INVARIANT THEORY OF MATRICES

Helmer Aslaksen[1], Eng-Chye Tan and Chen-bo Zhu

Department of Mathematics
National University of Singapore
Singapore 0511
Republic of Singapore
e-mail: mathelmr(mattanec, matzhucb)@nus.sg

INTRODUCTION

Let F be a field of characteristic 0, let $M(n,m) = M(n,m,F)$ denote the set of $n \times m$ matrices over F and let $W = W(n,m,F)$ be the vector space of m-tuples of $n \times n$ matrices over F. Let $V \subset W$ be a vector space on which a group $G \subset GL(n,F)$ acts by simultaneous conjugation. We will denote the polynomial functions on V by $P(V)$ and the G invariants by $P(V)^G$.

We will first consider affine invariants, i.e., invariants under the general linear group. Expressions of the form $\operatorname{tr} A^2B^2AB$ are invariant, since

$$\operatorname{tr}(gAg^{-1})^2(gBg^{-1})^2gAg^{-1}gBg^{-1} = \operatorname{tr} gA^2B^2ABg^{-1} = \operatorname{tr} A^2B^2AB.$$

The following theorem is due to Gurevich [6], Sibirskii [14] and Procesi [12].

Theorem 1 *Let $G = GL(n,F)$ act on $W(n,m,F)$ by conjugation. The invariants are of the form $\operatorname{tr} P(A_1,\dots,A_m)$, where P is a noncommutative polynomial in m matrix variables.*

We will next consider invariants under the orthogonal group. The conjugation action is now the same as the congruence action ($gMg^{-1} = gMg^t$), and we see that in additions to the affine invariants, expressions of the form $\operatorname{tr} A^tB$ are now also invariants. Namely,

$$\operatorname{tr}(gAg^t)^t gBg^t = \operatorname{tr} gA^tg^tgBg^t = \operatorname{tr} A^tB.$$

The next theorem is due to Sibirskii [14] and Procesi [12].

Theorem 2 *Let $G = O(n,F)$ act on $W(n,m,F)$ by conjugation. The invariants are of the form $\operatorname{tr} P(A_1,\dots,A_m,A_1^t,\dots,A_m^t)$, where P is a noncommutative polynomial in $2m$ matrix variables.*

[1]Presenter of the paper at the conference.

Symmetries in Science VII, Edited by
B. Gruber, Plenum Press, New York, 1995

We can similarly consider invariants under the unitary group. If we set $M^* = \overline{M}^t$, then we get the following theorem due to Sibirskii [14] and Procesi [12].

Theorem 3 *Let $G = U(n)$ act on $W(n,m,\mathbb{C})$ by conjugation. The invariants are of the form $\operatorname{tr} P(A_1,\dots,A_m,A_1^*,\dots,A_m^*)$, where P is a noncommutative polynomial in $2m$ matrix variables.*

We can also consider symplectic invariants. If $n = 2k$, we let

$$J_n = \begin{pmatrix} 0 & -I_k \\ I_k & 0 \end{pmatrix},$$

and define the symplectic involution by

$$M^* = J_n^{-1} M^t J_n = -J_n M^t J_n,$$

then we can state the following theorem due to Procesi [12].

Theorem 4 *Let $G = Sp(k,F)$ act on $W(2k,m,F)$ by conjugation. The invariants are of the form $\operatorname{tr} P(A_1,\dots,A_m,A_1^*,\dots,A_m^*)$, where P is a noncommutative polynomial in $2m$ matrix variables.*

Unfortunately, these theorems give infinite sets of generators for the invariants. We will outline how to find finite sets of generators.

For $n = 2$ the Cayley-Hamilton Theorem says that

$$A^2 - A \operatorname{tr} A + 1/2I[(\operatorname{tr} A)^2 - \operatorname{tr} A^2] = 0$$

or in its polarized version

$$AB + BA - A \operatorname{tr} B - B \operatorname{tr} A + I[\operatorname{tr} A \operatorname{tr} B - \operatorname{tr} AB] = 0.$$

