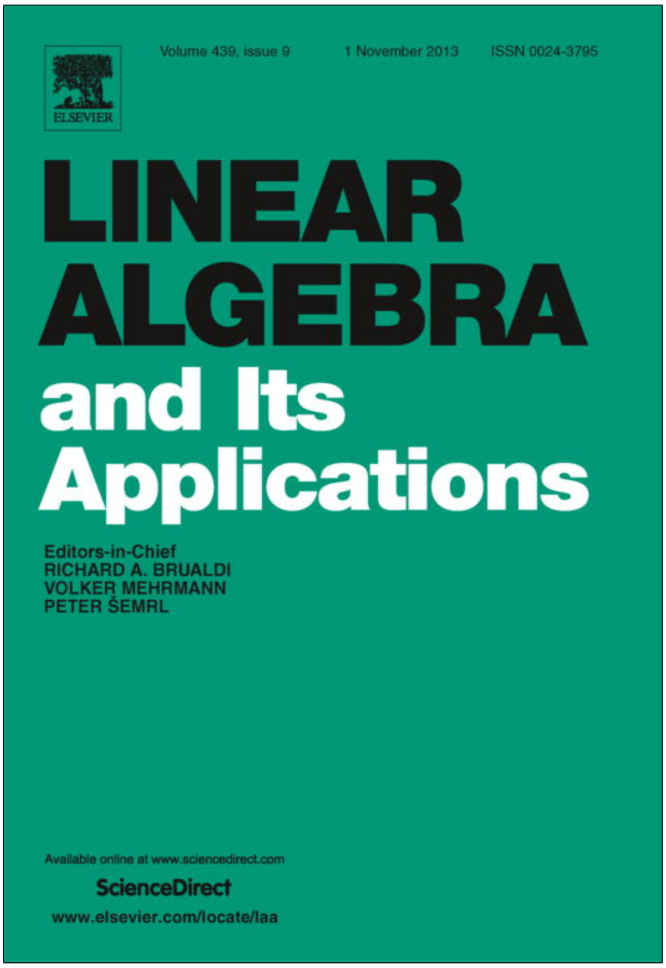

Volume 439, issue 9
1 November 2013
ISSN 0024-3795
ELSEVIER
LINEAR ALGEBRA and Its Applications
Editors-in-Chief
RICHARD A. BRUALDI
VOLKER MEHRMANN
PETER ŠEMRL
Available online at www.sciencedirect.com
ScienceDirect
www.elsevier.com/locate/laa

Linear Algebra and its Applications 439 (2013) 2651–2658

Contents lists available at SciVerse ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

# A multilinear algebra proof of the Cauchy–Binet formula and a multilinear version of Parseval's identity

Takis Konstantopoulos

*Department of Mathematics, Uppsala University, 751 06 Uppsala, Sweden*

ARTICLE INFO

*Article history:*
Received 3 May 2013
Accepted 5 July 2013
Available online 9 August 2013
Submitted by P. Butkovic

*MSC:*
15A15
15A69
15A24
46C05

*Keywords:*
Cauchy–Binet theorem
Determinant
Matrix identities
Hilbert space
Parseval's identity
Multilinear algebra
Exterior products
Projections
Pythagorean theorem
Fock space

ABSTRACT

We prove the Cauchy–Binet determinantal formula using multilinear algebra by first generalizing it to an identity not involving determinants. By extending the formula to abstract Hilbert spaces we obtain, as a corollary, a generalization of the classical Parseval identity.



*E-mail address:* takis@math.uu.se.
*URL:* http://www.math.uu.se/~takis.

0024-3795/$ – see front matter 
http://dx.doi.org/10.1016/j.laa.2013.07.009

## 1. Introduction and overview

The classical Cauchy–Binet formula states that if $A, B$ are two matrices over $\mathbb{R}$ (or any field) of sizes $n \times N$, $N \times n$, respectively, with $n \leqslant N$, then

$$\det(AB) = \sum_{\sigma} \det(A_\sigma) \det\big(B^\sigma\big) \tag{1}$$

where the sum is taken over all $\sigma = (\sigma_1 < \sigma_2 < \cdots < \sigma_n)$, with $\sigma_i \in \{1, \ldots, N\}$, and where $A_\sigma$ (respectively $B^\sigma$) is the $n \times n$ submatrix of $A$ (respectively submatrix of $B$) obtained by deleting all columns (respectively all rows) except those with indices in $\sigma$.

