A Geometric Approach to Classical Galois Theory

R.C. Churchill

Hunter College and the Graduate Center of CUNY, and the
University of Calgary
August 2010

Prepared for the

Kolchin Seminar on Differential Algebra

Graduate Center, City University of New York
27 August 2010

Contents

Throughout the notes K denotes a field, n is a positive integer, and V denotes an n-dimensional vector space over K. The characteristic and minimal polynomials of a K-linear operator $T : V \to V$ are denoted $\mathrm{char}_{T,K}(x)$ and $\mathrm{min}_{T,K}(x)$ respectively: both are elements of $K[x]$. The order of a finite group G is denoted $|G|$.

1. Introduction

There are striking parallels between standard techniques for studying the zeros of single-variable polynomials with field coefficients and standard techniques for studying the zeros of ordinary linear differential operators $\sum_j f_j(x)\frac{d^j}{dx^j}$ on appropriate fields of functions. These become even more apparent when the latter are formulated algebraically, and that is our first item of business.

Let R be a ring, not necessarily commutative, with unity (i.e., multiplicative identity). A function $\delta : r \in R \mapsto r' \in R$ is a *derivation* (*on* or *of* R) if the following two properties hold for all $r_1, r_2 \in R$:

- ("additivity") $(r_1 + r_2)' = r_1' + r_2'$, and
- ("the Leibniz rule") $(r_1 r_2)' = r_1 r_2' + r_1' r_2$.

Note from the Leibniz rule and induction that

$$(\prod_{j=1}^n r_j)' = \sum_j r_1 r_2 \cdots r_{j-1} r_j' r_{j+1} \cdots r_n \tag{1.1}$$

for any $n \geq 2$ and any $r_1, r_2, \ldots, r_n \in R$.

A pair (R, δ) as in the previous paragraph is called a *differential ring*, or a *differential field* when R is a field, although when δ is clear from context one generally refers to R as the differential ring (or field, as the case may be). Indeed, the custom is to minimize specific reference to δ, within reason, by writing $\delta^n r$ as $r^{(n)}$ for any non-negative integer n and any $r \in R$, where $\delta^0 r := r$. For small $n > 0$ primes are also used, e.g., $r'' := \delta^2 r = (r')'$ and $r''' := \delta^3 r = (r'')'$. By regarding an element $r \in R$ as the left multiplication function $\mu_r : t \in R \mapsto rt \in R$ one can view R as sitting within the ring $\mathcal{F}_R$ of (set-theoretic) functions $f : R \to R$, and $\mathcal{F}_R$ is thereby endowed with the structure of an R-algebra. The derivation δ is obviously an element of $\mathcal{F}_R$, and as a result one can consider the R-subalgebra $R[\delta]$ of $\mathcal{F}_R$ generated by R and δ. Any element of $R[\delta]$ can be expressed in the "polynomial" form[1] $\sum_{j=0}^m r_j \delta^j$, and these are our generalizations of "linear ordinary differential operators."

The "parallels" alluded to in the opening paragraph should henceforth be regarded as being between (ordinary) polynomials $p = \sum_{j=0}^n r_j x^j \in R[x]$ (with no derivation assumed on R), and ordinary linear differential operators $\mathcal{L} = \mathcal{L}_p = \sum_{j=0}^n r_j \delta^j \in R[\delta]$.

[1] This form will be unique under quite mild assumptions which we will simply assume are satisfied, and, to avoid a lengthy digression, will not state explicitly. Uniqueness holds, for example, when R is the field $\mathbb{C}(z)$ of rational functions (quotients of polynomial functions) and $\delta = \frac{d}{dz}$.

However, before beginning that discussion it might be a good idea to point out one major difference between the R-algebras $R[x]$ and $R[\delta]$, i.e., that the former is commutative (as readers are surely aware), while the latter is generally not (as readers may not be aware). Indeed, for any $r, t \in R$ one sees from the Leibniz rule that

$$\begin{aligned}(\delta \circ \mu_r)(t) &= \delta(\mu_r(t)) \\ &= \delta(rt) \\ &= r\,\delta t + \delta r\, t \\ &= (\mu_r \circ \delta)(t) + \delta r t \\ &= (\mu_r \circ \delta + \mu_{\delta r})(t),\end{aligned}$$

and we therefore have

$$\delta \circ \mu_r = \mu_r \circ \delta + \mu_{\delta r}. \tag{1.2}$$

In keeping with blurring all distinction between r and μ_r, and minimizing specific references to δ, this would generally be written

$$\delta r = r\delta + r' \qquad \text{for all} \qquad r \in R. \tag{1.3}$$

A derivation $r \in R \mapsto r' \in R$ is *trivial* when $r' = 0$ for all $r \in R$, and in that case we see from (1.3) that $R[\delta]$ is commutative. Indeed, under that hypothesis the correspondence $x \leftrightarrow \delta$ induces an isomorphism of the R-algebras $R[x]$ and $R[\delta]$ which one learns to exploit in a first course on ordinary differential equations, wherein one generally assumes that $R = \mathbb{R}$ or $\mathbb{C}$ and that $\delta = \frac{d}{dx}$. In that context the first hint of any parallelism between polynomials $p = \sum_{j=0}^{n} r_j x^j$ and the corresponding linear operators $\mathcal{L}_p = \sum_{j=0}^{n} r_j \delta^j$ takes the form of the observation that for any $\lambda \in R$ one has $\mathcal{L}_p e^{\lambda x} = p(\lambda) e^{\lambda x}$, hence that *for any root λ of p the function $y = e^{\lambda x}$ is a solution of the n^{th}-order equation $\mathcal{L}_p y = 0$.*

For the remainder of the introduction R denotes a (commutative) *integral domain with unity $1 = 1_R$.*

For our purposes the deeper analogies between polynomials and linear differential operators are most easily revealed by treating the former in terms of Vandermonde matrices and Vandermonde determinants. For ease of reference, and to establish our notation, we recall the definitions.

Suppose $n \geq 1$ and $r_1, r_2, \ldots, r_n$ are (not necessarily distinct) elements of $R[x]$. The classical *Vandermonde matrix* of $r := (r_1, r_2, \ldots, r_n) \in R[x]^n$ is defined to be the $n \times n$ matrix

$$\text{vdmm}_{R[x],n}(r) := (r_j^{i-1}) \in R[x], \tag{1.4}$$

which is sometimes more conveniently expressed as $\text{vdmm}_{R[x],n}(r_1, r_2, \ldots, r_n)$. Using full matrix notation the definition is:

$$\text{vdmm}_{R[x],n}(r) = \text{vdmm}_{R[x],n}(r_1, r_2, \ldots, r_n) := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_n \\ \vdots & & & \vdots \\ r_1^{n-2} & r_2^{n-2} & \cdots & r_n^{n-2} \\ r_1^{n-1} & r_2^{n-1} & \cdots & r_n^{n-1} \end{bmatrix}, \tag{1.5}$$

The associated determinant

$$\text{vdmd}_{R[x],n}(r) = \text{vdmd}_{R[x],n}(r_1, r_2, \ldots, r_n) := \det(\text{vdmm}_{R[x],n}(r)) \tag{1.6}$$

is called the *Vandermonde determinant* of $r = (r_1, r_2, \ldots, r_n)$.

When $r \in R[x]^{n+1}$ has the form $(t_1, t_2, \ldots, t_n, x)$, with $t_j \in R$ for $j = 1, \ldots, n$, it proves convenient to first express the $(n+1) \times (n+1)$ Vandermonde matrix $\text{vdmm}_{R[x],n+1}(r)$ as $\text{vdmm}_{R,n+1}(t; x) = \text{vdmm}_{R,n+1}(t_1, t_2, \ldots, t_n, x)$, and to then relabel t as r and all t_j as r_j. In other words,

$$\left\{ \begin{array}{rcl} \text{vdmm}_{R,n+1}(r; x) & = & \text{vdmm}_{R,n+1}(r_1, r_2, \ldots, r_n; x) \\ & := & \text{vdmm}_{R[x],n+1}(r_1, r_2, \ldots, r_n, x) \\ & = & \begin{bmatrix} 1 & \cdots & 1 & 1 \\ r_1 & \cdots & r_n & x \\ \vdots & & & \vdots \\ r_1^{n-1} & \cdots & r_n^{n-1} & x^{n-1} \\ r_1^n & \cdots & r_n^n & x^n \end{bmatrix}. \end{array} \right. \tag{1.7}$$

The associated Vandermonde determinant is expressed accordingly, i.e.,

$$\text{vdmd}_{R,n+1}(r; x) := \text{vdmd}_{R[x],n+1}(r_1, r_2, \ldots, r_n, x) := \det(\text{vdmm}_{R,n+1}(r; x)). \tag{1.8}$$

Proposition 1.9 : *Assume* $n \geq 1$, *choose any* (not necessarily distinct) *elements* $r_1, r_2, \ldots, r_n \in R$, *and set* $r := (r_1, r_2, \ldots, r_n) \in R^n$. *Then the following assertions hold.*

(a) *The Vandermonde determinant* $\mathrm{vdmd}_{R[x],n}(r)$ *has value* $\prod_{i>j}(r_i - r_j)$.

(b) *The determinant* $\mathrm{vdmd}_{R,n+1}(r; x)$ *has value* $\mathrm{vdmd}_{R[x],n}(r) \cdot \prod_j (x - r_j)$.

(c) $\mathrm{vdmd}_{R[x],n}(r) \neq 0$ *if and only if the elements* r_j *are pairwise distinct.*

(d) *Suppose* R *is a field,* $n \geq 1$, $p = \sum_{j=0}^{n} a_j x^j \in R[x]$, *and* $a_n \neq 0$. *Then* p *has at most* n *roots in* R. *If the number of distinct roots of* p *in* R *is* $k \geq 1$, *then a collection of roots* $t_1, t_2, \ldots, t_k \in R$ *of* p *is precisely that collection of distinct roots if and only if* $\mathrm{vdmd}_{R[x],k}(t_1, t_2, \ldots, t_k) \neq 0$.

(e) *The determinant* $\mathrm{vdmd}_{R,n+1}(r; x)$ *is a degree* n *polynomial in* $R[x]$ *having* $r_1, r_2, \ldots, r_n$ *as a complete set of roots. Moreover, when* $\mathrm{vdmd}_{R[x],n}(r)$ *is a unit of* R *the* r_j *are pairwise distinct and the product*

(i) $$\big(\mathrm{vdmd}_{R[x],n}(r)\big)^{-1} \cdot \mathrm{vdmd}_{R,n+1}(r; x)$$

is a monic degree n *polynomial in* $R[x]$ *having precisely these roots.*

(f) *Suppose* $p \in R[x]$ *is a monic polynomial of degree* $n \geq 1$ *having* $r_1, r_2, \ldots, r_n$ *as a complete set of roots and* $\mathrm{vdmd}_{R[x],n}(r)$ *is a unit of* R. *Then* p *must be the polynomial defined in* (i) *of* (e).

