the number of connectivity components, we see that a spanning forest with $n - k$ edges has k base points. These are the roots of the individual trees. The sum also includes the case $k = m$, where all trees are seeds and which corresponds to the case $|P| = 0$ in the sum. \square

See [48, 46] for the discovery and [28], for more details and examples.

6. PROOF

While the proof of Theorem 7) uses standard multi-linear algebra, we need to fix some notation.

Recall that $\Lambda^k F$ is the k 'th exterior power of a $n \times m$ matrix F . It is a linear transformation from the vector space $\Lambda^k \mathbb{R}^m$ to the vector space $\Lambda^k \mathbb{R}^n$. The linear transformation $\Lambda^k F$ can be written as a $\binom{n}{k} \times \binom{m}{k}$ matrix if a basis is given in \mathbb{R}^m and \mathbb{R}^n . If $k = 0$, then $\Lambda^0 F$ is just the scalar 1.

While the matrix entries F_{ij} are defined for $1 \leq i \leq n, 1 \leq j \leq m$, the indices of $\Lambda^k F$ are given by subsets I of $\{1, \dots, n\}$ and subsets J of $\{1, \dots, m\}$ which both have cardinality k . If e_j are basis elements in \mathbb{R}^m and f_j are basis elements in \mathbb{R}^n , then $e_I = e_{i_1} \wedge \dots \wedge e_{i_k}$ and $f_J = f_{j_1} \wedge \dots \wedge f_{j_k}$ are basis vectors in $\Lambda^k \mathbb{R}^m$ and $\Lambda^k \mathbb{R}^n$. Write $\Lambda^k F e_I$ for $F e_{i_1} \wedge \dots \wedge F e_{i_k}$. It is possible to simplify notation and just write F_{IJ} for the matrix entry of $(\Lambda^k F)_{IJ}$. It is a real number given by $F e_I \cdot f_J$. Define also $F_P = F_{P(IJ)}$ for the matrix with pattern $P = P(IJ)$ defined by the sets $I \subset \{1, \dots, n\}$ and $J \subset \{1, \dots, m\}$. Finally, in the case of a square matrix A , write $\text{Tr}(A) = \sum_K A_{KK}$ when summing over all subsets K .

The following lemma tells that F_{IJ} is a minor

Lemma 11 (Minor). *If $\Lambda^k \mathbb{R}^n$ and $\Lambda^k \mathbb{R}^m$ are equipped with the standard basis $e_I = e_{i_1} \wedge \dots \wedge e_{i_k}$ obtained from a basis e_i , then*

$$F_{IJ} = \det(F_{P(IJ)})$$

for any sets I, J of the same cardinality $|I| = |J| = k$.

Proof. This is just rewriting $F_{IJ} = \langle F e_I, f_J \rangle$ using the basis vectors $e_I = e_{i_1} \wedge \dots \wedge e_{i_k}$ and $f_J = f_{j_1} \wedge \dots \wedge f_{j_k}$ and the definition of $\Lambda^k F : \Lambda^k \mathbb{R}^m \rightarrow \Lambda^k \mathbb{R}^n$. \square

Lemma 11) implies the trace identity:

Corollary 12 (Trace). *For any $n \times n$ square matrix A and $0 \leq k \leq n$ we have*

$$\mathrm{tr}(\Lambda^k A) = \sum_P \det(A_P) ,$$

where the sum is over all $k \times k$ sub matrices A_P .

Note that unlike in Cauchy-Binet, we have just summed over diagonal minors $A_{P(II)}$. In the case $k = 0$, we understand the left hand side to be the 1×1 matrix $\Lambda^0 A = 1$ and the right hand side to sum over the empty pattern P which is 1 as we postulated $\det(F_P) = 1$ for $|P| = 0$.

Lemma 13 (Multiplication). *For $n \times m$ matrix F and an $m \times n$ matrix G*

$$(FG)_{IJ} = \sum_K F_{IK} G_{KJ}$$

where $I, J \subset \{1, \dots, n\}$ of cardinality k and where the sum is over all $K \subset \{1, \dots, m\}$ of cardinality k .