There are many proofs of this formula, each telling its own story, explaining the formula from a different point of view. The most direct way of proving the formula is by writing down the determinant as a sum over permutations and performing algebraic manipulations. This is the approach taken in many linear algebra books; see Marcus and Minc [7, Theorem 6.1, p. 128] and Gohberg et al. [6, Theorem A.2.1, p. 651]. A probabilistic interpretation and proof of the formula (which starts by using the formula for a determinant) is also available [4]. On the other hand, there are many combinatorial proofs. Suffice, perhaps, to refer to the one chosen to be included in the "Proofs from The Book" [2] by Aigner and Ziegler. This is a nice proof (after all, it is a proof from The Book) based on the beautiful Gessel–Vienot lemma which states that, in a finite weighted acyclic directed graph, the determinant of the path matrix between two sets of vertices of cardinality $n$ each equals a sum over all possible vertex-disjoint path systems; see [2, Chapter 29, p. 196] and [1] for details. Another very simple proof appears in the recent book by Terence Tao [11, p. 298] on random matrices. An outline of proof is as follows: Start from Sylvester's determinant identity,

$$\det(I_n + AB) = \det(I_N + BA), \tag{2}$$

where $I_n, I_N$ are identity matrices. This can be proved in a number of ways, one of which is sketched in [11, p. 298]. Then

$$z^{N-n} \det(zI_n + AB) = \det(zI_N + BA).$$

When $N \geqslant n$, an inspection of the $z^{N-n}$ coefficient gives the Cauchy–Binet formula (1).

On the other hand, the Cauchy–Binet formula is a generalization of the Pythagorean theorem. Indeed, let $A$ be an $n \times N$ real matrix, $n \leqslant N$, and take $B = A^T$, the transpose of $A$. Since $B^\sigma = (A^T)^\sigma = (A_\sigma)^T$, the formula gives

$$\det\big(AA^T\big) = \sum_{\sigma} \det(A_\sigma)^2,$$

which can be interpreted geometrically as follows: The parallelotope in $\mathbb{R}^N$ generated by the $n$ row vectors of $A$ has $n$-dimensional Lebesgue measure $\sqrt{\det(AA^T)}$. Therefore the formula says that the square of the $n$-dimensional measure of an $n$-dimensional parallelotope, embedded in a higher-dimensional Euclidean space, equals the sum of the squares of the measures of its projections onto all possible $n$-dimensional coordinate hyperplanes. If $n = 1$ this reduces to the Pythagorean theorem.

The goal of this short article is to give a proof of the Cauchy–Binet formula, from an algebraic-geometric viewpoint. If $n = 1$, the Cauchy–Binet formula is a triviality: it states that the inner product of two $N$-dimensional vectors equals the sum of the products of their components:

$$(a_1, \ldots, a_N) \cdot (b_1, \ldots, b_N)^T = \sum_{\sigma=1}^{N} a_\sigma b_\sigma.$$

There is no need to take determinants here, because both sides involve $1 \times 1$ matrices, i.e., real numbers. What we show is that the general case, when $n \geqslant 1$, is the same, but on bigger vector spaces. In Section 2 we give an account of the ingredients we need, and, in Section 3, we state and prove the main formula (Theorem 1) without determinants and in a more general setup; a corollary of it is the classical Cauchy–Binet formula. Then, in Section 4, we see that the formula can be extended to a Hilbert space, giving a generalization of the classical Parseval identity. We conclude with a few bibliographic remarks.

## 2. The main ingredients

The main theorem, Theorem 1, requires two ingredients.