(g) *Suppose* $r_1, r_2, \ldots, r_n$ *are the zeros of a monic polynomial* $p = x^n + \sum_{k=0}^{n-1} a_k x^k \in R[x]$. *Then*

(ii) $$\theta(\mathrm{vdmd}_{R[x],n}(r)) = -a_{n-1} \cdot \mathrm{vdmd}_{R[x],n}(r),$$

where $\theta(\mathrm{vdmd}_{R[x],n}(r))$ *denotes the determinant of the "modified" Vandermonde matrix*

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_n \\ \vdots & & & \vdots \\ r_1^{n-2} & r_2^{n-2} & \cdots & r_n^{n-2} \\ r_1^{n} & r_2^{n} & \cdots & r_n^{n} \end{bmatrix},$$

(h) **(Reduction of Degree)** *Suppose* $r, s \in R$, r *is a zero of a polynomial* $p = \sum_{k=0}^{n} a_k x^k \in R[x]$, *and* $s \neq 0$. *Then* $r+s$ *is a zero of* p *if and only if* s *is a zero of the polynomial*

(iii) $$\sum_{j=0}^{n-1} \left(\sum_{k=j+1}^{n} \binom{k}{j+1} a_k r^{k-1-j} \right) x^j.$$

Assertions (d) and (e) can be summarized by the statement: a monic polynomial in $R[x]$ is uniquely determined by its roots.

As the reader may suspect, there are much simpler ways to state several of these results. We have emphasized the Vandermonde formulation because, as we will see in Proposition 1.18, it offers a very effective means for highlighting analogies between polynomials and linear differential operators.

Proof :

(a) This is standard. A proof, generally by induction, can be found in any reasonable[2] text on linear algebra.

(b) Since the indicated determinant is the same as $\mathrm{vdmd}_{R[x],n+1}(r_1, r_2, \ldots, r_n, x)$ we see from (a) (with slight relabeling) that the value is $\prod_j (x - r_j) \prod_{n \geq i > j} (r_i - r_j) = \prod_j (x - r_j) \cdot \mathrm{vdmd}_{R[x],n}(r)$.

(c) Immediate from (a) and the integral domain hypothesis on R.

(d) Suppose $t_1, t_2, \ldots, t_{n+1} \in R$ are roots of p, i.e., that $\sum_{i=0}^{n} a_i t_j^i = 0$ for $j = 1, 2, \ldots, n+1$. Express this collection of $n+1$ equalities in the matrix form

$$\begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^n \\ 1 & t_2 & t_2^2 & \cdots & t_2^n \\ \vdots & & & & \vdots \\ 1 & t_{n+1} & t_{n+1}^2 & \cdots & t_{n+1}^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The matrix on the left is immediately recognized as the transpose of the Vandermonde matrix $\mathrm{vdmm}_{R[x],n=1}(t_1, t_2, \ldots, t_n)$. Since determinants are unaffected by transposition, the corresponding determinant must be $\mathrm{vdmd}_{R[x],n+1}(t_1, t_2, \ldots, t_{n+1})$, which for simplicity we abbreviate as v. If the collection $t_1, t_2, \ldots, t_{n+1}$ consists of distinct roots of p in R then $v \neq 0$ by (c), and Cramer's rule then forces $a_0 = a_1 = \cdots = a_n = 0$.

[2]In the opinion of this author, a text on linear algebra is *reasonable* if and only if it contains a proof of this result.

Since $a_n \neq 0$, this contradiction proves that $v = 0$, hence by (c) that $t_1, t_2, \ldots, t_{n+1}$ are not all distinct. The collection of roots of p therefore contains at most n elements.

The final assertion is evident from (c).

(e) The initial assertion is immediate from (b). If $\mathrm{vdmd}_{R[x],n}(r)$ is a unit it cannot be 0, and the pairwise distinct assertion is then seen from (c). By (a) and (b) the product appearing in (i) is equal to $\prod_j (x - r_j)$, and the final assertion follows.

(f) Denote the monic polynomial defined in (i) by q. Then either $p = q$, in which case we are done, or $p \neq q$ and $\deg(p - q) < n$. The second alternative is impossible by the integral domain hypothesis on R: a non-zero polynomial in $R[x]$ can have no more roots than its degree, whereas $p - q$ admits the n distinct roots $r_1, r_2, \ldots, r_n$.

(g) By hypothesis we can replace each r_j^n in the displayed matrix with $-\sum_k^{n-1} a_k r_j^k$, and then, using standard properties of determinants, reduce the expression to

$$\theta(\mathrm{vdmd}_{R[x],n}(r)) = \sum_k^{n-1} a_k \cdot \det\left(\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ r_1 & r_2 & r_3 & \cdots & r_n \\ \vdots & & & \vdots & \\ r_1^{n-2} & r_2^{n-2} & r_3^{n-2} & \cdots & r_n^{n-2} \\ r_1^k & r_2^k & r_3^k & \cdots & r_n^k \end{bmatrix} \right).$$

The matrices corresponding to $1 \leq k < n - 1$ each have two equal rows, hence determinant zero, and the matrix corresponding to $k = n - 1$ is $\mathrm{vdmm}_{R[x],n}(r)$. The result follows.

(h) From the binomial theorem we have

$$(r + s)^k = \sum_{j=0}^k \begin{pmatrix} k \\ j \end{pmatrix} r^{k-j} s^k$$

for any integer $k \geq 1$. Therefore,

$$\begin{aligned}
\textstyle\sum_{k=0}^{n} a_k(r+s)^k &= a_0(r+s) + \textstyle\sum_{k=1}^{n} a_k(r+s)^k \\
&= a_0(r+s) + \textstyle\sum_{k=1}^{n} a_k \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) r^{k-j} s^j \\
&= a_0(r+s) + \textstyle\sum_{k=1}^{n} a_k \left[r^k + \sum_{j=1}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) r^{k-j} s^j \right] \\
&= \left(\textstyle\sum_{k=0}^{n} a_k r^k \right) + \textstyle\sum_{k=1}^{n} \sum_{j=1}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a_k r^{k-j} s^j \\
&= 0 + \textstyle\sum_{j=1}^{n} \left(\sum_{k=j}^{n} \left(\begin{array}{c} k \\ j \end{array} \right) a_k r^{k-j} \right) s^j \\
&= \textstyle\sum_{j=0}^{n-1} \left(\sum_{k=j+1}^{n} \left(\begin{array}{c} k \\ j+1 \end{array} \right) a_k r^{k-(j+1)} \right) s^{j+1} \\
&= s \cdot \textstyle\sum_{j=0}^{n-1} \left(\sum_{k=j+1}^{n} \left(\begin{array}{c} k \\ j+1 \end{array} \right) a_k r^{k-1-j} \right) s^j,
\end{aligned}$$

and (h) follows.

q.e.d.

Examples 1.10 : Proposition 1.9 has been formulated to stress analogies with linear differential operators, and if that thought is not kept in mind several of assertions can seem rather silly. This comment applies, in particular, to assertion (h). After all, if we know that r is a root of p, the remaining roots will be the roots of the lower degree polynomial $p/(x-r)$, so why bother looking at the polynomial satisfied by the difference of two roots? The answer from this author's perspective is: because the analogous concept for linear differential operators, i.e., "reduction of order," is a very important technique. In the polynomial context what the concept seems to offer is an alternate approach to certain classes of problems, and therefore, perhaps, a means to alleviate boredom.

(a) Let $p \in \mathbb{Q}[x]$ be the Lagrange interpolation polynomial of the data $(x,y) = (2,0), (4,7), (5,0)$ and $(8,4)$, i.e., $f = \frac{67}{72}x^3 - \frac{989}{72}x^2 + \frac{2{,}155}{36}x - \frac{650}{9}$. From the given initial data we see that p has roots $r_1 := 2$ and $r_2 := 5$, and by making

the (arbitrary) choice $r = r_1$ the polynomial in (iii) of Proposition 1.9(h) is calculated to be

$$\tfrac{67}{72} \cdot (x^2 - \tfrac{587}{67}x + \tfrac{1158}{67} = \tfrac{67}{72} \cdot (x-3)(x - \tfrac{386}{67}).$$

The root $s = 3$ of this lower degree polynomial corresponds to the known root $2 + s = 5$ of f, and the choice $s = \frac{386}{67}$ gives the remaining root $2 + \frac{386}{67} = \frac{520}{67}$ of f.

(b) **(Tartaglia-Cardano)** In combination with a few classically known variable substitutions, Proposition 1.9(h) can be used to produce all the solutions of any cubic equation

(i) $$x^3 + a_2x^2 + a_1x + a_0 = 0, \qquad \text{with} \qquad a_0, a_1, a_2 \in \mathbb{Q}.$$

(In fact the method works for $a_0, a_1, a_2 \in \mathbb{R}$, but the more general case is of less historical interest.) We first note that a complex number y_0 is a solution to the equation

(ii) $$y^3 + (a_1 - a_2^2/3)y + a_0 + 2a_2^3/27 - a_1a_3/3 = 0$$

if and only if

(iii) $$x_0 = y - a_2/3$$

is a solution of (i), and we are therefore reduced to considering cubic equations of the form

(iv) $$y^3 + py + q = 0, \qquad p, q \in \mathbb{Q}.$$

In order to apply Proposition 1.9(h) to this last equation we need only one solution, and since $y = (-q)^{1/3}$ is such when $p = 0$, we can assume w.l.o.g. that $p \neq 0$. In that case the classical trick is to make the substitution

(v) $$y = 2 \cdot \tfrac{\sqrt{3}}{3} \cdot \sqrt{|p|} \cdot z$$

in (iv), so as to convert that equation to

$$\tfrac{2}{9} \cdot \sqrt{3} \cdot |p|^{3/2} \cdot \Big(4z^3 + 3 \cdot \tfrac{p}{|p|} \cdot z + \tfrac{3}{2} \cdot \tfrac{\sqrt{3}}{|p|^2} \cdot q\Big) = 0,$$

and the problem is thereby reduced to the consideration of

(vi) $$4z^3 \pm 3z + r = 0, \qquad \text{where} \qquad r := \tfrac{3}{2} \cdot \tfrac{\sqrt{3}}{|p|^2} \cdot q,$$

and the choice of the plus or minus sign is made to agree with the sign of $p/|p|$.

The Plus Case

In this case one sees from the hyperbolic function identity $4\sinh^3\theta + 3\sinh\theta = \sinh 3\theta$ that $z = \sinh(\frac{1}{3}\sinh^{-1} r)$ is a solution of (vi).