Proof. This lemma appears as Theorem A.2.1) in [21]. As pointed out in [22] Lemma 10.16), it rephrases the classical matrix multiplication of the two exterior products $\Lambda^k A, \Lambda^k B$ as a composition of maps $\Lambda^k \mathbb{R}^n \rightarrow \Lambda^k \mathbb{R}^m \rightarrow \Lambda^k \mathbb{R}^n$. For more on the exterior algebra, see [1, 20]. \square

We can now prove Theorem 7):

Proof. We use $p_k(A) = \mathrm{tr}(\Lambda^k A)$ for any square matrix A , Lemma 11) and Lemma 13):

$$\begin{aligned} p_k(F^T G) &= \mathrm{tr}(\Lambda^k(F^T G)) = \mathrm{tr}(\Lambda^k F^T \Lambda^k G) \\ &= \sum_J \left(\sum_I F_{JI}^T G_{IJ} \right) \\ &= \sum_J \sum_I F_{IJ} G_{IJ} \\ &= \sum_J \sum_I \det(F_{P(IJ)}) \det(G_{P(IJ)}) \\ &= \sum_P \det(F_P) \det(G_P) . \end{aligned}$$

\square

Theorem 7) implies

Corollary 14 (Pythagoras). *For any self-adjoint $n \times n$ matrix A and any $0 \leq k \leq n$ we have*

$$p_k = \mathrm{tr}(\Lambda^k A^2) = \sum_{|P|=k} \det^2(A_P) ,$$

where the sum is taken over all minors $A_{P(IJ)}$ with $|I| = k, |J| = k$ and where p_k is a coefficient of the characteristic polynomial $p_0(-x)^n + p_1(-x)^{n-1} + \dots + p_n$ of A^2 .

Remarks.

1) Despite the simplicity of the proof and similar looking results for minor expansion, formulas in multi-linear algebra [33, 4], condensation formulas, trace ideals [53], matrix tree results [7], formulas for the characteristic polynomial [37, 47, 11], pseudo inverses, non-commutative generalizations [51], we are not aware that even the special case of Corollary 6) for the pseudo determinant has appeared anywhere already. In the classical case, where A is invertible, the Pythagorean identity is also called Lagrange Identity [4].

2) Pythagoras (14) should be compared with the **trace identity** given in Lemma 12). The former deals with a product of matrices and is therefore a “quadratic” identity. The trace identity deals with one matrix only and does not explain Pythagoras yet. For $k = 1$, the Cauchy-Binet formula and the trace identity reduce to the definition of the matrix multiplication and the definition of the trace: $\text{tr}(F^T G) = \sum_{i,j} F_{ij} G_{ij}$ and $\text{tr}(A) = \sum_i A_{ii}$. If k is the rank of A , then Cauchy-Binet is Theorem 1) and the trace identity is the known formula $\text{Det}(A) = \text{tr}(\Lambda^k A)$, where k is the rank of A .

7. ROW REDUCTION

One can try to prove Theorem 1) by simplifying both sides of

$$\text{Det}(F^T G) = \sum_P \det(F_P) \det(G_P),$$

by applying row operations on F and G and using that both sides of the identity are basis independent. We will only illustrate this here as a proof along those lines is probably much longer. This section might have some pedagogical merit however and also traces an earlier proof attempt. The strategy of row reduction is traditionally used in the proof of Cauchy-Binet. In the special case of the product identity $\det(AB) = \det(A)\det(B)$ already, there is a proof which row reduces the $n \times 2n$ matrix $[A|B]$. This strategy does not generalize, as any of the three row reduction steps are false for pseudo-determinants. Lets explain the difficulty: classical row reduction of a matrix A consists of applying swap, scale or subtract operations to A to bring a $n \times m$ matrix into row reduced echelon form. Exercises in textbooks like [36, 10]) ask to prove that the end result $\text{rref}(A)$ is independent of the strategy with which these steps are applied. When applying row

reduction to a nonsingular matrix until the identity matrix is obtained, then $\det(A) = (-1)^r / \prod \lambda_j$, where r is the number of swap operations used and λ_j are the scaling constants which were applied during the elimination process. While this is all fine for \det , for Det it is simply false. Multiplying a zero row with λ or swapping two zero rows does not alter the matrix and does therefore not contribute to the $(-1)^r$ or scaling factors. For example, swapping the two rows of

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

does not change the pseudo-determinant, nor does a multiplication of the second row by λ . As far as the “subtraction” part of row reduction, there are more bad news: unlike for the determinant, subtracting a row from an other row can change the pseudo-determinant. For example,

$$\text{Det}\left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) = 2, \text{Det}\left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\right) = 1.$$