(i) The first is the notion of the determinant of a linear transformation $F: X \to X$ on a vector space $X$ of dimension $d$. The dimension of the linear space $\bigwedge^m X$ of alternating $m$-linear maps $\omega: X^m \to \mathbb{R}$ is $\binom{d}{m}$. For each $m$, the $m$-th level dual $F^*: \bigwedge^m X \to \bigwedge^m X$ of $F$ is defined by

$$F^*\omega[x_1, \dots, x_m] := \omega[Fx_1, \dots, Fx_m]; \tag{3}$$

see [10]. (Duals obey the standard composition rules: $(GF)^* = F^*G^*$.) Since $\bigwedge^d X$ is 1-dimensional, the $d$-th level dual $F^*$ is multiplication by a constant. This constant is, *by definition*, the determinant of $F$:

$$F^*\omega = (\det F) \cdot \omega, \quad \omega \in \bigwedge^d X. \tag{4}$$

(ii) The second ingredient is very simple too. Let $X, Y, Z$ be vector spaces, and $F: X \to Y$, $G: Y \to Z$ linear maps. Suppose $Y$ is the direct sum of $Y_1, \dots, Y_K$. Let $P_i: Y \to Y_i$, $1 \leqslant i \leqslant K$, be the projections corresponding to this direct sum (so $\mathrm{id}_V = P_1 + \dots + P_K$ is a partition of the identity on $V$), and let $E_i: Y_i \to Y$ be the natural embedding of $Y_i$ into $Y$. Then, clearly,

$$GF = \sum_{i=1}^{K} (GE_i)(P_iF). \tag{5}$$

See Diagram 1.

## 3. An abstract version of the Cauchy–Binet formula

Let $U, V, W$ be finite-dimensional vector spaces of arbitrary dimensions, and let $B: U \to V$, $A: V \to W$ be two linear maps. Fix $n \in \mathbb{N}$ and consider the $n$-th level duals $B^*: \bigwedge^n V \to \bigwedge^n U$, $A^*: \bigwedge^n W \to \bigwedge^n V$. Let $N$ be the dimension of $V$ and let $f_1, \dots, f_N$ be a basis for $V$. See Diagram 2. Denote by $\mathcal{S}_n(N)$ the set of subsets of $\{1, \dots, N\}$ of size $n$. For each $\sigma \in \mathcal{S}_n(N)$, let $V_\sigma$ be the subspace of $V$ spanned by $\{f_i, i \in \sigma\}$ and consider the direct sum

$$V = V_\sigma \oplus V_{\overline{\sigma}}, \tag{6}$$

where $\overline{\sigma} := \{1, \dots, N\} \setminus \sigma$,

$$P_\sigma: V \to V_\sigma$$

is the projection of $V$ onto $V_\sigma$ along $V_{\overline{\sigma}}$, and

$$E_\sigma: V_\sigma \to V$$

is the natural embedding of $V_\sigma$ into $V$.

**Theorem 1.**

$$(AB)^* = \sum_{\sigma \in \mathcal{S}_n(N)} (P_\sigma B)^* (AE_\sigma)^*, \tag{7}$$

**Proof.** The $\binom{N}{n}$-dimensional space $\bigwedge^n V$ is the direct sum of the 1-dimensional spaces $\bigwedge^n V_\sigma$, where $\sigma$ ranges in $\mathcal{S}_n(N)$:

$$\bigwedge^n V = \bigoplus_{\sigma \in \mathcal{S}_n(N)} \bigwedge^n V_\sigma. \tag{8}$$

Let $\mathscr{P}_\sigma : \bigwedge^n V \to \bigwedge^n V_\sigma$ be projections corresponding to this direct sum, and let $\mathscr{E}_\sigma : \bigwedge^n V_\sigma \to \bigwedge^n V$ be natural embedding. Using (5) (with $A^*$, $B^*$ in place of $F$, $G$, respectively, and $K = \binom{N}{n}$), we have

$$B^* A^* = \sum_{\sigma \in \mathcal{S}_n(N)} \big(B^* \mathscr{E}_\sigma\big)\big(\mathscr{P}_\sigma A^*\big).$$