The Minus Cases

- If $r \geq 1$ one sees from the identity $4\cosh^3\theta - 3\cosh\theta = \cosh 3\theta$ that $z = \cosh(\frac{1}{3}\cosh^{-1}(r))$ is a solution.
- If $r \leq -1$ we are reduced to the $+$ case by replacing z by $-z$.
- If $|r| < 1$ one sees from the trigonometric identity $4\cos^3\theta - 3\cos\theta = \cos 3\theta$ that $z = \cos(\frac{1}{3}\cos^{-1}(r))$ is a solution. (In fact in this subcase one can produce three roots with this method.)

In summary[3], we now see that we can always produce a solution z_0 to (vi), whereupon tracing backwards through substitutions (v) and (iii), and noting that

$$\begin{aligned} p &:= a_1 - a_2^2/3, \\ q &:= a_0 + 2a_2^3/27 - a_1 a_3/3, \end{aligned}$$

we conclude that

$$x_0 := \begin{cases} (-q)^{1/3} - \frac{a_0}{3} & \text{if } p = 0; \text{ and} \\ 2 \cdot \frac{\sqrt{3}}{3} \cdot \sqrt{|a_1 - a_2^2/3|} \cdot z_0 \; - \; \frac{a_0}{3} & \text{otherwise,} \end{cases}$$

where the square root in the last line denotes the non-negative square root, will be a solution of (i).

To obtain two additional solutions check that the polynomial in (iii) of Proposition 1.9(h), with $n := 3$ and $r := x_0$, is given by

$$x^2 + (a_2 + 3x_0)x + 3x_0^2 + 2a_2 x_0 + a_1,$$

and that the roots of this polynomial are

$$\tfrac{1}{2} \cdot \left(-a_2 - 3x_0 \pm \sqrt{a_2^2 - 2a_2 x_0 - 3x_0^2 - 4a_1} \right).$$

[3]The argument thus far has been adapted from [B-M, Chapter IV, §4, Appendix, pp. 90-1].

The three (not necessarily distinct) solutions of (i) are therefore

$$\begin{aligned} &x_0, \\ &x_0 + \tfrac{1}{2}\left(-a_2 - 3x_0 + \sqrt{a_2^2 - 2a_2x_0 - 3x_0^2 - 4a_1}\right), \quad \text{and} \\ &x_0 + \tfrac{1}{2}\left(a_2 - 3x_0 - \sqrt{a_2^2 - 2a_2x_0 - 3x_0^2 - 4a_1}\right). \end{aligned}$$

Now assume, for the next four paragraphs, that $R = (R, \delta)$ is a differential ring (or field).

Elements $r \in R$ satisfying $r' $ are the *constants* (of the derivation δ). They form a subring $R_C \subset R$ (a subfield if R is a field) called the *ring* (resp. *field*) *of constants* (*of* R, or of (R, δ)). In particular, $0, 1 \in R_C$. Note from the definition of a derivation that $\delta : R \to R$ is an R_C-linear map (when R is regarded as an R_C-module).

Suppose $\mathcal{L} \in R[\delta]$, say $\mathcal{L} = \sum_{j=0}^{n} a_j \delta^j$. Since $R[\delta] \subset \mathcal{F}_R$, we can view $\mathcal{L}$ as a function from R into R, which from the conclusion of the previous paragraph must be R_C-linear, and it is therefore reasonable to refer to the kernel $\ker(\mathcal{L})$ of $\mathcal{L}$. An element $r \in R$ within that kernel is called a *zero* of $\mathcal{L}$, and also a *solution* of the ordinary linear differential equation

$$\sum_{j=0}^{n} a_j y^{(j)} = 0. \tag{1.11}$$

Indeed, the condition $\mathcal{L}(r) = 0$ for an element $r \in R$ to be in $\ker(\mathcal{L})$ is precisely

$$\sum_{j=0}^{n} a_j s^{(j)} = 0. \tag{1.12}$$

The analogue for the differential ring R of the Vandermonde matrix the *Wronski matrix*[4]

$$\mathrm{wrm}_{R[\delta],n}(r) := \left(r_j^{(i-1)}\right), \qquad r := (r_1, r_2, \ldots, r_n) \in R[\delta]^n, \tag{1.13}$$

which is sometimes more conveniently expressed as $\mathrm{wrm}_{R[\delta],n}(r_1, r_2, \ldots, r_n)$. In full matrix notation:

$$\mathrm{wrm}_{R[\delta],n}(r) = \mathrm{wrm}_{R[\delta],n}(r_1, r_2, \ldots, r_n) := \begin{bmatrix} r_1 & r_2 & \cdots & r_n \\ r_1' & r_2' & \cdots & r_n' \\ \vdots & & & \vdots \\ r_1^{(n-2)} & r_2^{(n-2)} & \cdots & r_n^{(n-2)} \\ r_1^{(n-1)} & r_2^{(n-1)} & \cdots & r_n^{(n-1)} \end{bmatrix}. \tag{1.14}$$

[4]This name is not common in the literature; it is used since it mimics the terminology "Vandermonde matrix," which one does encounter elsewhere.

The associated determinant

$$\text{wron}_{S[\delta],n}(r) := \text{wron}_{S[\delta],n}(r_1, r_2, \ldots, r_n) := \det(\text{wrm}_{S[\delta],n}(r)) \tag{1.15}$$

is called the[5] *Wronskian* of r.

When $r \in R[\delta]^{n+1}$ has the form $(t_1, t_2, \ldots, t_n, \delta)$, with $t_j \in R$ for $j = 1, \ldots, n$, it proves convenient to first express the $(n+1)\times(n+1)$ Wronski matrix $\text{wrm}_{R[\delta],n+1}(r)$ as $\text{wrm}_{R,n+1}(t;\delta) = \text{wrm}_{R,n+1}(t_1, t_2, \ldots, t_n; \delta)$ and to then relabel all t_j as r_j. In other words,

$$\left\{ \begin{array}{rcl} \text{wrm}_{R,n+1}(r;\delta) & = & \text{wrm}_{R,n+1}(r_1, r_2, \ldots, r_n; \delta) \\ & := & \text{wrm}_{R[\delta],n+1}(r_1, r_2, \ldots, r_n, \delta) \\ & = & \left[\begin{array}{ccccc} r_1 & r_2 & \cdots & r_n & 1 \\ r_1' & r_2' & \cdots & r_n' & \delta \\ \vdots & & & \vdots & \\ r_1^{(n-1)} & r_2^{(n-1)} & \cdots & r_n^{(n-1)} & \delta^{n-1} \\ r_1^{(n)} & r_2^{(n)} & \cdots & r_n^{(n)} & \delta^n \end{array} \right] . \end{array} \right. \tag{1.16}$$

Many of the familiar rules of determinants fail with this matrix since $R[\delta]$ is not commutative, and as a result one needs to exercise caution. For example, when A is a square matrix with entries in a commutative ring interchanging two columns of a given matrix, and then two rows, will not change the determinant, but this fails when the entries of A are in $R[\delta]$. For example, if $A = \left[\begin{array}{cc} r & 1 \\ r' & \delta \end{array} \right]$ these successive interchanges result in the matrix $B := \left[\begin{array}{cc} \delta & r' \\ 1 & r \end{array} \right]$, and, assuming the usual definition of determinant, we would find that

$$\begin{array}{rcll} \det(A) & = & r\delta - r' & \\ & = & (\delta r - r') - r' & \text{(by (1.3))} \\ & = & \delta r - 2r' & \\ & \neq & \delta r - r' & \text{(assuming, of course, that } 2r' \neq r') \\ & = & \det(B). & \end{array}$$

[5]To be consistent with the terminology "Vandermonde determinant" one should here refer to the "Wronski determinant," but "Wronskian" is the accepted standard. (Alternatively, why not "Vandermondian?") This should also explain why this author did not use the notation $\text{wrd}_{S[\delta],n}(r)$ in place of $\text{wron}_{S[\delta],n}(r)$.

For our purposes it suffices to define the "determinant" of the displayed matrix in (1.16) by

(1.17)

$$\begin{cases} \mathrm{wron}_{R,n+1}(r;\delta) & := & \sum_{k=1}^{n+1}(-1)^{n+k}\det((r_j^{(\sigma(i)-1)};k))\delta^{k-1} \\ & = & \mathrm{wron}_{R[\delta],n}(r)\delta^n + \sum_{k=1}^{n}(-1)^{n+k}\det((r_j^{(\sigma(i)-1)};k))\delta^{k-1}, \end{cases}$$

where $(r_j^{(i-1)};k)$ denotes the $n\times n$ matrix (with entries in the commutative ring R) obtained by removing row k from the $(n+1)\times n$ matrix $(r_j^{(i-1)})$. In other words, the determinant of this matrix is defined to be the result of the usual calculation of a determinant by "expanding down the final column," with the added proviso that *in each product appearing within the sum we must place the term from the last column in the extreme right position.* (Since R is assumed commutative, the minors appearing in this definition obey the usual properties.) Thus, for example,

$$\begin{aligned} \mathrm{wron}_{R,3}(r;\delta) &= \text{“}\det\text{“}(\begin{bmatrix} r_1 & r_2 & 1 \\ r_1' & r_2' & \delta \\ r_1'' & r_2'' & \delta^2 \end{bmatrix}) \\ &:= \det(\begin{bmatrix} r_1' & r_2' \\ r_1'' & r_2'' \end{bmatrix}) - \det(\begin{bmatrix} r_1 & r_2 \\ r_1'' & r_2'' \end{bmatrix})\cdot\delta + \det(\begin{bmatrix} r_1 & r_2 \\ r_1' & r_2' \end{bmatrix})\cdot\delta^2 \\ &= (r_1'r_2'' - r_2'r_1'') - (r_1r_2'' - r_2r_1'')\cdot\delta + (r_1r_2' - r_2r_1')\cdot\delta^2. \end{aligned}$$

In the following statement, and henceforth, S_n denotes the symmetric group on n letters, i.e., the group (under composition) of bijections $\sigma : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$. The sign of a permutation $\sigma \in S_n$ is denoted $\mathrm{sgn}(\sigma)$.

Proposition 1.18 : *Assume $\delta : R \to R$ is a derivation, let $n \geq 1$, choose any* (not necessarily distinct) *elements $r_1, r_2, \ldots, r_n \in R$, and set $r := (r_1, r_2, \ldots, r_n) \in R^n$. Then the following assertions hold.*

(a) *The Wronskian determinant $\mathrm{wron}_{R[\delta],n}(r)$ has value $\sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)\prod_j r_j^{(\sigma(i)-1)}$.*

(b) *The "determinant" $\mathrm{wron}_{R,n+1}(\delta, r)$ has value $\sum_{k=1}^{n+1}(-1)^{k+1}\det((r_j^{(\sigma(i)-1)};k))\delta^{k-1}$.*

(c) *Suppose R is a field. Then $\mathrm{wron}_{R[\delta],n}(r) \neq 0$ if and only if the elements r_j are linearly independent over R_C.*

(d) *Suppose R is a field, $n \geq 0$, $\mathcal{L} = \sum_{j=0}^{n} a_j\delta^j \in R[\delta]$, and $a_n \neq 0$. Then $\ker(\mathcal{L})$ is a vector space over R_C of dimension at most n. If $\dim_{K_C}(\ker(\mathfrak{L})) = k$, a*

collection $t_1, t_2, \ldots, t_k \in \ker(\mathfrak{L})$ *is a basis if and only if* $\operatorname{wron}_{K[x],k}(t_1, t_2, \ldots, t_k) \neq 0$.