These problems can be overcome partly with a safe Gauss elimination which honors a “Pauli exclusion principle”: we can do row reduction, as long as we do not deal with pairs of parallel vectors. This analogy leads to the use of alternating multi-linear forms to get a geometric characterization of the pseudo-determinant. Lets assume for simplicity that A is a square matrix. In this multi-linear setup, A acts on k -forms $f = f_1 \wedge f_2 \wedge \dots \wedge f_k$ as $Af = Af_1 \wedge Af_2 \wedge \dots \wedge Af_k$. The length of the vector $f = f_1 \wedge \dots \wedge f_k$ is defined as the k -volume $|f|$ of the parallelepiped spanned by the k vectors f_1, \dots, f_k . We have $|f| = \det(F)$, where F is the matrix $F_{ij} = f_i \cdot f_j$ and where $f \cdot g$ denotes the Euclidean dot product. Lets call a row in A independent within A , if it is not parallel to any other nonzero row. Safe row reduction consists of three steps $A \rightarrow B$ using the following rules: **a)** if an independent row of A is scaled by a factor $\lambda \neq 0$, then $\text{Det}(B) = \lambda \text{Det}(A)$, **b)** if two independent rows in A are swapped, then $\text{Det}(B) = -\text{Det}(A)$, **c)** if an independent row is subtracted from an other independent row, then $\text{Det}(B) = \text{Det}(A)$. We would like to apply this to the pseudo-determinant of $A = F^T G$: lets see what happens if we append a multiple of a given row, starting with the assumption that all of rows are already independent. While it is difficult to predict the effect of adding a multiple of a row to an other row, it is possible to see when what happens to $A = F^T G$ if such an operation is performed for F and for G .

i) Appending a parallel row.

Given two $n \times m$ matrices F, G such that $F^T G$ is nonsingular. Assume A^T is the $n \times (m + 1)$ matrix obtained from F^T by appending λ times

the l 'th row of F at the end. Assume that B^T is the $n \times (m+1)$ matrix obtained from G^T by appending μ times the l 'th row of G^T at the end.

Then both sides of the Cauchy-Binet formula are multiplied by $1 + \lambda\mu$.

Proof. First show that $\det(A^T B) = \text{Det}(A^T B) = (1 + \lambda\mu)\text{Det}(F^T G)$.

First bring the m row vectors in row reduced echelon form so that we end up both for F and G with matrices which are row reduced in the first $m - 1$ rows and for which the m 'th and $m + 1$ th row are parallel. If we now reduce the last two rows, $F^T G$ is block diagonal with

$\begin{bmatrix} 1 & \mu \\ \lambda & \lambda + \mu \end{bmatrix}$ at the end which has pseudo-determinant

$1 + \lambda\mu$. For every pattern P which does not involve the l 'th row,

we have $\det(F_P) = \det(A_P)$ and $\det(G_P) = \det(B_P)$. For every pattern

P which does involve the l ' row and not the second last we have $\det(A_P) = \lambda\det(F_P)$ and $\det(B_P) = \mu\det(G_P)$. For every pattern

which involves the appended last row as well as the l 'th row, we have $\det(A_P) = \det(B_P) = 0$.

ii) Given two $n \times m$ matrices F, G such that $F^T G$ and FG^T have maximal rank. Given $1 \leq l \leq m$. Assume the row $v = \sum_{j=1}^l \lambda_j v_j$ is appended to F^T and $w = \sum_{j=1}^l \mu_j w_j$ is appended to G^T , where v_j are the rows of F^T and w_j are the rows of G^T . Then both sides of the Cauchy-Binet formula are multiplied by $1 + \sum_{j=1}^l \lambda_j \mu_j$.

Proof. Use induction with respect to l , where $l = 1$ was case (i). When adding a new vector $l \rightarrow l + 1$, the determinant increases by $\lambda_{l+1} \mu_{l+1}$. On the right hand side this is clear by looking at the patterns which involve the last $(m + 1)$ 'th row.

When adding more rows we do not have explicit formulas any more.

1) This example hopes to illustrate, how difficult it can be to predict what scaling of a row does to the pseudo-determinant. Lets take the

rank-2 matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \\ 2 & 3 & 4 \end{bmatrix}$ and scale successively the first, second

or third row by a factor $\lambda = 3$, to get

$$B = \begin{bmatrix} 3 & 2 & 3 \\ 3 & 1 & 1 \\ 6 & 3 & 4 \end{bmatrix}, C = \begin{bmatrix} 1 & 6 & 3 \\ 1 & 3 & 1 \\ 2 & 9 & 4 \end{bmatrix}, D = \begin{bmatrix} 1 & 2 & 9 \\ 1 & 1 & 3 \\ 2 & 3 & 12 \end{bmatrix}.$$

While $\det(A) = -2$ we have $\det(B) = -8$, $\det(C) = -2$ and $\det(D) = -4$. A look at the eigenvalues $\sigma(A) = \{6.31662, -0.316625, 0\}$, $\sigma(B) = \{8.89898, -0.898979, 0\}$, $\sigma(C) = \{8.24264, -0.242641, 0\}$ and $\sigma(D) =$

$\{14.2801, -0.28011, 0\}$ confirms how scaling of rows with the same factor can be tossed around the spectrum all over the place. In this 3×3 example, we can visualize the situation geometrically. The pseudo-determinant can be interpreted as the area of parallelogram in the image plane. Scaling rows deforms the triangle in different directions and it depends on the location of the triangle, how the area is scaled.