See Diagram 2. Since (see Lemma 1)

$$\mathscr{E}_\sigma = P_\sigma^*, \qquad \mathscr{P}_\sigma = E_\sigma^*,$$

the theorem follows from the composition rules of the duals. □

**Lemma 1.**

$$\mathscr{P}_\sigma = E_\sigma^*, \qquad \mathscr{E}_\sigma = P_\sigma^*.$$

**Proof.** We identify $\mathcal{S}_n(N)$ with the set of strictly increasing sequences of length $n$ with values in $\{1, \dots, N\}$. Thus, if $\sigma$ is a subset of $\{1, \dots, N\}$ we let $(\sigma_1, \dots, \sigma_n)$ be a listing of its elements in increasing order. To prove the first equality it suffices to show that

$$\mathscr{P}_\sigma \omega[v_1, \dots, v_n] = \omega[E_\sigma v_1, \dots, E_\sigma v_n],$$

for all $\omega \in \Lambda_n(V)$ and all $v_1, \dots, v_n \in V_\sigma$. But then $E_\sigma v_i = v_i$ and, since $V_\sigma$ is spanned by $f_{\sigma_1}, \dots, f_{\sigma_n}$, it suffices to show that

$$\mathscr{P}_\sigma \omega[f_{\sigma_{\pi(1)}}, \dots, f_{\sigma_{\pi(n)}}] = \omega[f_{\sigma_{\pi(1)}}, \dots, f_{\sigma_{\pi(n)}}], \tag{9}$$

where $\pi$ is a permutation of $\{1, \dots, n\}$. Since $\omega = \sum_{\tau \in \mathcal{S}_n(N)} \mathscr{P}_\tau \omega$ [this is the partition of the identity on $\bigwedge^n V$ corresponding to (8)] we may replace $\omega$ by $\mathscr{P}_\tau \omega$ in (9):

$$\mathscr{P}_\sigma \mathscr{P}_\tau \omega[f_{\sigma_{\pi(1)}}, \dots, f_{\sigma_{\pi(n)}}] = \mathscr{P}_\tau \omega[f_{\sigma_{\pi(1)}}, \dots, f_{\sigma_{\pi(n)}}].$$

But then, if $\tau = \sigma$ the two sides are equal, and if $\tau \neq \sigma$ the left-hand side equals zero and $\mathscr{P}_\tau \omega[f_{\sigma_{\pi(1)}}, \dots, f_{\sigma_{\pi(n)}}] = 0$.

To prove the second equality it suffices to show that

$$\mathscr{E}_\sigma \omega[v_1, \dots, v_n] = \omega[P_\sigma v_1, \dots, P_\sigma v_n],$$

for all $\omega \in \Lambda_n(V_\sigma)$ and all $v_1, \dots, v_n \in V$. But then $\mathscr{E}_\sigma \omega = \omega$. Since $v_i = P_\sigma v_i + P_{\overline{\sigma}} v_i$ [corresponding to (6)], we have

$$\mathscr{E}_\sigma \omega[v_1, \dots, v_n] = \omega[P_\sigma v_1 + P_{\overline{\sigma}} v_1, \dots, P_\sigma v_n + P_{\overline{\sigma}} v_n]. \tag{10}$$

Using the multilinearity of $\omega$ we split the right-hand side of (10) into $2^n$ terms, all of which are zero except the one involving only $P_\sigma v_i$ as arguments. □

Consider now the case where $W = U$. Moreover, take the number $n$ in Theorem 1 to be equal to their common dimension. Assume $n \leqslant N = \dim V$ to avoid trivialities. Then the linear maps $(AB)^*$, $(P_\sigma B)^*$, and $(AE_\sigma)^*$, appearing in formula (7), are maps between 1-dimensional spaces. Since the spaces $V_\sigma$ and $U$ have common dimension $n$, we can identify them by means of a linear bijection

$$\varphi_\sigma : V_\sigma \to U.$$

Then

$$(P_\sigma B)^*(AE_\sigma)^* = (P_\sigma B)^*\varphi_\sigma^*\big(\varphi_\sigma^{-1}\big)^*(AE_\sigma)^* = (\varphi_\sigma P_\sigma B)^*\big(AE_\sigma\varphi_\sigma^{-1}\big)^*,$$

and so

$$(AB)^* = \sum_{\sigma\in\mathcal{S}_n(N)} (\varphi_\sigma P_\sigma B)^*\big(AE_\sigma\varphi_\sigma^{-1}\big)^*. \tag{11}$$