(e) *The "determinant"* $\operatorname{wron}_{R,n+1}(r;\delta)$ *is a linear differential operator of order at most* n *having* $r_1, r_2, \ldots, r_n$ *as zeros. Moreover, when* $\operatorname{wron}_{R[\delta],n}(r)$ *is a unit of* R *the* r_j *are pairwise distinct and the product*

(i) $$\left(\operatorname{wron}_{R[\delta],n}(r)\right)^{-1} \cdot \operatorname{wron}_{R,n+1}(r;\delta)$$

is a monic linear differential operator of order n *in* $R[\delta]$ *with kernel having* $r_1, r_2, \ldots, r_n$ *as a basis.*

(f) *Suppose* R *is a field, that* $\mathcal{L} \in R[\delta]$ *is a monic linear differential operator of order* $n \geq 1$ *with kernel containing* $r_1, r_2, \ldots, r_n$, *and that* $\operatorname{wron}_{R[\delta],n}(r) \neq 0$. *Then* $\mathcal{L}$ *must be the linear differential operator defined in* (i) *of* (e), *and* $r_1, r_2, \ldots, r_n$ *must be a basis of* $\ker(\mathcal{L})$.

(g) **(Abel-Liouville)** *Suppose* $r_1, r_2, \ldots, r_n \in R$ *are zeros of a linear differential operator* $\mathcal{L} = \delta^n + \sum_{k=0}^{n-1} a_k \delta^k \in R[\delta]$. *Then*

(ii) $$\delta(\operatorname{wron}_{R[\delta],n}(r)) = -a_{n-1} \cdot \operatorname{wron}_{R[\delta],n}(r).$$

(h) **(Reduction of Order)** *Suppose* $r, s \in R$ *and* r *is a zero of a linear differential operator* $\mathcal{L} = \sum_{k=0}^{n} a_k \delta^k \in R[\delta]$. *Then* rs *is a zero of* $\mathcal{L}$ *if and only if* s' *is a zero of the linear differential operator*

(iii) $$\sum_{j=0}^{n-1} \left(\sum_{k=j+1}^{n} \binom{k}{j+1} a_k r^{(k-1-j)} \right) \delta^j.$$

Proof :

(a) Obvious from (1.13) (and the formula $\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(i)j}$ for computing the determinant of a square matrix $A = (a_{ij})$).

(b) This simply repeats definition (1.17).

(c) We will prove the contrapositive equivalence: that the Wronskian is zero if and only if the elements $r_1, r_2, \ldots, r_n$ are linearly dependent over R_C.

The vanishing of that Wronskian is equivalent to the existence of a dependence relation (over R) among the columns of $\operatorname{wrm}_{R[x],n}(r)$. From the structure exhibited

in (1.13) this in turn is seen to have the following equivalent formulation: there are elements $c_1, c_2, \ldots, c_n \in R$, not all 0, such that

(iv) $$\sum_{j=1}^{n} c_j r_j^{(m)} = 0 \qquad \text{for} \qquad m = 0, \ldots, n-1.$$

We are thereby reduced to proving:

> *Equalities* (iv) *hold, for some collection* $c_1, c_2, \ldots, c_n \in R$, *not all of which vanish, if and only if the collection* $r_1, r_2, \ldots, r_n$ *is linearly dependent over* R_C.

$\Rightarrow$ We argue by induction on $n \geq 1$. As the case $n = 1$ is trivial, we may assume that $n > 1$ in (iv) and that the result holds for any subset of R with at most $n-1$ elements.

If $c_1 = 0$ in (iv) the given sums can be expressed as $\sum_{j=2}^{n} c_j r_n^{(m)}$, with not all of $c_2, c_3, \ldots, c_n \in R$ equal to 0, and from the induction hypothesis we can then conclude that $r_2, r_3, \ldots, r_n$ are linearly dependent over R_C. The same then holds for $r_1, r_2, \ldots, r_n$.

If $c_1 \neq 0$ in (iv) then[6] w.l.o.g. we may assume that $c_1 = 1$. For $m = 0, 1, \ldots, n-2$ we then see from (iv) and the membership $1 \in R_C$ that

$$\begin{aligned} 0 &= (\textstyle\sum_{j=1}^{n} c_j r_j^{(m)})' \\ &= \textstyle\sum_{j=1}^{n} c_j r_j^{(m+1)} + \sum_{j=2}^{n} c_j' r_j^{(m)} \\ &= 0 + \textstyle\sum_{j=2}^{n} c_j' r_j^{(m)} \quad \text{(again by (iv))} \\ &= \textstyle\sum_{j=2}^{n} c_j' r_j^{(m)}. \end{aligned}$$

There are now two possibilities: $c_j' \neq 0$ for at least one j between 2 and n; or $c_j' = 0$ for all such j.

In the first case the calculation establishes a dependence relation over R for the collection $r_2, r_3, \ldots, r_n$, and by induction there is then such a relation over R_C, say $0 = \sum_{j=2}^{n} \hat{c}_j r_j$, where $\hat{c}_j \in R_C$ for $j = 2, 3, \ldots, n$ and not all vanish. By expressing the relation as $0 = 0 \cdot r_1 + \sum_{j=2}^{n} \hat{c}_j r_j$, and recalling that $0 \in R_C$, we conclude that the collection $r_1, r_2, \ldots, r_n$ is linearly dependent over R_C.

In the second case we have $c_j \in R_C$ for $j = 2, 3, \ldots, n$, and, because we are assuming $c_1 = 1$, equality (iv) for $m = 0$ is simply $r_1 + \sum_{j=1}^{n} c_j r_j = 0$. Since $1 \in R_C$, and we have therefore constructed a dependence relation for $r_1, r_2, \ldots, r_n$ over R_C.

[6] This is where the field hypothesis is used.

$\Leftarrow$ If the collection $r_1, r_2, \ldots, r_n$ is linearly dependent over R_C there are elements $c_1, c_2, \ldots, c_n \in R_C$, not all 0, such that $\sum_{j=1}^n c_j r_j = 0$. Using the R_C-linearity of δ we then see that m applications of this derivation, for any integer $1 \leq m \leq n-1$, gives $\sum_{j=1}^n c_j r_j^{(m)} = 0$, precisely as required.

(d) As already noted, following the proof of Proposition 1.9, the mapping $\mathcal{L} : R \to R$ is linear over R_C, and by elementary linear algebra the kernel of a linear operator is always a subspace. To prove the dimension statement suppose $t_1, t_2, \ldots, t_{n+1} \in R$ are in $\ker(\mathcal{L})$, i.e., that $\sum_i a_i t_j^{(i)} = 0$ for $j = 1, 2, \ldots, n+1$. Express this collection of $n+1$ equations in the matrix form

$$\begin{bmatrix} t_1 & t_1' & t_1'' & \cdots & t_1^{(n)} \\ t_2 & t_2' & t_2'' & \cdots & t_2^{(n)} \\ \vdots & & & & \vdots \\ t_{n+1} & t_{n+1}' & t_{n+1}'' & \cdots & t_{n+1}^{(n)} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

and note that the determinant of the matrix on the left is $\mathrm{wron}_{R[x],n+1}(t_1, t_2, \ldots, t_{n+1})$, which for simplicity we abbreviate as w. If the collection $t_1, t_2, \ldots, t_{n+1}$ is linearly independent over R_C then $w \neq 0$ by (c), and Cramer's rule then forces $a_0 = a_1 = \cdots = a_n = 0$. Since $a_n \neq 0$, this contradiction proves that $w = 0$, hence by (c) that $t_1, t_2, \ldots, t_{n+1}$ are linearly dependent over K_C. A basis for $\ker(\mathcal{L})$ therefore contains at most n elements. In particular, $\dim_{K_C}(\ker(\mathcal{L})) \leq n$.

The final assertion of (d) is, by (c), equivalent to the assertion that $t_1, t_2, \ldots, t_k$ is a basis of $\ker(\mathcal{L})$ if and only if this collection is linearly independent over R_C. Since a basis is, by definition, linearly independent over the ground field, the forward implication is trivial. Conversely, by elementary linear algebra any collection $t_1, t_2, \ldots, t_k$ within $\ker(\mathcal{L})$ which is linearly independent over R_C can be extended to a basis of $\ker(\mathcal{L})$, and therefore must be a basis since $\dim_{K_C}(\ker(\mathcal{L})) = k$.

(e) The initial assertion is immediate from (b). If $\mathrm{wron}_{R[\delta],n}(r)$ is a unit of R then from the second line in definition (1.17) we see that this linear differential operator will have order n, and that the linear differential operator defined in (i) of (e) is monic. The basis assertion is a special case of (d).

(f) First note by (d) that the collection $r_1, r_2, \ldots, r_n$ must be a basis of $\ker(\mathcal{L})$.

Denote the monic linear differential operator defined in (i) by $\hat{\mathcal{L}}$. Then either $\mathcal{L} = \hat{\mathcal{L}}$, in which case we are done, or $\mathcal{L} \neq \hat{\mathcal{L}}$ and the order of the non-zero linear differential operator $\mathcal{L} - \hat{\mathcal{L}} \in R[\delta]$ is less than n. But this operator admits $r_1, r_2, \ldots, r_n$ as zeros, and the second alternative is therefore impossible by (d) (applied to $\mathcal{L} - \hat{\mathcal{L}}$).