It is well known (see for example [5] or [11]) that the polarized Cayley-Hamilton Theorem implies that any product of three 2×2 matrices is reducible. That is, ABC can be written as a linear combination of matrix products with fewer factors and coefficients expressible in terms of traces. Writing

$$\operatorname{tr}(A,B) = \operatorname{tr} AB - \operatorname{tr} A \operatorname{tr} B,$$

we have

$$\begin{aligned} 2ABC &= A(BC+CB) + (AB+BA)C - [B(AC) + (AC)B] \\ &= AB \operatorname{tr} C + AC \operatorname{tr} B + A\operatorname{tr}(B,C) + AC \operatorname{tr} B + BC \operatorname{tr} A \\ &\quad + C\operatorname{tr}(A,B) - AC \operatorname{tr} B - B\operatorname{tr}(AC) - I\operatorname{tr}(AC,B) \\ &= AB \operatorname{tr} C + AC \operatorname{tr} B + BC \operatorname{tr} A \\ &\quad + A\operatorname{tr}(B,C) - B\operatorname{tr}(AC) + C\operatorname{tr}(A,B) - I\operatorname{tr}(AC,B). \end{aligned}$$

More generally, it follows from the work of Procesi [12] and Razmyslov [13] that the product of $n^2 - 1$ matrices of order n is reducible. For 3×3 matrices the product of 6 matrices is reducible [5]. Using these results, we get finite sets of generators for the algebras of invariants.

SPECIAL ORTHOGONAL GROUP

We will now try to generalize these results to other classical groups. We will start with $SO(n)$, where $n = 2k$. Recall that if M is a skewsymmetric matrix, then $\det M$ is a square and is 0 if the size of M is odd (see for example [8]). We can thus define the Pfaffian of an $n \times n$ skewsymmetric matrix by

$$(\operatorname{pf} M)^2 = \det M, \qquad \operatorname{pf} J_n = 1.$$

The Pfaffian satisfies the relation

$$\operatorname{pf}(gMg^t) = \operatorname{pf} M.$$

For an arbitrary matrix, we let $\widetilde{\operatorname{pf}}\, M$ denote the Pfaffian of the skewsymmetric projection,

$$\widetilde{\operatorname{pf}}\, M = \operatorname{pf}(M - M^t).$$

This is clearly an $SO(2k,F)$ invariant. By abuse of notation we will refer to $\widetilde{\operatorname{pf}}$ as the Pfaffian, too. We have

$$\operatorname{pf}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = b - c.$$

For $G = SO(2,F)$, the invariants $P[W(2,m,F)]^G$ were determined in [1]. They are generated by the invariants $\operatorname{tr} P(A, A^t)$ and $\widetilde{\operatorname{pf}}\, P(A,A^t)$ where $A \in W(2,m,F)$ and P is a noncommutative polynomial.

We will see that for n odd we do not get any more invariants when we restrict $O(n,F)$ to $SO(n,F)$. In the even case we have the following crucial Lemma that indicates why the Pfaffian appears.

Lemma 5 *Let $x_1,\dots,x_{2k} \in F^{2k}$ and let $[x_1,\dots,x_{2k}]$ denote the determinant of the matrix with columns $x_1,\dots,x_{2k}$. Then*

$$[x_1,\dots,x_{2k}] = \widetilde{\operatorname{pf}}(x_1 x_2^t + \dots + x_{2k-1} x_{2k}^t).$$

Proof. Let X be the matrix with columns $x_1,\dots,x_{2k}$ and set

$$\tilde{J} = \begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{pmatrix}.$$

Then

$$\widetilde{\operatorname{pf}}(x_1 x_2^t + \dots + x_{2k-1} x_{2k}^t) = \operatorname{pf}(X\tilde{J}X^t) = \det X \operatorname{pf} \tilde{J} = \det X.$$

□

When $k > 1$ the Pfaffian will no longer be linear. Instead we will consider the polarized Pfaffian, which we will denote by pl. Thus $\operatorname{pl}(A_1,\dots,A_k)$ is the coefficient of $t_1 \cdots t_k$ in the expansion of $\widetilde{\operatorname{pf}}(t_1 A_1 + \dots + t_k A_k)$. It is a symmetric, multilinear function of k matrices and satisfies $\operatorname{pl}(A,\dots,A) = k!\, \widetilde{\operatorname{pf}}(A)$.

When proving our main theorem about the invariants of $SO(2k,F)$ we will use classical invariant theory and consider decomposable matrices of the form $A_i = u_i v_i^t$. Notice that rank $u_i v_i^t \le 1$. We will need the following simple Lemma about the polarized Pfaffian.