Since all three linear maps $AB$, $\varphi_\sigma P_\sigma B$, $AE_\sigma\varphi_\sigma^{-1}$ are linear maps on the same 1-dimensional vector space $U$, it follows, from the definition of the determinant, that

$$\det(AB) = \sum_{\sigma\in\mathcal{S}_n(N)} \det(\varphi_\sigma P_\sigma B)\det\big(AE_\sigma\varphi_\sigma^{-1}\big). \tag{12}$$

(The role of $\varphi_\sigma$ is to force all maps be on the same space, so we can talk about determinants.) If $U=\mathbb{R}^n$, $V=\mathbb{R}^N$, this proves the classical Cauchy–Binet formula (1). If $N=n$, then we have shown that the determinant of the product is the product of the determinants.

Therefore (1) follows from (12). The latter is a restatement of (11). But (11) is a special case of (7) because in (7) we allow $U, V, W$ to be different with dimensions that may be distinct from $n$.

## 4. Multilinear Parseval's identity

We are now going to replace the middle space $V$ of the previous setup by a separable Hilbert space $H$ over the complex numbers $\mathbb{C}$, having inner product $\langle x, y\rangle$. Let $f_1, f_2, \ldots$ be an orthonormal basis for $H$. Let $\bigwedge^n H$ be the collection of all *continuous* alternating multilinear functionals $\omega : H^n \to \mathbb{C}$. In particular, $\bigwedge^1 H = H^*$ is the Hilbert space dual of $H$. By the Riesz–Fischer theorem, $f_1, f_2, \ldots$ forms a basis for $\bigwedge^1 H$ in the sense that every $\omega \in \bigwedge^1 H$ can be uniquely written as $\omega[x] = \sum_{\sigma=1}^\infty a_\sigma\langle f_\sigma, x\rangle$, for $a_\sigma \in \mathbb{C}$ such that $\sum_\sigma |a_\sigma|^2 < \infty$. More generally, $\bigwedge^n H$ is a separable Hilbert space with orthonormal (with respect to a suitably defined inner product) basis

$$f_{\sigma_1}\wedge\cdots\wedge f_{\sigma_n}, \quad \sigma = (\sigma_1,\ldots,\sigma_n)\in\mathcal{S}_n(\mathbb{N}),$$

where $\mathcal{S}_n(\mathbb{N})$ is the collection of all $n$-tuples $(\sigma_1,\ldots,\sigma_n)$ of positive integers such that $\sigma_1 < \cdots < \sigma_n$. Recall that the wedge product satisfies, by definition,

$$(f_1\wedge f_2)[x,y] = f_1[x]f_2[y] - f_1[y]f_2[x],$$

and, more generally, $f_{\sigma_1}\wedge\cdots\wedge f_{\sigma_n}$ is obtained by antisymmetrization of the tensor product of $f_{\sigma_1},\ldots,f_{\sigma_n}$. Incidentally, the direct sum of $\bigoplus_{n=0}^\infty \bigwedge^n H$ (where $\bigwedge^0 H := \mathbb{C}$) is the so-called alternating Fock (or fermionic) space [9]. Wedge products can be defined, by linearity, between any finite number of elements of this space.

If $H_1, H_2$ are two Hilbert spaces and $F : H_1 \to H_2$ is a continuous linear function then $F^* : \bigwedge^n H_2 \to \bigwedge^n H_1$ is defined as before–see (3)–and is, moreover, continuous.