(g) For any $n \times n$ matrix $S = (s_{ij})$ with entries in R one sees from the determinant expansion $\det(S) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_j s_{\sigma(j)j}$ and (1.1) that

$$\begin{aligned}
\delta(\det(S)) &= \textstyle\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \sum_{i=1}^{n} s_{\sigma(1)1} \cdots s_{\sigma(i-1),i-1} \cdot s'_{\sigma(i)i} \cdot s_{\sigma(i+1),i+1} \cdots s_{\sigma(n)n} \\
&= \textstyle\sum_{i=1}^{n} \Big(\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) s_{\sigma(1)1} \cdots s_{\sigma(i-1),i-1} \cdot s'_{\sigma(i)i} \cdot s_{\sigma(i+1),i+1} \cdots s_{\sigma(n),n} \Big) \\
&= \textstyle\sum_{i=1}^{n} \det\left(\begin{bmatrix}
s_{11} & s_{12} & s_{13} & \cdots & s_{1n} \\
\vdots & & & & \vdots \\
s_{i-1,1} & s_{i-1,2} & s_{i-1,3} & \cdots & s_{i-1,n} \\
s'_{i,1} & s'_{i,2} & s'_{i,3} & \cdots & s'_{i,n} \\
s_{i+1,1} & s_{i+1,2} & s_{i+1,3} & \cdots & s_{i+1,n} \\
\vdots & & & & \vdots \\
s_{n1} & s_{n2} & s_{n3} & \cdots & s_{nn}
\end{bmatrix} \right).
\end{aligned}$$

When $S = \operatorname{wron}_{R[\delta],n}(r)$ we see from (1.13) that each of the matrices involved in this sum has two equal rows when $j \neq n$, from which we deduce that

$$\delta(\operatorname{wron}_{R[\delta],n}(r)) = \det\left(\begin{bmatrix}
r_1 & r_2 & r_3 & \cdots & r_n \\
r'_1 & r'_2 & r'_3 & \cdots & r'_n \\
\vdots & & & \vdots & \\
r_1^{(n-2)} & r_2^{(n-2)} & r_3^{(n-2)} & \cdots & r_n^{(n-2)} \\
r_1^{(n)} & r_2^{(n)} & r_3^{(n)} & \cdots & r_n^{(n)}
\end{bmatrix} \right).$$

However, from the solution hypothesis we can replace each $r_j^{(n)}$ in this matrix with $-\sum_k^{n-1} a_k r_j^{(k)}$, and then, using standard properties of determinants, reduce the expression to

$$\delta(\operatorname{wron}_{R[\delta],n}(r)) = \textstyle\sum_k^{n-1} a_k \cdot \det\left(\begin{bmatrix}
r_1 & r_2 & r_3 & \cdots & r_n \\
r'_1 & r'_2 & r'_3 & \cdots & r'_n \\
\vdots & & & \vdots & \\
r_1^{(n-2)} & r_2^{(n-2)} & r_3^{(n-2)} & \cdots & r_n^{(n-2)} \\
r_1^{(k)} & r_2^{(k)} & r_3^{(k)} & \cdots & r_n^{(k)}
\end{bmatrix} \right).$$

The matrices corresponding to $1 \leq k < n-1$ each have two equal rows, hence determinant zero, and the matrix corresponding to $k = n-1$ is $\mathrm{wron}_{R[\delta],n}(r)$. The Abel-Liouville formula follows.

(h) First note, by induction, that

$$(rs)^{(k)} = \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) r^{(k-j)} s^{(k)}$$

for any integer $k \geq 1$. We therefore have

$$\begin{aligned}
\sum_{k=0}^{n} a_k (rs)^{(k)} &= a_0(rs) + \sum_{k=1}^{n} a_k (rs)^{(k)} \\
&= a_0(rs) + \sum_{k=1}^{n} a_k \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) r^{(k-j)} s^{(j)} \\
&= a_0(rs) + \sum_{k=1}^{n} a_k \left[r^{(k)} s + \sum_{j=1}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) r^{(k-j)} s^{(j)} \right] \\
&= \left(\sum_{k=0}^{n} a_k r^{(k)} \right) s + \sum_{k=1}^{n} \sum_{j=1}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) a_k r^{(k-j)} s^{(j)} \\
&= 0 \cdot s + \sum_{j=1}^{n} \left(\sum_{k=j}^{n} \left(\begin{array}{c} k \\ j \end{array} \right) a_k r^{(k-j)} \right) s^{(j)} \\
&= \sum_{j=0}^{n-1} \left(\sum_{k=j+1}^{n} \left(\begin{array}{c} k \\ j+1 \end{array} \right) a_k r^{(k-(j+1))} \right) s^{(j+1)} \\
&= \sum_{j=0}^{n-1} \left(\sum_{k=j+1}^{n} \left(\begin{array}{c} k \\ j+1 \end{array} \right) a_k r^{(k-1-j)} \right) (s')^{(j)},
\end{aligned}$$

and (h) follows.

q.e.d.

The parallels between polynomials and ordinary linear differential operators extend to Galois theory. We assume readers are familiar with the Galois theory of separable polynomials, which we will refer to as "classical Galois theory," but not with that of ordinary linear differential operators, commonly called *differential Galois theory*. To indicate the parallels for these theories, and to give the motivation behind these notes, we offer a quick (and technically incomplete) sketch of the latter theory, always assuming that R is a differential field.

One begins with a monic linear differential operator $\mathcal{L} = \delta^n + \sum_{k=0}^{n-1} r_k \delta^k \in R[\delta]$ of order $n \geq 1$, and, to make the situation challenging, assumes that R does not contain a basis of solutions. To achieve such a basis one constructs an appropriate field T containing R which admits a derivation extending that on R and does contain such a basis. The subfield $S \subset T$ generated by R and this basis, which is the analogue of a splitting field of a polynomial, is called a *Picard-Vessiot extension* for $\mathcal{L}$. The *differential Galois group* of $\mathcal{L}$ is defined to be the group of automorphisms of S over R which commute with the (extended) derivation on S. Thus far, the parallels with classical Galois theory are clear[7], but we will not venture, at least temporarily, further from the shore.

It is well-known that the n^{th}-order equation $\mathcal{L} = 0$ is equivalent to the first-order system

$$x' = Ax \tag{1.19}$$

of ordinary linear differential equations, where

$$A := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -r_0 & -r_1 & \cdots & -r_{n-2} & -r_{n-1} \end{bmatrix} \tag{1.20}$$

is the transpose of the companion matrix of the "obvious" polynomial $x^n + \sum_{k=1}^{n-1} r_k x^k$ which one would associate with $\mathcal{L}$. One might therefore ask: is there a differential Galois theory for arbitrary first-order linear systems as in (1.19), but with the $n \times n$ matrix A not necessarily of the form (1.20)? The answer is yes, and the generalization involves practically no additional work, but the analogies with classical Galois theory begin to blur.

Things become even murkier when one realizes that when $P \in \mathrm{GL}(n, R)$ and one substitutes $x = Py$ in (1.19), obtaining

$$y' = By, \qquad \text{where} \qquad B := P^{-1}AP - P^{-1}P, \tag{1.21}$$

precisely the same Picard-Vessiot extension and Galois group result. In fancier language: *the differential Galois group is invariant under "gauge transformations,"* i.e.,

[7]But there are major differences, e.g., the differential Galois group can be infinite, but always has a faithful representation as a linear algebraic group.

under the action of $\mathrm{GL}(n,R)$ on the space of $n \times n$ matrices A (with entries in R) defined by $P \cdot A := P^{-1}AP - P^{-1}P'$. This seems to have no analogue in classical Galois theory, but we will find, by reformulating the classical theory, that this is not the case.

What this invariance suggests is that the differential Galois group should not be associated with first-order systems, but with some other entity which admits first-order systems as "coordinate" representations. And this works out beautifully.

Assuming V is an n-dimensional vector space over R, define a *differential structure* on V to be a mapping $D : V \to V$ such that the following two properties hold for all $v_1, v_2, v \in V$ and all $r \in R$:

- ("additivity") $D(v_1 + v_2) = Dv_1 + Dv_2$ and
- ("the Leibniz rule") $D(rv) = rDv + r'v$.

When $D : V \to V$ is a differential structure and V^* denotes the dual space of V a corresponding differential structure $D^* : V^* \to V^*$ is defined by

$$D^*v^*(v) := (v^*(v))' - v^*(Dv), \qquad v^* \in V^*, v \in V, \tag{1.22}$$

as the reader can easily check. This is the *dual* (*differential*) *structure* of D, or the (*differential*) *structure dual to* D. If for $w^* \in V^*$ and $v \in V$ we write $w^*v \in R$ as $\langle v, w^* \rangle$, then (1.22) can be expressed as the classical Lagrange identity

$$\langle v, v^* \rangle' = \langle Dv, v^* \rangle + \langle v, D^*v^* \rangle, \tag{1.23}$$

which has a well-known counterpart in differential geometry.

When $D : V \to V$ is a differential structure a vector $v \in V$ is called[8] *horizontal*, or *D-horizontal* when D needs clarification, if $Dv = 0$. When $\mathbf{e} = (e_j)_{j=1}^n$ is a(n ordered) basis of V one defines an $n \times n$ matrix $A = (a_{ij})$ with entries in R by

$$De_j := \textstyle\sum_{i=1}^n -a_{ij}e_i, \qquad j = 1, 2, \ldots, n, \tag{1.24}$$

and one then has

$$Dv = w \qquad \Leftrightarrow \qquad v_{\mathbf{e}}' - Av_{\mathbf{e}} = w_{\mathbf{e}}, \tag{1.25}$$

[8]The terminology is from differential geometry: our differential structures lie somewhere between first order systems of linear differential equations and the "connections" one learns about in that subject.

where $u_{\mathbf{e}}$ denotes the column vector of coefficients u_j of a typical vector $u = \sum_{j=1}^{n} u_j e_j \in V$ and $u'_{\mathbf{e}}$ is the column vector with entries u'_j. The search for solutions for

$$x' = Ax \tag{1.26}$$

is now seen to be a search, by means of a choice of basis, for horizontal vectors. We refer to the matrix A in (1.24) as the $\mathbf{e}$-*matrix* of D, and to (1.26) as the $\mathbf{e}$-*basis representation* of D. When $\mathbf{e}$ is clear from context, or specifying $\mathbf{e}$ is not essential, we simply refer to A as the (*corresponding*) *matrix* of D, and to (1.26) as the (*corresponding*) *basis representation* of D. When A is the $\mathbf{e}$-matrix of D, and $\mathbf{e}^*$ is the dual basis V^*, one checks that the $\mathbf{e}^*$-matrix of the dual structure $D^* : V^* \to V^*$ to D is $-A^\tau$, where the Greek letter τ (tau) denotes transposition. The $\mathbf{e}^*$-basis representation of D^* is therefore given by the so-called "adjoint equation"

$$x' = -A^\tau x \tag{1.27}$$

of (1.26).

We also use the "$\mathbf{e}$-matrix" terminology with linear operators $T : V \to V$: if $\mathbf{e}$ is a basis of V the $\mathbf{e}$-*matrix* of T is the $n \times n$ matrix $A = (a_{ij})$ with entries in R defined by

$$Te_j = \textstyle\sum_{j=1}^{n} a_{ij} e_i, \qquad j = 1, 2, \ldots, n. \tag{1.28}$$

As one learns in elementary linear algebra (although perhaps with different notation and terminology), T and A are related by

$$Tv = w \qquad \Leftrightarrow \qquad (Tv)_{\mathbf{e}} = Av_{\mathbf{e}} = w_{\mathbf{e}} \tag{1.29}$$

for all vector $v, w \in V$. The equivalence seen in (1.25) is the direct analogue for differential structures.

When phrased in terms of differential structures an extension $S \supset R$ of differential fields[9] is a *Picard-Vessiot extension* for D if the following three conditions hold:

(I) ("no new constants") $S_C = R_C$;

(II) the S-vector space[10] $V_S := S \otimes_R V$ admits a basis consisting of horizontal vectors for the unique differential structure[11] $D_S : \sum_k s_j \otimes v_j \in V_S \mapsto \sum_k (s_j \otimes Dv_j + s_j' \otimes v_j) \in V_S$ extending D.