2) The two 3×2 matrices $F = \begin{bmatrix} 1 & 1 \\ 0 & 3 \\ 2 & 1 \end{bmatrix}$, $G = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ both have rank

2. If we append 3 times the first row to F^T and 2 times the first row to G^T , then the pseudo-determinant changes by a factor $1 + 2 \cdot 3 = 7$. Indeed, we have $\det(F^T F) = -9$ and $\det(A^T B) = -63$ with $A = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 3 & 0 \\ 2 & 1 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 2 \\ 1 & 2 & 4 \end{bmatrix}$.

If we append 3 times the first row of F^T to F^T and 2 times the **second** row of G^T to G^T however, then the pseudo-determinant does not change because $1 + \sum_i \mu_i \lambda_i = 1$. Indeed we compute $\det(A^T B) = -9$ with

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 3 & 0 \\ 2 & 1 & 6 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

3) This is an example of 4×2 matrices F, G , where the first matrix

has rank 2 and the second matrix has rank 1. If $F = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}$, $G =$

$\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \\ 4 & 4 \end{bmatrix}$, Then $F^T G = \begin{bmatrix} 30 & 30 \\ 10 & 10 \end{bmatrix}$ which has pseudo-determinant 40.

If we multiply the second row of F^T with a factor 2, we get $F_1 = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 3 & 2 \\ 4 & 2 \end{bmatrix}$, and $F_1^T G = \begin{bmatrix} 30 & 30 \\ 20 & 20 \end{bmatrix}$ which has pseudo-determinant 50. If

we swap two rows of F^T , we get $F_2 = \begin{bmatrix} 2 & 1 \\ 2 & 2 \\ 2 & 3 \\ 2 & 4 \end{bmatrix}$ and $F_2^T G = \begin{bmatrix} 10 & 10 \\ 30 & 30 \end{bmatrix}$

which has the same pseudo-determinant 40. If we subtract the second

row of F^T from the first row of F^T , we get $F_3 = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix}$ and $F_3^T G = \begin{bmatrix} 20 & 20 \\ 10 & 10 \end{bmatrix}$ which has pseudo-determinant 30.

Finally, let's append twice the first row of F^T to F^T and three times the first row of G^T to G^T to get $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 4 \\ 3 & 1 & 6 \\ 4 & 1 & 8 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 2 & 6 \\ 3 & 3 & 9 \\ 4 & 4 & 12 \end{bmatrix}$,

so that $A^T B = \begin{bmatrix} 30 & 30 & 90 \\ 10 & 10 & 30 \\ 60 & 60 & 180 \end{bmatrix}$ which has rank 1 and $\text{Det}(A^T B) =$

220. For random 4×2 matrices F, G , scaling a row of F^T by λ changes the determinant of $F^T G$ by λ , swapping two rows of F^T swaps the determinant of $F^T G$ and subtracting a row of F^T from an other row does not change the determinant. After appending λ times a rows to F^T and μ times a row to G^T , we would get the pseudo-determinant scaled by a factor $1 + \lambda\mu = 7$.

8. REMARKS

A) Pseudo-determinants are useful also in infinite dimensions. In infinite dimensions, one has regularized Fredholm determinants $\det(1 + A) = \sum_{k=0}^{\infty} \text{tr}(\Lambda^k A)$ for trace class operators A . If A and B are trace class, then they satisfy the Cauchy-Binet formula $\det((1 + A)(1 + B)) = \det(1 + A)\det(1 + B)$ and Fredholm determinants are continuous on trace class operators. (See i.e. [52, 53]). Pseudo-determinants make an important appearance for zeta regularized Ray-Singer determinants [35] of Laplacians for compact manifolds using the Minakshisundaraman-Pleijel zeta function defined by the Laplacian. These determinants are always pseudo-determinants because Laplacians on compact Riemannian manifolds always have a kernel. For the Dirac operator $A = i \frac{d}{dx}$ on the circle, which has eigenvalues n to the eigenvalues e^{-in} the Dirac zeta function is $\zeta(s) = \sum_{n>0} n^{-s}$. Since $\zeta'(0) = -1$, the circle has the Dirac Ray-Singer determinant $\text{Det}(D) = e$. This must be seen as a regularized pseudo-determinant because only the nonzero eigenvalues have made it into the zeta function.