**Theorem 2.** *Let $H$ be a separable Hilbert space over $\mathbb{C}$ with orthonormal basis $f_1, f_2, \ldots,$ and let $n$ be a positive integer. For each $\sigma \in \mathcal{S}_n(\mathbb{N})$, let $H_\sigma$ be the subspace spanned by $f_{\sigma_1},\ldots,f_{\sigma_n}$. Let $E_\sigma : H_\sigma \to H$ be the natural embedding of $H_\sigma$ into $H$ and let $P_\sigma : H \to H_\sigma$ be the orthogonal projection of $H$ onto $H_\sigma$. If $U$, $W$ are finite-dimensional vector spaces over $\mathbb{C}$ and $B : U \to H$, $A : H \to W$ continuous linear maps, then*

$$(AB)^* = \sum_{\sigma\in\mathcal{S}_n(\mathbb{N})} (P_\sigma B)^*(AE_\sigma)^*.$$

*If $W = U$ with common dimension $n$, and if $\varphi_\sigma : H_\sigma \to U$ is any linear bijection, then*

$$(AB)^* = \sum_{\sigma\in\mathcal{S}_n(\mathbb{N})} (\varphi_\sigma P_\sigma B)^*\big(AE_\sigma\varphi_\sigma^{-1}\big)^*.$$

*In particular,*

$$\det(AB) = \sum_{\sigma \in \mathcal{S}_n(\mathbb{N})} \det(\varphi_\sigma P_\sigma B) \det\big(A E_\sigma \varphi_\sigma^{-1}\big).$$

The proof of this theorem is exactly as in the finite-dimensional case. Infinite sums have to be understood in the Hilbert space sense.

Consider now $H = L^2[0,1]$ with inner product $\langle x, y\rangle = \int_0^1 x(t)\overline{y(t)}\,dt$ and the standard orthonormal basis $e_k(t) = \exp(i2\pi kt)$, $k \in \mathbb{Z}$, and let $U = W = \mathbb{C}^n$, for a given positive integer $n$. A continuous linear map $A : L^2[0,1] \to \mathbb{C}^n$ is necessarily (Riesz representation theorem) of the form

$$Ax = \big(\langle x, a_1\rangle, \dots, \langle x, a_n\rangle\big) = \left(\int_0^1 \overline{a_1(t)} x(t)\,dt, \dots, \int_0^1 \overline{a_n(t)} x(t)\,dt\right), \quad x \in L^2[0,1],$$

where $a_1, \dots, a_n \in L^2[0,1]$. A linear map $B : \mathbb{C}^n \to L^2[0,1]$ is of the form

$$(Bu)(t) = u_1 b_1(t) + \dots + u_n b_n(t), \quad u \in \mathbb{C}^n,$$

where $b_1, \dots, b_n \in L^2[0,1]$. Hence the $jk$-entry of the matrix of $AB : \mathbb{C}^n \to \mathbb{C}^n$, with respect to the standard basis on $\mathbb{C}^n$, is given by

$$(AB)_{jk} = \int_0^1 \overline{a_j(t)} b_k(t)\,dt.$$

Consider now $\sigma \in \mathcal{S}_n(\mathbb{Z})$, i.e., $\sigma = (\sigma_1, \dots, \sigma_n) \in \mathbb{Z}^n$ with $\sigma_1 < \dots < \sigma_n$. (There is no difficulty in replacing $\mathbb{N}$ in the preceding theorem by $\mathbb{Z}$.) Then $H_\sigma$ is the subspace of $L^2[0,1]$ spanned by $e_{\sigma_1}, \dots, e_{\sigma_n}$. So the orthogonal projection $P_\sigma : H \to H_\sigma$ is given by

$$P_\sigma x = \widehat{x}(\sigma_1) e_{\sigma_1} + \dots + \widehat{x}(\sigma_n) e_{\sigma_n},$$