(III) ("minimality") If $T \supset R$ is any differential field extension satisfying the first two items there is a differential embedding $\eta : S \to T$ over[12] R.

Assuming such an extension exists[13] we define the *differential Galois group* of D exactly as before, i.e., as the group of automorphisms of S over R which commute with the derivation on S.

Assuming $S \supset R$ is a Picard-Vessiot extension for D and G is the associated differential Galois group, define a representation $\rho : G \to \mathrm{GL}(V_S, S)$, which we often express as an action, by

$$\rho(g)(s \otimes v) = g \cdot (s \otimes v) := (g \cdot s) \otimes v, \qquad s \otimes v \in V_S. \tag{1.30}$$

Theorem 1.31 : *For any $g \in G$ one has*

$$\rho(g) \circ D_S = D_S \circ \rho(g). \tag{i}$$

In fancier language: D_S is equivariant w.r.t. the given G-action on V_S.

---

[9]The definition of a differential field extension requires that the derivation on S restricts to that on R.

[10]Those not familiar with tensor products can think of V_S as what results by choosing a basis $\mathbf{e}$ for V and then allowing vector coefficients to be in S. In particular, one can regard V as a subspace of V_S (when the latter is viewed as an R-space), and the basis $\mathbf{e}$ of the R-space V as a basis of the S-space V_S, which has the important consequence that $\dim_S(V_S) = \dim_R(V)$. The tensor product achieves such results in a purely geometric way, i.e., without reference to bases.

[11]Since $D : V \to V$ is not R-linear, one must do a bit of work to see that this function is well-defined. See [C, §8, Proposition 8.4].

[12]That is, a field embedding $\eta : S \to T$ which fixes R pointwise and satisfies $\delta_T \circ \eta = \eta \circ \delta_S$, where δ_T and δ_S are the respective derivations on T and S extending the given derivation on R.

[13]Which is not guaranteed. For a general existence theorem one needs assume the characteristic of R is zero and that the field of constants R_C is algebraically closed. See, e.g., [vdP-S].

Proof : For any $s \otimes v \in V_S$ one has

$$\begin{aligned}(\rho(g) \circ D_S)(s \otimes v) &= g \cdot (s \otimes Dv + s' \otimes v) \\ &= (g \cdot s) \otimes Dv + g \cdot (s' \otimes v) \\ &= (g \cdot s) \otimes Dv + (g \cdot s') \otimes v \\ &= (g \cdot s) \otimes Dv + (g \cdot s)' \otimes v \\ &= D_S(g \cdot (s \otimes v)) \\ &= (D_S \circ \rho(g))(s \otimes v).\end{aligned}$$

q.e.d.

Corollary 1.32 : *G permutes the horizontal vectors of V_S.*

When $S \supset R$ is a Picard-Vessiot extension for a differential structure $D : V \to V$ a matrix $\alpha = (\alpha_{ij}) \in \mathrm{GL}(n, S)$ is called a *fundamental matrix* for D if there is a basis $\mathbf{e}$ of V such that α and the $\mathbf{e}$-matrix A of D are related by

$$\alpha' = A\alpha, \tag{1.33}$$

where

$$\alpha' := (\alpha'_{ij}), \tag{1.34}$$

Any such α is also called a *fundamental matrix* for the $\mathbf{e}$-basis representation $x' = Ax$ of D. Indeed, both α and the columns α_j of this matrix satisfy this last system of equations. When we need to make $\mathbf{e}$ explicit we refer to such an α as a *fundamental* $\mathbf{e}$*-matrix* for D.

Proposition 1.35 : *When $D : V \to V$ is a differential structure and $S \supset R$ is a Picard-Vessiot extension, the field S is generated by R and any fundamental matrix for D.*

Proof : See [C, Proposition 9.3(c_2)].

Let $D : V \to V$ and $S \supset R$ be as in Proposition 1.35. Define an action of the associated differential Galois group G of D on the collection of $n \times n$ matrices with entries in S by

$$g \cdot (s_{ij}) := (g \cdot s_{ij}). \tag{1.36}$$

Now pick a basis $\mathbf{e}$ for V, let

$$x' = Ax \tag{1.37}$$

be the associated $\mathbf{e}$-matrix representation of D, and let $\alpha \in \mathrm{GL}(n, S)$ be a fundamental matrix of D.

Theorem 1.38 : *Assume the notation of the previous paragraph and let* $\alpha \in \mathrm{GL}(n, S)$ *be a fundamental* $\mathbf{e}$*-matrix for* D*. Then:*

(a) *for any fundamental* $\mathbf{e}$*-matrix for* D *one has* $\alpha^{-1}\beta \in \mathrm{GL}(n, R_C)$*;*

(b) *for any* $g \in G$ *the matrix* $g \cdot \alpha$ *is also a fundamental* $\mathbf{e}$*-matrix for* D*; and*

(c) *the mapping* $\rho : g \in G \mapsto \alpha^{-1}(g \cdot \alpha) \in \mathrm{GL}(n, R_C)$ *is a faithful matrix representation.*

In the final assertion $\alpha^{-1}(g \cdot \alpha)$ denotes the product of the matrices α^{-1} and $g \cdot \alpha$.

Proof : For (a) and (b) see, e.g., [C]. Assertion (c) is implicit in [Lev], and the proof is fairly straightforward. (This author has worked through the details, but is not aware of a published reference.) **q.e.d**.

The goal of these notes is to formulate and prove analogues of Theorem 1.38(b) and (c) for classical Galois theory. For (b) the goal is achieved in Proposition 9.3; for (c) in Theorem 9.16 and Corollary 9.18. The obvious analogue of Theorem 1.38(a) fails with our approach (see Example 6.4(a)), but this causes little trouble[14].

The basic idea is *not* to associate the classical Galois extension with a separable polynomial, but rather with a linear operator having that polynomial as characteristic polynomial. The $\mathbf{e}$-matrices for a differential structure then correspond to the $\mathbf{e}$-matrices of a linear operator. We develop the analogue of a fundamental matrix, and use these matrices as in Theorem 1.38 to define a faithful matrix representation of the Galois group. As the reader might expect, the representing matrices are nothing but permutation matrices, and in that regard we obtain nothing new. In fact computing the group by classical methods is probably much easier, but the approach herein makes the connections with differential Galois theory far more transparent.

[14]The mapping γ defined in (9.10) enables one to circumvent any difficulties.

Fundamental to our approach is the notion of a "cyclic vector." For a linear operator $T : V \to V$ on an n-dimensional space a vector $v \in V$ is *cyclic* if the collection $v, Tv, T^2v, \ldots, T^{n-1}v$ is a basis. By replacing T by D, one obtains the definition for a differential structure. For linear operators such vectors are not to be expected, but for differential structures they are plentiful: their existence is guaranteed under very mild hypotheses (see, e.g., [C-K]). In the differential case such vectors for dual structures $D^* : V^* \to V^*$ guarantee that first-order systems can be expressed as n^{th}-order equations, which is of great practical importance. One can say a bit more (as will be seen in Corollary 1.41).

Proposition 1.39 : *When $D : V \to V$ is a differential structure on an R-space V of dimension n the following statements are equivalent.*

(a) *$D^* : V^* \to V^*$ admits a cyclic vector.*

(b) *There is a basis $\mathbf{e}$ of V such that the $\mathbf{e}$-matrix of D has the form*

(i)
$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-1} & -a_n \end{bmatrix}.$$

Proof : Let $\mathbf{e}$ be a basis of V and let $\mathbf{e}^* = (e_j^*)_{j=1}^n$ be the (corresponding) dual basis of V^*. Then from the discussion leading to (1.27) we see that the $\mathbf{e}$-matrix of D has the form seen in (i) if and only if the $\mathbf{e}^*$-matrix of D^* has the form

(ii)
$$B := \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & a_0 \\ -1 & 0 & 0 & \cdots & 0 & a_1 \\ 0 & -1 & 0 & \cdots & 0 & a_2 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & -1 & 0 & a_{n-1} \\ 0 & \cdots & 0 & 0 & -1 & a_n \end{bmatrix}.$$

and from (1.24) we see that the $\mathbf{e}^*$-matrix of D^* has this form if and only if e_1^* is a cyclic vector. **q.e.d.**

Proposition 1.40 : *Let* $D : V \to V$ *be a differential structure, let* $S \supset R$ *be a Picard-Vessiot extension for* D*, and let* A *be as in* (i) *of* Proposition 1.39. *Choose any elements* $s_1, s_2, \ldots, s_n \in S$*, and set* $s := [s_1 \ s_2 \ \cdots \ s_n]^\tau$*. Then*

$$\text{(i)} \quad s' = As \qquad \Leftrightarrow \qquad \begin{cases} s = \begin{bmatrix} s_1 \\ s_1' \\ \vdots \\ s_1^{(n-1)} \end{bmatrix} \\ \textit{and} \\ s_1 \ \textit{is a solution of the linear differential equation} \\ \quad y^{(n)} + \sum_{k=0}^{n-1} a_k y^{(k)} = 0. \end{cases}$$

Proof : From

$$As = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-1} & -a_n \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{n-1} \\ s_n \end{bmatrix} = \begin{bmatrix} s_2 \\ s_3 \\ \vdots \\ s_n \\ -\sum_{j=0}^{n} a_j s_{j+1} \end{bmatrix}$$

one sees that

$$\begin{cases} \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{cases} s' = As \quad \Leftrightarrow \quad \begin{cases} s_2 = s_1' \\ s_3 = s_2' = s_1'' \\ \vdots \\ s_n = s_{n-1}' = s_1^{(n-1)} \\ \text{and} \\ s_n' = s_1^{(n)} = -\sum_{j=0}^{n-1} a_j s_1^{(j)} \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} s_1 \text{ is a solution of } y^{(n)} + \sum_{j=0}^{n-1} a_n y^{(j)} \text{ and} \\ s = \begin{bmatrix} s_1 \\ s_1' \\ \vdots \\ s_1^{(n-1)} \end{bmatrix}, \end{cases}$$

precisely as claimed. **q.e.d.**

Corollary 1.41 : *Suppose* D^* *admits a cyclic vector,* $s_1, s_2, \ldots, s_n \in S$ *are linearly independent over* R_C*, and* $s := (s_1, s_2, \ldots, s_n) \in S^n$*. Then the following assertions are equivalent.*

(a) *The Wronski matrix* $\mathrm{wrm}_{R[x],n}(s)$ *is a fundamental matrix of* D*.*

(b) *The elements* $s_1, s_2, \ldots, s_n \in S$ *form a basis of solutions for the* n^{th}*-order linear differential equation* $\sum_{k=0}^{n} a_k y^{(k)} = 0$*, where the* a_k *are as in* (i) *of* Proposition 1.40*.*

Proof :

(a) $\Rightarrow$ (b) : When (a) holds each column s of $\mathrm{wrm}_{R[x],n}(r)$ satisfies $s' = As$, and (b) then follows from Propositions 1.40 and 1.18(c).