B) Theorem 1) will allow to give a combinatorial description of the pseudo-determinant of the Laplace-Beltrami operator $L = (d + d^*)^2$

acting on discrete differential forms of a finite simple graph. The right hand side of Cauchy-Binet for pseudo determinants can be expressed in terms of trees. If $D = d + d^*$ is the Dirac matrix of a finite simple graph, then $L = D^2$ is the Laplace-Beltrami operator L acting on discrete differential forms. It is a $v \times v$ matrix if v is the total number of simplices in the graph. The formula implies there that $\text{Det}(L)$ is the number of maximal complete trees in a double cover of G branched at a single vertex. Like the classical matrix tree theorem which gives an interpretation of the $\text{Det}(L_0)$ where L_0 is the matrix Laplacian on function, the Dirac matrix theorem gives an interpretation of $\text{Det}(L)$, where L is the Laplace-Beltrami operator acting on discrete differential forms. Things are a bit more interesting in the Dirac case because trees come with a positive or negative sign and unlike in the classical matrix tree theorem, $\det(D_P^2)$ does no more correspond to a single tree only. One has to look at a branched double cover of the graph so that we have to interpret $\det(D_P)\det(D_P)$ a signed number of maximal labeled trees in a double cover of the simplex graph \mathcal{G} branched at the base point. Now, $\text{Det}(D)$ is divisible by the number of vertices $|V|$ in the original graph G , if G is connected. Note that for the scalar Laplacian \mathcal{L}_0 of the simplex graph \mathcal{G} , the classical matrix tree just counts the number of trees in \mathcal{G} but this is a different matrix than the form Laplacian $(d + d^*)^2$, even so they are both square $v \times v$ matrices. As \mathcal{G} is a connected graph with v vertices, the kernel of \mathcal{L}_0 is one dimensional and $\text{Det}(\mathcal{L}_0)$ is divisible by v . The kernel of L however has a dimension $\sum_i b_i$ where $b_k = \dim(H^k(G))$ of G . In this linear algebra context it is worth mentioning that the Euler-Poincaré identity $\sum_k (-1)^k b_k = \sum_k (-1)^k v_k$ for the Euler characteristic of G or the Hodge formula $b_k = \dim(\ker(L_k))$ for the dimension of the k -harmonic forms are elementary (see [27, 29]).

C) For circle graphs C_n , the scalar Laplacian $L = L_0$ satisfies

$$\text{Det}(L) = \prod_{k=1}^{n-1} 4 \sin\left(\frac{k\pi}{n}\right)^2 = n^2 .$$

We have looked at Birkhoff sums $\sum_{k=1}^{n-1} \log(\sin^2(\pi k\alpha/n))$ for Diophantine α in [31, 26] and studied the roots of the zeta functions of these graphs which are entire functions

$$\zeta_n(s) = \sum_{k=1}^{n-1} 2^{-s} \sin^{-s}\left(\pi \frac{k}{n}\right)$$

in [30].

D) Let A be the adjacency matrix of a finite simple weighted graph. The later means that real values $A_{ij} = A_{ji}$ called weights are assigned to the edges of the graph. A sub-matrix $P = P_{K,L}$ is obtained by restricting to a sub pattern. If the square of $\det(P)^2$ is called the benefit of the sub-pattern and $\text{Det}^2(A)$ the benefit of the matrix, then the Pythagorean pseudo-determinant formula (6) tells that the square $\text{Det}^2(A)$ is the sum of the benefits of all $k \times k$ sub patterns, where k is the rank of A . The formula $\text{Det}(1 + A^2)$ is the sum of the benefits over all possible sub patterns. This picture confirms that both Det or \det are interesting functionals.

E) Lets look at a probability space of symmetric $n \times n$ matrices which take values in a finite set. We can ask which matrices maximize or minimize the pseudo-determinant. Pseudo-determinants can be larger than expected: for all 2×2 matrices taking values in $0, 1$, the maximal determinant is 1 while the maximal pseudo-determinant is 2, obtained for the matrix where all entries are 1. On a probability space (Ω, P) of matrices, where matrix entries have continuous distribution, it does of course not matter whether we take the determinant functional or pseudo-determinant functional because non-invertible matrices have zero probability. But we can ask for which $n \times n$ matrices taking values in a finite set, the pseudo-determinant $\text{Det}(A)$ is maximal.