where

$$\widehat{x}(k) := \int_0^1 x(t) \exp(-i2\pi kt)\,dt, \quad k \in \mathbb{Z},$$

are the Fourier coefficients of $x$. Letting $\varphi_\sigma : H_\sigma \to \mathbb{C}^n$ be the linear bijection that takes $e_{\sigma_r}$ into the $r$-th standard basis vector of $\mathbb{C}^n$, for $r = 1, \dots, n$, we see that the $jk$-entry of the matrix of $\varphi_\sigma P_\sigma B$ is

$$(\varphi_\sigma P_\sigma B)_{jk} = \widehat{b}_k(\sigma_j).$$

Arguing analogously, the $jk$-entry of the matrix of $A E_\sigma \varphi_\sigma^{-1}$ is

$$\big(A E_\sigma \varphi_\sigma^{-1}\big)_{jk} = \overline{\widehat{a}_j(\sigma_k)}.$$

Hence the last formula of Theorem 2 gives

$$\begin{aligned}\det_{1 \leqslant j,k \leqslant n} \int_0^1 \overline{a_j(t)} b_k(t)\,dt &= \sum_{\sigma \in \mathcal{S}_n(\mathbb{Z})} \det_{1 \leqslant j,k \leqslant n} \big[\overline{\widehat{a}_j(\sigma_k)}\big] \det_{1 \leqslant j,k \leqslant n} \big[\widehat{b}_j(\sigma_k)\big] \\ &= \frac{1}{n!} \sum_{\sigma_1 \in \mathbb{Z}} \cdots \sum_{\sigma_n \in \mathbb{Z}} \det_{1 \leqslant j,k \leqslant n} \big[\overline{\widehat{a}_j(\sigma_k)}\big] \det_{1 \leqslant j,k \leqslant n} \big[\widehat{b}_j(\sigma_k)\big],\end{aligned}$$

where the second equality follows from the fact that applying the permutation of $(\sigma_1, \ldots, \sigma_n)$ to both matrices changes the sign of both determinants simultaneously and the fact that repeated indices result into zero determinants. For $n = 1$, this is the standard Parseval identity.

Of course, there is nothing special about the Lebesgue measure. We can obtain formulas for any other $L^2$ space or other separable Hilbert spaces.

## 5. Remarks

My motivation for this article was due to my desire to understand some elements of random matrix theory [3] and determinantal point processes [5]. In particular, the derivation of the ubiquitous Tracy–Widom probability distribution [3] involves several applications of Cauchy–Binet type formulas. When I looked at it first, a standard computational proof was not too satisfactory. I discovered that there are many proofs, which can be roughly classified into combinatorial and algebraic ones. The version presented in this article was inspired by the observation that the Cauchy–Binet formula is a version of Pythagorean theorem: it is a version of the Pythagorean theorem on $\bigwedge^n \mathbb{R}^N$, with $n \leqslant N$ (which is of course isomorphic to $\mathbb{R}^{\binom{N}{n}}$).

Several years ago, Zeilberger [12] "complained" that, to most contemporary mathematicians, matrices and linear transformations are practically interchangeable notions and that the mainstream 'Bourbakian' establishment, with its profound disdain for the concrete, goes as far as to frown at the mere mention of the word 'matrix'. He then explains how "to [him], as well as to other 'dissidents' called 'combinatorialists', a matrix has nothing whatsoever to do with that intimidating abstract concept called 'a linear transformation between linear vector spaces'" and, by thinking of matrices as putting weights on a graph, he develops a combinatorial way of interpreting and proving fundamental results such as the Cayley–Hamilton theorem. The Cauchy–Binet formula has found a nice proof, in the Zeilberger sense, as a corollary of the Gessel–Vienot lemma. We also mention Zeng's proof [13] which also uses Zeilberger's methods.

In a sense then, what we have done here is in exactly the opposite of Zeilberger's spirit, because the proof presented uses nothing else but the concept of a linear map between vector spaces (and lots of definitions). Each point of view leads to different kinds of extensions. (Extensions to infinite matrices are not easy when the combinatorial point of view is adopted.)

There are generalizations of the Cauchy–Binet formula to matrices that contain elements of a non-commutative ring [8]. We do not know how to extend our proof to this case.