(b) $\Rightarrow$ (a) : If (b) holds we see from Proposition 1.40 that $(\mathrm{wrm}_{R[x],n}(r))' = A \cdot \mathrm{wrm}_{R[x],n}(r)$, and from Proposition 1.18(c) that $\mathrm{wrm}_{R[x],n}(r) \in \mathrm{GL}(n, R)$.

q.e.d.

If the notes we will develop the analogue for a class of operators $T : V \to V$ having duals $T^* : V^* \to V^*$ which admit cyclic vectors.

2. Preliminaries on Cyclic Vectors

In this section $T : V \to V$ *is a* K*-linear operator.*

This section is a self-contained introduction (modulo several proofs) to cyclic vectors for linear operators. For ease of reference some of the work in the introduction will be repeated.

A vector $v \in V$ is *cyclic* (w.r.t. T) if $(v, Tv, T^2v, \ldots, T^{n-1}v)$ is a basis of V. When such a v exists we say that T is *cyclic*, or that T *admits a cyclic vector*. The following observation will simplify the presentation of examples.

Proposition 2.1 : *The following assertions are equivalent:*

(a) T *is cyclic;*

(b) *there is a basis* $\mathbf{e}$ *of* V *such that the* $\mathbf{e}$*-matrix of* T *has the companion matrix form*

(i)
$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -k_0 \\ 1 & 0 & 0 & & 0 & -k_1 \\ 0 & 1 & 0 & \ddots & \vdots & -k_2 \\ & \vdots & \ddots & & & \\ 0 & 0 & & 1 & 0 & -k_{n-2} \\ 0 & 0 & \cdots & 0 & 1 & -k_{n-1} \end{bmatrix};$$

(c) *the characteristic and minimal polynomials of* T *are identical; and*

(d) *the minimal polynomial of* T *has degree* $n = \dim_K(V)$.

Moreover, when any (and therefore all) *of these assertions holds the characteristic and minimal polynomials of* T *are both given by*

(ii) $$\mathrm{char}_{T,K}(x) = \mathrm{min}_{T,K}(x) = x^n + \textstyle\sum_{j=0}^{n-1} k_j x^j,$$

where the k_j *are as in* (i). *In particular, the companion matrix in* (i) *is the unique such matrix representation of* T.

Note that assertions (a), (c) and (d) are independent of bases. We view assertion (a) as geometric; (c) and (d) as algebraic. In the statements of (c) and (d) the polynomials are understood to be in $K[x]$.

Also note that the matrix appearing in (i) is the rational (canonical) form of T, and recall from elementary linear algebra that the rational form of a linear operator is unique.

Proof :

(a) $\Leftrightarrow$ (b) : Obvious.

(b) $\Rightarrow$ (c) : See, e.g., [M-B, Chapter IX, §6, Proposition 13, p. 315].

(c) $\Rightarrow$ (b) : See, e.g., [M-B, Chapter X, §4, Theorem 8 and Corollary 1, p. 351].

(c) $\Rightarrow$ (d) : Obvious.

(d) $\Rightarrow$ (c) : The characteristic polynomial has degree $n = \dim_K(V)$, and both polynomials are monic. If they do not agree, subtraction would produce a polynomial of lower degree satisfied by T, thereby contradicting minimality.

The final assertions are easy consequences of (i).

q.e.d.

Examples 2.2 : In these examples we take $K = \mathbb{Q}$, $V = K^n$, where $n = 2, 3$ or 4, we let $\mathbf{e} = (e_j)_{j=1}^n$ be the usual basis, and we let $T : V \to V$ denote the K-linear operator with $\mathbf{e}$-matrix A.

(a) Take $A = \begin{bmatrix} 1 & -2 \\ 1 & 4 \end{bmatrix}$. The rational form is $\begin{bmatrix} 0 & -6 \\ 1 & 5 \end{bmatrix}$, and from Proposition 2.1 we conclude that T admits a cyclic vector. In fact $e_1 \in \mathbb{Q}^2$ is such a vector, as the reader can easily check, but there are many more, e.g., e_2 and $e_1 + e_2$.

(b) Let $A := \begin{bmatrix} -3 & 5 & -1 \\ 10 & -15 & 3 \\ 61 & -89 & 18 \end{bmatrix}$. Then $\mathrm{char}_{T,\mathbb{Q}}(x) = \min_{T,\mathbb{Q}}(x) = x^3 - x + 1$, and we conclude from Proposition 2.1(c) (or (d)) that T is cyclic. In this case $e_1 \in \mathbb{Q}^3$ is a cyclic vector and, once again, there are many more.

(c) The rational form of $A := \begin{bmatrix} -35 & -7 & -19 & -3 \\ 31 & -35 & -34 & -10 \\ 216 & 8 & 75 & 8 \\ -957 & 82 & -186 & 1 \end{bmatrix}$ is $\begin{bmatrix} 0 & 0 & 3 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$.
Since this rational form is not a companion matrix, T is not cyclic (Proposi-

tion 2.1(b)). This conclusion can also be seen from the fact that the minimal polynomial $x^3 - 3x^2 + x - 3$ does not have degree $4 = \dim_{\mathbb{Q}}(\mathbb{Q}^4)$ (Proposition 2.1(d)).

Proposition 2.3 : *A linear operator $T : V \to V$ is cyclic if and only if the dual operator $T^* : V^* \to V^*$ is cyclic.*

By the *dual operator* of T we mean the linear operator on V^* defined by $v^* \mapsto (v \in V \mapsto v^*(Tv)) \in V^*$. The concept (perhaps by some other name) is assumed familiar, as is the fact that when $\mathbf{e}$ is a basis of V, and $\mathbf{e}^*$ is the dual basis of V^*, the $\mathbf{e}$-matrix of T is A if and only if the $\mathbf{e}^*$-matrix of T^* is A^τ.

We give two proofs of this proposition. The first views the result a corollary of Proposition 2.1; the second motivates our proof of the corresponding result for differential structures (Proposition 2.7).

First Proof : One needs the fact that A and A^τ are similar when A is a square matrix. It is clearly enough to prove this when A is an elementary Jordan block, and in that case one sees that for $P := (\delta_{n+1-i,j})$ one has $P^{-1}AP = A^\tau$.

Now recall the well-known results that similar matrices have the same characteristic polynomial and the same minimal polynomial[15]. The corollary is now immediate from Proposition 2.1(c) (or (d)). **q.e.d.**

The value $v^*(v)$ of a linear functional $v^* \in V^*$ on a vector $v \in V$ is often more conveniently expressed as $\langle v, v^* \rangle$, i.e.,

$$\langle v, v^* \rangle := v^*(v). \tag{2.4}$$

The relationship between $T : V \to V$ and the dual operator can then be written

$$\langle Tv, v^* \rangle = \langle v, T^* v^* \rangle, \tag{2.5}$$

whereupon by induction one immediately sees that

$$\langle T^k v, v^* \rangle = \langle v, (T^*)^k v^* \rangle, \qquad k = 1, 2, 3, \ldots \ . \tag{2.6}$$

[15]See, e.g., [M-B, Chapter IX, §5, Theorem 11, p. 312 and §2, Corollary 2, p. 314].

Second Proof of Proposition 2.3 :

$\Rightarrow$: Let $e_1 \in V$ be cyclic w.r.t. T and let $\mathbf{e} = (e_j)_{j=1}^n$ be the corresponding basis of V, i.e., $e_j := T^{j-1}e_1$ for $j = 1, 2 \ldots, n$. Denote the dual basis of V^* by $(e^*)_{j=1}^n$. Then for any $1 \leq i, j \leq n$ we have

$$\begin{aligned} \langle e_i, T^* e_j^* \rangle &= \langle T e_i, e_j^* \rangle \\ &= \langle e_{i+1}, e_j^* \rangle \\ &= \delta_{i+1,j} \\ &= \delta_{i,j-1} \\ &= \langle e_i, e_{j-1}^* \rangle , \end{aligned}$$

from which we see[16] that $T^* e_j^* = e_{j-1}^*$. The vector e_n^* is therefore cyclic for T^*.

$\Leftarrow$ The argument is completely analogous to that for the forward implication.

q.e.d.

Now suppose, until the end of the proof of Proposition 2.7, that K is a differential field with derivation $\delta : k \in K \mapsto k' \in K$, W is an n-dimensional K-space, and $\hat{D} : W \to W$ is a differential structure on W. A vector $w \in W$ is *cyclic* w.r.t. $\hat{D}$ if $w, \hat{D}w, \hat{D}^2 w, \ldots, \hat{D}^{n-1} w$ is a basis of W. $\hat{D}$ is *cyclic* when such a vector exists.

Proposition 2.7 : *A differential structure $D : V \to V$ is cyclic if and only if the dual structure $D^* : V^* \to V^*$ is cyclic.*

The dual structure D^* was defined in (1.22).

Proof :

$\Rightarrow$ For the proof it is useful to express the Lagrange identity (1.23) in the form

$$\text{(i)} \qquad \langle v, D^* v^* \rangle = \langle v, v^* \rangle' - \langle Dv, v^* \rangle.$$

Let e_1 be cyclic for D, let $(e_i := D^{i-1} e_1)_{i=1}^n$ be the corresponding basis, and let $(e_j^*)_{j=1}^n$ be the dual basis of V^*. Then for any $1 \leq i, j \leq n$ we see from (i) that

$$\begin{aligned} \langle e_i, D^* e_j^* \rangle &= \langle e_i, e_j \rangle' - \langle D e_i, e_j^* \rangle \\ &= \delta_{ij}' - \langle e_{i+1}, e_j^* \rangle \\ &= 0 - \delta_{i+1,j} \\ &= -\delta_{i,j-1} \\ &= \langle e_i, -e_{j-1}^* \rangle , \end{aligned}$$

[16] From the fact that that a linear functional (or, for that matter, any linear transformation) is completely determined by its values on a basis.

and as a result that $D^*e_j^* = -e_{j-1}^*$. The vector e_n^* is therefore cyclic for D^*.

$\Leftarrow$ The argument is completely analogous to that for the forward implication.

q.e.d.

We now return to the purely linear algebraic context. We will show that the existence of one cyclic vector for a linear operator implies the existence of many more[17]. We formulate the result assuming $K = \mathbb{R}$, so as to be able to introduce the norm topology, but readers familiar with algebraic geometry will immediately see from the proof that the result holds for any K when Zariski topology is assumed on V.