F) We can look at the statistics of the pseudo-determinant on the Erdős-Renyi probability space of all finite simple graphs $G = (V, E)$ of order $|V| = n$ similar than for the Euler characteristic or the dimension of graph in [25]. Considering the pseudo-determinant functional on the subset of all connected graphs could be interesting. While the minimal pseudo-determinant is achieved for the complete graphs where $\text{Det}(D(K_n)) = -n^{2^{n-1}-1}$, it grows linearly for linear graphs $\text{Det}(D(L_n)) = n(-1)^{n-1}$ and quadratically for cycle graphs $\text{Det}(D(C_n)) = n^2(-1)^{n-1}$, where $D = D(C_n)$ is a $2n \times 2n$ matrix and D^2 is a block matrix containing two copies L_0, L_1 of the $n \times n$ scalar Laplacian L_0 and $\text{Det}(D^2) = \text{Det}(L_0)\text{Det}(L_1) = \text{Det}^2(L_0) = n^4$. The complete graph with one added spike seems to lead to the largest pseudo determinant. We computed the pseudo-determinant for all graphs up to order $n = 7$, where there are 1'866'256 connected graphs. It suggests that a limiting distribution of the random variable $X(G) = \log(|\text{Det}(L(G))|)$ might exist on the probability space of all connected graphs G in the limit $n \rightarrow \infty$.

G) The results were stated over fields of characteristic zero. Since multi-linear algebra can be done over any field K , Theorem 7) generalizes. Determinants over a commutative ring K can be characterized as an alternating n -linear function D on $M(n, K)$ satisfying $D(1) = 1$ (see e.g. [23, 18]). As discussed, this does not apply to pseudo-determinants. Is there an elegant axiomatic description of pseudo-determinants? I asked this Fuzhen Zhang after his plenary talk in Providence who informed me that it is not a **generalized matrix function** in the sense of Marcus and Minc [38], who introduced functions of the type $d(A) = \sum_{x \in H} \chi(x) \prod A_{i,x(i)}$, where H is a subgroup of the symmetric group and χ is a character on H . Indeed, the latter is continuous in A , while the pseudo determinant is not continuous as a function on matrices.

APPENDIX

We add some Mathematica routines which illustrate the results in this paper. First, we list an implementation of the pseudo determinant, using the first non zero entry p_k of the Characteristic polynomial multiplied by $(-1)^k$

```
FirstNon0[s_]:= -(-1)^ArrayRules[s][[1,1,1]] ArrayRules[s][[1,2]];
PDet[A_]:= FirstNon0[CoefficientList[CharacteristicPolynomial[A,x],x]];
```

Next we look at the procedure “PTrace” - the P stands for Pauli trace - which produces the pairing $\langle F, G \rangle_k$, which is the sum over all products $\det(F_P)\det(G_P)$ where P runs over $k \times k$ sub matrices. We compute the list of Pauli traces and a generating function $t(x)$ which will match the characteristic polynomial $p(x)$ of $F^T G$. There is a perfect match as our theorem tells that $p(x) = t(x)$.

```
PTrace[A_, B_, k_] := Module[{U, V, m=Length[A], n=Length[A[[1]]], s={}, t={},
  U=Partition[Flatten[Subsets[Range[m], {k,k}]], k]; u=Length[U];
  V=Partition[Flatten[Subsets[Range[n], {k,k}]], k]; v=Length[V];
  Do[s=Append[s, Table[A[[U[[i,u]], V[[j,v]]]], {u,k}, {v,k}], {i,u}, {j,v}];
  Do[t=Append[t, Table[B[[U[[i,u]], V[[j,v]]]], {u,k}, {v,k}], {i,u}, {j,v}];
  Sum[Det[s[[1]]] * Det[t[[1]]], {1, Length[s]}];
PTraces[A_, B_] := (-x)^m + Sum[(-x)^(m-k) PTrace[A, B, k], {k, 1, Min[n, m]}];
n=5; m=7; r=4; RandomMatrix[n_, m_] := Table[RandomInteger[2r-r, {n}, {m}];
F=RandomMatrix[n, m]; G=RandomMatrix[n, m];
CharacteristicPolynomial[Transpose[F].G, x]
PTraces[F, G]
```

The code for experiments made to discover Theorem 7) generates matrices F, G and then $A = F^T G$ and $B = F G^T$ for which both have zero determinant. This assures that we are in a situation, which cannot be explained by the classical Cauchy-Binet theorem. Then we compute both polynomials:

```

R:=RandomMatrix[n,m];
Shuffle:=Module[{},F=R;G=R;A=Transpose[F].G;B=F.Transpose[G];
Shuffle; While[Abs[Det[A]]+Abs[Det[B]]>0,Shuffle];
{CharacteristicPolynomial[Transpose[F].G,x],PTraces[F,G]}