Lemma 6 *Let $A_1, \dots, A_k$ be of the form $A_i = u_i v_i^t$. The polarized Pfaffian is alternating in the A_i, i.e.,*

$$\mathrm{pl}(A_{i_1}, \dots, A_{i_k}) = 0$$

if two of the A_{i_j} are equal. Hence

$$\widetilde{\mathrm{pf}}(A_1 + \cdots + A_k) = \mathrm{pl}(A_1, \dots, A_k).$$

Proof. Set $A = t_1 A_{i_1} + \cdots + t_k A_{i_k}$. Then $\mathrm{pl}(A_{i_1}, \dots, A_{i_k})$ is the coefficient of $t_1 \cdots t_k$ in the expansion of $\widetilde{\mathrm{pf}}(A)$. But if the A_{i_j} are not all distinct, the rank of A will be at most $k-1$. But then the rank of $A - A^t$ will be at most $2k-2$, so $\det(A - A^t) = 0$. It follows that

$$\widetilde{\mathrm{pf}}(A) = \mathrm{pf}(A - A^t) = 0,$$

and hence $\mathrm{pl}(A_{i_1}, \dots, A_{i_k}) = 0$ if two of the A_{i_j} are equal. Combining this with the multilinearity of pl, we get

$$\widetilde{\mathrm{pf}}(A_1 + \cdots + A_k) = \mathrm{pl}(A_1 + \cdots + A_k, \dots, A_1 + \cdots + A_k)/k! = \mathrm{pl}(A_1, \dots, A_k).$$

□

We can now prove our main theorem.

Theorem 7 *When $G = SO(2k+1, F)$, the invariants $P[W(2k+1, m, F)]^G$ are the same as the $O(2k+1, F)$ invariants. When $G = SO(2k, F)$, the invariants $P[W(2k, m, F)]^G$ are generated by traces and polarized Pfaffians of the A_i and A_i^t, $i = 1, \dots, m$, i.e.,*

$$\mathrm{tr}\, P(A, A^t) \quad \text{and} \quad \mathrm{pl}\big(P_1(A, A^t), \dots, P_k(A, A^t)\big),$$

where $P, P_1, \dots, P_k$ are noncommutative polynomials and $A \in W(2k, m, F)$.

Proof. The proof will follow from classical invariant theory. We can first reduce the problems to finding the multihomogeneous invariants of order $(d_1, \dots, d_r)$ in m matrices, and then reduce further to studying multilinear invariants of dm matrices where $d = \sum_{i=1}^r d_i$. We will identify $F^n \otimes F^n$ and $M(n, F)$ using

$$u \otimes v \to uv^t.$$

We can assume that $A_i = u_i \otimes v_i$ (the symbolic method). The invariants of $(F^n \otimes F^n)^{\otimes dm}$ are generated by inner products and determinants, i.e., invariants of the form

$$\begin{aligned} &\phi(x_1 \otimes \cdots \otimes x_{2dm}) \\ &= [x_1, \dots, x_n] \cdots [x_{n(l-1)+1}, \dots, x_{nl}] \langle x_{nl+1}, x_{nl+2} \rangle \cdots \langle x_{2dm-1}, x_{2dm} \rangle \end{aligned}$$

where $\langle x_i, x_j \rangle = x_i^t x_j$ and $[x_1, \dots, x_n]$ denotes the determinant of the matrix with columns $x_1, \dots, x_n$.

We first observe that if n is odd, we must have an even number of determinants. But then the whole expression will be an $O(n, F)$ invariant. This proves the first part of the theorem.

We can therefore assume that $n = 2k$. We observe that

$$\langle x_i, x_j \rangle = x_i^t x_j = \mathrm{tr}\, x_i x_j^t = \mathrm{tr}\, x_i \otimes x_j.$$

If w_i is u_i or v_i and w_i' the other one, then a product of the form

$$\langle w_{i_1}, w'_{i_2} \rangle \langle w_{i_2}, w'_{i_3} \rangle \cdots \langle w_{i_r}, w'_{i_1} \rangle$$

is equal to

$$\mathrm{tr}\, B_{i_1} \cdots B_{i_r},$$

where B_i equals A_i or A_i^t depending on whether w_i is u_i or v_i.

We will now show that ϕ can be written in terms of traces and polarized Pfaffians of the A_i and the A_i^t.