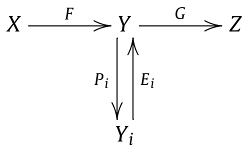
## 6. Note added in proof

Craig Tracy has recently pointed out to me that a proof of the Cauchy–Binet formula can also be found in [C.A. Tracy, H. Widom, On the distributions of the lengths of the longest monotone subsequences in random words, Probab. Theory Related Fields 119 (2001) 350–380]. The proof is essentially the same as the one proof appearing in [11].

## Acknowledgements

I thank Svante Janson for pointing out Ref. [8] to me and for his comments on this article; Richard Ehrenborg for kindly making his notes available to me; Terence Tao for informing me that his proof of the Cauchy–Binet theorem was inspired by Percy Deift's comment on the central importance of Sylvester's determinant theorem (2) in mathematics; and an anonymous referee for his/her careful reading of the manuscript as well as for informing me that he/she learned the Cauchy–Binet proof based on Sylvester's identity from Alan Hoffman.

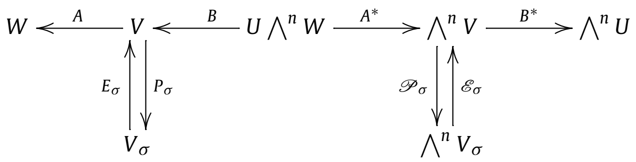
## Appendix A. Diagrams




$$Y = \bigoplus_{i=1}^{K} Y_i$$

$$GF = \sum_{i=1}^{K} (GE_i)(P_i F)$$

**Diagram 1.**




$$\bigwedge^n V = \bigoplus_{\sigma \in \mathcal{S}_n(N)} \bigwedge^n V_\sigma$$

$$B^* A^* = \sum_{\sigma \in \mathcal{S}_n(N)} (B^* \mathscr{E}_\sigma)(\mathscr{P}_\sigma A^*)$$

**Diagram 2.**

## References


[1] Martin Aigner, Lattice paths and determinants, in: Comput. Discrete Mathematics, vols. 1–12, in: Lect. Notes Comput. Sci., vol. 2122, Springer, Berlin, 2001.
[2] Martin Aigner, Günter Ziegler, Proofs from the Book, 4th ed., Springer, New York, 2009.
[3] Greg Anderson, Alice Guionnet, Ofer Zeitouni, An Introduction to Random Matrices, Cambridge Univ. Press, 2009.
[4] Richard Ehrenborg, Gian-Carlo Rota, Topics on Polynomials, Lect. Notes, MIT, 1991.
[5] John Ben Hough, Manjunath Krishnapur, Yuval Peres, Balint Virag, Zeros of Gaussian Analytic Functions and Determinantal Point Processes, Univ. Lecture Ser., vol. 51, Amer. Math. Society, 2009.
[6] Israel Gohberg, Peter Lancaster, Leiba Rodman, Invariant Subspaces of Matrices with Applications, SIAM Classics in Applied Mathematics, vol. 51, SIAM, 1986.
[7] Marvin Marcus, Henryk Minc, Introduction to Linear Algebra, MacMillan, New York, 1965, reprinted by Dover, New York, 1965.
[8] Sergio Caracciolo, Alan Sokal, Andrea Sportiello, Noncommutative determinants, Cauchy–Binet formulae, and Capelli-type identities. I. Generalizations of the Capelli and Turnbull identities, Electron. J. Comb. 16 (1) (2009), 43 pp., research paper 103.
[9] Michael Reed, Barry Simon, Methods of Modern Mathematical Physics, vol. II, Academic Press, 1975.
[10] Michael Spivak, Calculus on Manifolds, Addison-Wesley, 1965.
[11] Terence Tao, Topics in random matrix theory, http://terrytao.wordpress.com/books/topics-in-random-matrix-theory/.
[12] Doron Zeilberger, A combinatorial approach to matrix algebra, Discrete Math. 56 (1985) 61–72.
[13] Jiang Zeng, A bijective proof of Muir's identity and the Cauchy–Binet formula, Linear Algebra Appl. 184 (1993) 79–82.