```

Here we compare the coefficients stated first in Theorem 7):

```

F=RandomMatrix[n,m]; G=RandomMatrix[n,m];
PTr[A_.,B_.]:=1+Sum[z^k PTrace[A,B,k],{k,1,Min[n,m]}];
{Det[IdentityMatrix[m]+z Transpose[F].G],PTr[F,G]}

```

And now we give examples of the identity in Corollary 8):

```

A=RandomMatrix[n,m]; B=RandomMatrix[n,m];
{Det[IdentityMatrix[m]+Transpose[A].B],PTraces[A,B] /. x->-1}

```

Finally, lets illustrate the multi-linear Lemma 13) by building two matrices F, G and then comparing the matrix product of the exterior powers $\Lambda^k F, \Lambda^k G$ with the exterior power $\Lambda^k(FG)$ of the product FG :

```

WedgeP[F_.,k_.]:=Module[{U,V,m=Length[F],n=Length[F[[1]]],
U=Partition[Flatten[Subsets[Range[m],{k,k}],k],u=Length[U];
V=Partition[Flatten[Subsets[Range[n],{k,k}],k],v=Length[V];
Table[Det[Table[F[[U[[i,u]],V[[j,v]]]],{u,k},{v,k}],{i,u},{j,v}]];
F=RandomMatrix[n,m]; G=RandomMatrix[n,m];
Transpose[WedgeP[F,3].WedgeP[G,3]-WedgeP[Transpose[F].G,3]

```

REFERENCES

- [1] R. Abraham, J.E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis and Applications*. Applied Mathematical Sciences, 75. Springer Verlag, New York etc., second edition, 1988.
- [2] M. Aigner. *Combinatorial Theory*. Springer Verlag, 1979.
- [3] M. Aigner and G.M. Ziegler. *Proofs from the book*. Springer Verlag, Berlin, 2 edition, 2010. Chapter 29.
- [4] W. Willems B. Huppert. *Lineare Algebra*. Vieweg,Teubner, 2 edition, 2010.
- [5] R.B. Bapat. Outer inverses: Jacobi type identities and nullities of submatrices. *Linear Algebra and Its Applications*, 361:107–120, 2003.
- [6] J. Biao and A. Ben-Israel. Minors of the Moore-Penrose inverse. *Linear Algebra and Its Applications*, 195:191–207, 1993.
- [7] N. Biggs. *Algebraic Graph Theory*. Cambridge University Press, 1974.
- [8] J.P.M. Binet. Mémoire sur un système de formules analytiques, et leur application à des considerations géométriques. *J. de l'Ecole Polytechnique IX*, Cahier 16:280–302, page 287, 1813.
- [9] N. Bourbaki. *Elements d'histoire des Mathématiques*. Springer, 1984.
- [10] O. Bretscher. *Linear algebra with applications*. Prentice Hall, 2009.
- [11] B. Brooks. The coefficients of the characteristic polynomial in terms of the eigenvalues and the elements of an $n \times n$ matrix. *Appl. Math. Lett.*, 19(6):511–515, 2006.