Since the product of two determinants is an $O(2k, F)$ invariant and hence expressible in terms of scalar products, we can assume that we have only one determinant. By reordering the A_i and replacing A_i by A_i^t, i.e., interchanging u_i and v_i, we can assume that

$$\begin{aligned} \phi = \pm&[u_1, v_1, \dots, u_l, v_l, u_{l+1}, \dots, u_{2k-l}] \\ &\langle v_{l+1}, u_{2k-l+1} \rangle \langle v_{2k-l+1}, u_{2k-l+2} \rangle \cdots \langle v_{2k-l+d_1}, v_{l+2} \rangle \\ &\langle v_{l+3}, u_{2k-l+d_1+1} \rangle \dots \langle v_{2k-l+d_1+d_2}, v_{l+4} \rangle \\ &\cdots \langle v_{2k-l-1}, u_{2k-l+d_1+\cdots+d_{k-l-1}+1} \rangle \cdots \langle v_{2k-l+d_1+\cdots+d_{k-l}}, v_{2k-l} \rangle Q. \end{aligned}$$

where Q only involves scalar products of u_i and v_i for $i > 2k-l+d_1+\cdots+d_{k-l}$. But this can be expressed as the trace of a polynomial in A_i and A_i^t for $i > 2k-l+d_1+\cdots+d_{k-l}$. We can now apply Lemma 6 and we get that

$$\begin{aligned} [u_1, v_1, \dots, u_l, v_l, u_{l+1}, \dots, u_{2k-l}] &= \widetilde{\mathrm{pf}}(A_1 + \cdots A_l + u_{l+1}u_{l+2}^t + \cdots u_{2k-l-1}u_{2k-l}^t) \\ &= \mathrm{pl}(A_1, \dots, A_l, u_{l+1}u_{l+2}^t, \cdots, u_{2k-l-1}u_{2k-l}^t). \end{aligned} \tag{1}$$

But since the function pl is multilinear, we can combine

$$\langle v_{l+1}, u_{2k-l+1} \rangle \langle v_{2k-l+1}, u_{2k-l+2} \rangle \cdots \langle v_{2k-l+d_1}, v_{l+2} \rangle$$

with $u_{l+1}u_{l+2}^t$ to get

$$\begin{aligned} &u_{l+1}u_{l+2}^t v_{l+1}^t u_{2k-l+1} v_{2k-l+1}^t u_{2k-l+2} \cdots v_{2k-l+d_1}^t v_{l+2} \\ &= u_{l+1} v_{l+1}^t u_{2k-l+1} v_{2k-l+1}^t u_{2k-l+2} \cdots v_{2k-l+d_1}^t v_{l+2} u_{l+2}^t \\ &= A_{l+1} A_{2k-l+1} \cdots A_{2k-l+d_1} A_{l+2}^t. \end{aligned}$$

Repeating this for each of the last $k-l$ terms in (1), we get that

$$\begin{aligned} \phi/Q = \mathrm{pl}(&A_1, \dots, A_l, A_{l+1} A_{2k-l+1} \cdots A_{2k-l+d_1} A_{l+2}^t, \\ &\dots, A_{2k-l-1} A_{2k-l+d_1+\cdots+d_{k-l-1}+1} \cdots A_{2k-l+d_1+\cdots+d_{k-l}} A_{2k-l})^t. \end{aligned}$$

This shows that ϕ is of the required form. □

For similar results for the case of $SO(p, q)$, see [4].

CLASSICAL GROUPS

It turns out that we can use our results about $SO(n)$ invariants to deduce similar results for all the other classical groups. Let us first define what we mean by a classical group. Let F be either $\mathbb{R}$ or $\mathbb{C}$ and let D be a division algebra over F with involution $a \mapsto a^\#$. Let V be a vector space over D and let $(\,,)$ be a nondegenerate sesquilinear form on V that is Hermitian or skew-Hermitian relative to $^\#$. (If the involution is the identity we get bilinear symmetric and skewsymmetric forms.) Let G be the isometry group of $(\,,)$. We will call such groups classical isometry groups.

Table 1

	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$
bilinear and symmetric	$O(p,q)$	$O(n,\mathbb{C})$	–
bilinear and skewsymmetric	$Sp(k,\mathbb{R})$	$Sp(k,\mathbb{C})$	–
sesquilinear and Hermitian	–	$U(p,q)$	$Sp(p,q)$
sesquilinear and skew-Hermitian	–	$U(p,q)$	$Sp(k,\mathbb{H})$

If the involution is either the identity or complex or quaternionic conjugation, we can describe the classical isometry groups by Table 1. (We are assuming that $n = p+q$ or $n = 2k$ when appropriate.) A "–" means that there are no such forms. Helgason [7] writes $SO^*(2k)$ for the group we have called $Sp(k,\mathbb{H})$.