- [12] R.A. Brualdi and H. Schneider. Determinantal identities: Gauss, Schur, Cauchy, Sylvester, Kronecker, Jacobi, Binet, Laplace, Muir, and Cayley. *Linear Algebra Appl.*, 52/53:769–791, 1983.
- [13] K.A. Bush. Classroom Notes: An Interesting Result on Almost Square Matrices and the Cauchy-Binet Theorem. *Amer. Math. Monthly*, 69(7):648–649, 1962.
- [14] A. Cauchy. Memoire sur le nombre de valeurs qu'une fonction peut obtenir. *J. de l'Ecole Polytechnique X*, pages 51–112, 1815.
- [15] A. Cayley. On the theory of determinants. *Transactions of the Cambridge Philosophical Society*, VIII:1–16, 1843.
- [16] J.L. Coolidge. *A history of geometrical methods*. Oxford University Press, and Dover, 1940 and 2003.
- [17] P. A. Deift. Applications of a commutation formula. *Duke Math. J.*, 45(2):267–310, 1978.
- [18] J. Dieudonné. *Grundzüge der modernen Analysis*. Vieweg, dritte edition, 1985.
- [19] I. Gohberg and S. Goldberg. *Basic Operator Theory*. Birkhäuser, 1980.
- [20] W.H. Greub. *Multilinear Algebra*. Springer, 1967.
- [21] P. Lancaster I. Gohberg and L. Rodman. *Invariant Subspaces of Matrices with Applications*, volume 51 of *Classics in Applied Mathematics*. SIAM, 2006.
- [22] I.R. Shafarevich and A.O. Remizov. *Linear Algebra and Geometry*. Springer, 2009.
- [23] R. Kunze K. Hoffman. *Linear Algebra*. Prentice Hall, second edition, 1971.
- [24] M. Kline. *Mathematical Thought from Ancient to Modern Time (Volume 2)*. Oxford University Press, 1972.
- [25] O. Knill. The dimension and Euler characteristic of random graphs. <http://arxiv.org/abs/1112.5749>, 2011.
- [26] O. Knill. Selfsimilarity in the birkhoff sum of the cotangent function. <http://arxiv.org/abs/1206.5458>, 2012.
- [27] O. Knill. The McKean-Singer Formula in Graph Theory. <http://arxiv.org/abs/1301.1408>, 2012.
- [28] O. Knill. Counting rooted forests in a network. <http://arxiv.org/abs/1307.3810>, 2013.
- [29] O. Knill. The dirac operator of a graph. <http://arxiv.org/abs/1306.2166>, 2013.
- [30] O. Knill. The zeta function for circular graphs. <http://arxiv.org/abs/1312.4239>, December 2013.
- [31] O. Knill and F. Tangerman. Selfsimilarity and growth in Birkhoff sums for the golden rotation. *Nonlinearity*, 21, 2011.
- [32] E. Knobloch. *Der Beginn der Determinantentheorie*. Hildesheim, 1980.
- [33] T. Konstantopoulos. A multilinear algebra proof of the Cauchy-Binet formula and a multilinear version of Parseval's identity. <http://http://arxiv.org/abs/1305.0644>, 2013.
- [34] P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Computer Science and Applied Mathematics. Academic Press, 1985.
- [35] M.L. Lapidus and M. van Frankenhuysen. *Fractal Geometry, Complex Dimensions and Zeta Functions*. Springer, 2006.
- [36] D.C. Lay. *Linear Algebra and its applications*. Pearson, 2006.

- [37] M. Lewin. On the coefficients of the characteristic polynomial of a matrix. *Discrete Math.*, 125(1-3):255–262, 1994. 13th British Combinatorial Conference (Guildford, 1991).
- [38] M. Marcus and H. Minc. Generalized matrix functions. *Transactions of the AMS*, 116:316–329, 1965.
- [39] M. Marcus and H. Minc. *Introduction to Linear Algebra*. Dover Publications, Inc, New York, 1965.
- [40] G.A. Miller. On the History of Determinants. *Amer. Math. Monthly*, 37(5):216–219, 1930.
- [41] J. Morrow. Accelerated advanced calculus, online course notes http://www.math.washington.edu/~morrow/335_11/, 2009.
- [42] T. Muir. *A treatise on the theory of determinants*. MacMillan and Co, London, 1882.
- [43] T. Muir. Contributions to the history of determinants. *Nature*, 126, 1930.
- [44] K. Hunger Parshall. *James Joseph Sylvester: Jewish Mathematician in a Victorian World*. Johns Hopkins University Press, 2006.
- [45] E. Pascal. *Die Determinanten*. Teubner, Leipzig, 1900.
- [46] P.Chebotarev and E. Shamis. Matrix forest theorems. arXiv:0602575, 2006.
- [47] L. Pennisi. Coefficients of the Characteristic Polynomial. *Math. Mag.*, 60(1):31–33, 1987.
- [48] E.V. Shamis P.Yu, Chebotarev. A matrix forest theorem and the measurement of relations in small social groups. *Avtomat. i Telemekh.*, (9):125–137, 1997.
- [49] M. Reed and B. Simon. *Methods of modern mathematical physics , Volume IV*. Academic Press, Orlando, 1980.
- [50] M. Reiss. *Theorie der Determinanten*. Teubner, 1867.
- [51] A.D. Sokal S. Caracciolo and A. Sportiello. Noncommutative determinants, Cauchy-Binet formulae, and Capelli-type identities. I. Generalizations of the Capelli and Turnbull identities. *Electron. J. Combin.*, 16(1):Research Paper 103, 43, 2009.
- [52] B. Simon. Notes on infinite determinants of Hilbert space operators. *Adv. in Math.*, 24(3):244–273, 1977.
- [53] B. Simon. *Trace Ideals and Their Applications*. AMS, 2. edition, 2010.
- [54] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, forth edition, 2009.
- [55] J.J. Sylvester. Additions to the articles 'on a new class of theorems' and 'on pascal's theorem'. pages 363–370, 1850. The Collected Mathematical Papers of James Joseph Sylvester.
- [56] J.J. Sylvester. An essay on canonical forms, supplement to a sketch of a memoir on elimination, transformation and canonical forms. 1850. The Collected Mathematical Papers of James Joseph Sylvester.
- [57] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. AMS, 2012.