The remaining classical groups are the general and special linear groups $GL(n,F)$ and $SL(n,F)$ and the special isometry groups, i.e., the intersections of $SL(n,F)$ and the classical isometry groups.

As explained in the introduction, Sibirskii [14] gave a set of generators of $P[W]^G$ when G equals $GL(n,F)$, $O(n,F)$ or $U(n)$. His proof for the case of $U(n)$ is essentially Weyl's unitary trick, which can easily be adapted to work for $U(p,q)$. These results were also proved independently by Procesi [12], who also solved the problem for $Sp(k,F)$. His proof for the $O(n,\mathbb{R})$ case can also easily be modified for the case of $O(p,q)$. For the special isometry groups, all the cases are trivial except for the special orthogonal groups.

It therefore remains to discuss the quaternionic groups $Sp(k,\mathbb{H})$ and $Sp(p,q)$. This is discussed in [4].

QUIVERS

We are often interested in more general situations. We might be interested in invariants under block diagonal subgroups like $GL(m)\times GL(n)$ or $O(m)\times O(n)$. It turns out that again the invariants can be expressed in terms of traces of certain products of matrices. In order to express which products, we can form a quiver, and the allowed products will correspond to cycles in the quiver. For details, see [2, 9, 10], but we will outline one example here.

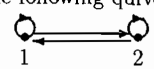
Consider the action of $GL(m)\times GL(n)$ on $M(n+m)$ given by

$$\begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix}^{-1} = \begin{pmatrix} g_1Ag_2^{-1} & g_1Bg_2^{-1} \\ g_2Cg_1^{-1} & g_2Dg_2^{-1} \end{pmatrix}.$$

We can think of this as an action on

$$V = M(m,m)\oplus M(m,n)\oplus M(n,m)\oplus M(n,n) = M_{1,1}\oplus M_{1,2}\oplus M_{2,1}\oplus M_{2,2}.$$

We will represent this action by the following quiver.



The invariants are now given by

$$\operatorname{tr} A_{i_1,i_2}A_{i_2,i_3}\cdots A_{i_{l-1},i_l},$$

where $i_1,\dots,i_l$ is a cycle in the quiver and $A_{i,j}\in M_{i,j}$.

For a classical group, its Levi subgroups are usually products of general linear, orthogonal or symplectic groups. Their invariant theory is useful in understanding the invariant theory in universal enveloping algebras [2].

References

[1] H. Aslaksen, *SO(2) invariants of a set of* 2×2 *matrices*, Math. Scand. **65** (1989), 59–66.

[2] H. Aslaksen, E.-C. Tan and C.-B. Zhu, *Quivers and the invariant theory of Levi subgroups*, J. Funct. Anal. **120** (1994), 163–187.

[3] H. Aslaksen, E.-C. Tan and C.-B. Zhu, *Generators and relations of invariants of* 2×2 *matrices*, Comm. Algebra **22** (1994), 1821–1832.

[4] H. Aslaksen, E.-C. Tan and C.-B. Zhu, *Invariant theory of special orthogonal groups*, Pacific J. Math. (to appear).

[5] J. Dubnov and V. Ivanov, *Sur l'abaissement du degré des polynômes en affineurs*, Dokl. Akad. Nauk SSSR **41** (1943), 95–98.

[6] G. B. Gurevich, *Foundations of the theory of algebraic invariants*, P. Noordhoff Ltd, Groningen, The Netherlands, 1964.

[7] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, Academic Press, New York, USA, 1978.

[8] N. Jacobson, *Basic Algebra I*, W. H. Freeman and Co., New York, USA, 1985.

[9] W. H. Klink and T. Ton-That, *Invariant theory of the block diagonal subgroups of* $GL(n,\mathbb{C})$ *and generalised Casimir operators*, J. of Algebra **145** (1992), 187 – 203.

[10] H. Kraft, *Invariants of quivers (following Procesi)*, private communication.

[11] L. Le Bruyn, *Trace rings of generic 2 by 2 matrices*, Memoirs of the AMS **66** (1987), no. 363.

[12] C. Procesi, *The invariant theory of* $n\times n$ *matrices*, Adv. Math. **19** (1976), 306–381.

[13] Ju. P. Razmyslov, *Trace identities of full matrix algebras over a field of characteristic zero*, Math. USSR Izv. **8** (1974), 727–760.

[14] K. S. Sibirskii, *Algebraic invariants for a set of matrices*, Siberian Math. J. **9** (1968), 115–124.