Proposition 2.8 : *Assume $K = \mathbb{R}$ and endow V with the norm topology[18]. Then the following assertions are equivalent:*

(a) *T is cyclic; and*

(b) *the collection of cyclic vectors for T forms a dense open subset of V.*

Proof :

(a) $\Rightarrow$ (b): Choose any basis $\mathbf{e} = (e_j)_{j=1}^n$ of V and let A denote the $\mathbf{e}$-matrix of T. For any $v = \sum_{j=1}^n k_j e_j$ let $v_{\mathbf{e}}$ denote the column vector[19] $[k_1 \;\; k_2 \;\; \cdots \;\; k_n]^\tau$. Then v is cyclic if and only if

$$\det(\begin{bmatrix} v_e & Av_e & \cdots & A^{n-1}v_e \end{bmatrix}) \neq 0.$$

Now simply observe that the hypersurface of $\mathbb{R}^n \simeq V$ where the polynomial function

$$x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \simeq V \mapsto \det(\begin{bmatrix} x^\tau & Ax^\tau & \cdots & A^{n-1}x^\tau \end{bmatrix})$$

vanishes is a closed nowhere dense subset.

(b) $\Rightarrow$ (a) : Obvious.

q.e.d.

---

[17] The analogous result holds for differential structures, but will not be established here. See, e.g., [C-K].

[18] By the finite-dimensionality assumption there is only one such topology.

[19] The tau (i.e., τ) denotes transposition.

A K-linear operator $S : V \to V$ is[20] *equivalent* to T if there is a K-linear isomorphism $Q : V \to V$ such that $S = Q^{-1} \circ T \circ Q$. For example, let S, T and Q be the linear operators on $\mathbb{R}^2$ with usual basis matrices

$$m_S := \begin{bmatrix} 6 & 8 \\ -4 & -6 \end{bmatrix}, \qquad m_T := \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \qquad \text{and} \qquad m_Q := \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

respectively. Then one sees directly from $m_S := m_{Q^{-1}} \cdot m_T \cdot m_Q$ that S and T are equivalent. Note that:

- equivalent linear operators have the same characteristic polynomial; that
- if S is a linear operator on V which is equivalent to T, then T admits a cyclic vector if and only if this is the case for S. (Indeed, $v \in V$ is cyclic for T if and only if $Q^{-1}v$ is cyclic for S.); and that
- "equivalence" is an equivalence relation.

Corollary 2.9 : *Suppose T admits a cyclic vector, $S : V \to V$ is a K-linear operator with the same property[21], and T and S have the same characteristic polynomial. Then T and S are equivalent.*

Proof : By Proposition 2.1 there are (ordered) bases $\mathbf{e} = (e_j)_{j=1}^n$ and $\hat{\mathbf{e}} = (\hat{e}_j)_{j=1}^n$ of V such that the $\mathbf{e}$ and $\hat{\mathbf{e}}$-matrices of T and S are the companion matrices of the associated characteristic polynomials; hence are identical by hypothesis. The K-linear mapping $Q : V \to V$ uniquely determined by assigning e_j to $\hat{e}_j$ will then be an equivalence as asserted. **q.e.d.**

Corollary 2.10 : *A bijective correspondence between equivalence classes $[L]$ of K-linear operators $L : V \to V$ admitting cyclic vectors and monic degree n polynomials in $K[x]$ is well-defined by assigning $[L]$ to the characteristic polynomial of L.*

Suppose $L \supset K$ is a field extension. Then an L-vector space structure is well-defined on the tensor product $L \otimes_K V$ by $\ell \cdot (m \otimes v) := (\ell m) \otimes v$ for any $\ell, m \in L$ and any $v \in V$, and an L-linear operator $T_L : L \otimes_K V \to L \otimes_K V$ is then given by

$$T_L := \mathrm{id}_L \otimes_K T, \qquad \text{i.e.,} \qquad T_L : \textstyle\sum_j \ell_j \otimes v_j \mapsto \sum_j \ell_j \otimes T v_j. \tag{2.11}$$

[20] One is tempted to say "similar" rather than "equivalent," but similarity is generally associated with matrices rather than with linear operators.

[21] The cyclic vector for S need not be the same as that for T.

One regards V as a K-subspace of $L \otimes_K V$ by means of the embedding $v \in V \mapsto 1 \otimes v \in L \otimes_K V$. In particular, for any such v one makes the identifications

$$Tv \simeq 1 \otimes Tv = (\mathrm{id} \otimes T)(1 \otimes v) = T_L(1 \otimes v) \simeq T_L v, \tag{2.12}$$

and thereby views T as the restriction of T_L to V or, equivalently, T_L as an extension of T to $L \otimes_K V$. If W is an L-space isomorphic to $L \otimes_K V$, and if $U : W \to W$ is a linear operator equivalent to T_L in the sense that the diagram

$$\begin{array}{ccc} W & \xrightarrow{U} & W \\ \simeq \,| & & |\, \simeq \\ V_L & \xrightarrow{T_L} & V_L \end{array} \tag{2.13}$$

commutes, then one says that T *ascends* to U and that U *descends* to T. The ascension process is often described as *extending the base*, which might be better described in this context as "extending the base field from K to L."

Proposition 2.14 : *Assuming the notation of the previous paragraph, the following assertions hold.*

(a) *Any basis* $\mathbf{e} = (e_j)_{j=1}^n$ *of* V *over* K *can be considered as a basis for* $L \otimes_K V$ *over* L *by means of the embedding* $v \mapsto 1 \otimes v$. *In particular,*

(i) $$\dim_L(L \otimes_K V) = \dim_K(V).$$

(b) *Let* $\mathbf{e}$ *be a basis of* V *and let* A *be the* $\mathbf{e}$*-matrix of* T. *Then* A *is also the* $\mathbf{e}$*-matrix of* T_L *when* $\mathbf{e}$ *is regarded* (as in (a)) *as a basis of* $L \otimes_K V$.

(c) T *and* T_L *have the same characteristic polynomial, i.e., when both are considered as polynomials in* $L[x]$ *one has* $\mathrm{char}_{T,K}(x) = \mathrm{char}_{T_L,L}(x)$.

(d) T *and* T_L *have the same rational form.*

(e) T *is cyclic if and only if* T_L *is cyclic.*

Proof :

(a) This is standard. See, e.g., [M-B, Chapter IX, §8, Proposition 16, p. 322] for a more general result.

(b) The $\mathbf{e}$-matrix of T is $A = (a_{ij})$, where

$$Te_j = \textstyle\sum_i a_{ij} e_i, \qquad j = 1, 2, \ldots, n.$$

Using the identifications surrounding (2.12) we therefore have

$$T_L(1 \otimes e_j) = 1 \otimes Te_j = 1 \otimes \sum_i a_{ij} e_i = \sum_i (a_{ij} \otimes e_i) = \sum_i a_{ij}(1 \otimes e_i),$$

and the result follows.

(c) Immediate from (b), since the characteristic polynomial can be computed using any matrix representation of T.

(d) Since the rational form of T is unique, and since any matrix with entries in K is also a matrix with entries in L, it follows from (b) that the rational form for T must be the rational form for T_L.

(e) Since the cyclic property can be determined from the rational form, this follows from (d).

q.e.d.

3. Preliminaries on Diagonalizability

Again $T : V \to V$ *denotes a* K*-linear operator.*

A subspace U of a K-space Y will be called *non-trivial* if the inclusion $U \subset Y$ is proper and $U \neq 0$.

A subspace $W \subset V$ is T*-invariant* if $T(W) \subset W$.

Lemma 3.1 : *When* $W \subset V$ *is a non-trivial* T*-invariant subspace the minimal polynomial in* $K[x]$ *of* $T|_W$ *divides the minimal polynomial in* $K[x]$ *of* T*.*

Proof : Let $m, m_W \in K[x]$ denote the minimal polynomials of T and $T|_W$ respectively; $m_W \neq 0$ by the non-triviality assumption on W. From $m(T) = 0$ we have $m(T)|_W = 0$, from T-invariance we have $m(T)|_W = m(T|_W)$, and $m(T|_W) = 0$ follows. Since $m_W(T|_W)$ also vanishes, and since m_W generates the principal ideal of $K[x]$ consisting of those polynomials vanishing on $T|_W$, the result follows. **q.e.d.**

The operator T is:

- *reducible* if there is a non-trivial T-invariant subspace;
- *irreducible* if it is not reducible;
- *completely reducible* if V is the direct sum of non-trivial T-invariant irreducible subspaces.

Theorem 3.2 : *The following assertions are equivalent:*

(a) T *is irreducible;*

(b) T *is cyclic and the minimal polynomial in* $K[x]$ *of* T *is irreducible; and*

(c) *the minimal polynomial in* $K[x]$ *of* T *is irreducible of degree* n*.*

Proof :

(a) $\Leftrightarrow$ (b) : See, e.g., [J, Chapter III, §7, Theorem 3, p. 128].

(b) $\Leftrightarrow$ (c) : By Proposition 2.1(d).

q.e.d.

Theorem 3.3 : *The following assertions are equivalent:*

(a) T *is completely reducible; and*

(b) *the minimal polynomial in* $K[x]$ *of* T *is a product of distinct irreducible polynomials in* $K[x]$.

Proof : See, e.g., [J, Chapter III, §7, Theorem 5, p. 129]. **q.e.d.**

The operator T is *diagonalizable* if there is a basis $\mathbf{e}$ of V such that the $\mathbf{e}$-matrix of T is diagonal.

Corollary 3.4 : *The following assertions are equivalent:*

(a) T *is diagonalizable; and*

(b) *the minimal polynomial in* $K[x]$ *of* T *factors into distinct monic linear polynomials in* $K[x]$.

Proof :

(a) $\Rightarrow$ (b) : Diagonalizability obviously implies complete reducibility, hence by Theorem 3.3 the minimal polynomial must be a product of distinct irreducible polynomials. On the other hand, by direct calculation using the diagonal matrix form the characteristic polynomial is seen to be a product of monic linear polynomials. Since the minimal polynomial divides the characteristic polynomial, (b) follows.

(b) $\Rightarrow$ (a) : By Theorem 3.3 T is completely reducible, i.e., $V = V_1 \oplus V_2 \oplus \cdots \oplus V_t$, where each V_j is T-invariant and irreducible w.r.t. $T|_{V_j}$. Fix any such j. Then by Theorem 3.2 the minimal polynomial in $K[x]$ of $T|_{V_j}$ must be irreducible. By Lemma 3.1 that polynomial must divide the minimal polynomial of T, hence must be linear by the factorization hypothesis. Proposition 2.1(d) then gives $\dim_K(V_j) = 1$, and (a) follows.

q.e.d.

Corollary 3.5 : *Suppose* $L \supset K$ *is an extension of fields and* $T_L : L \otimes_K V \to L \otimes_K V$ *is the unique* L*-linear operator extending* T. *Then the following statements are equivalent:*

(a) $L \otimes_K V$ *admits a basis consisting of eigenvectors of* T_L;

(b) T_L *is diagonalizable; and*

(c) *the minimal polynomial in* $K[x]$ *of* T *is separable, and* L *contains a complete set of roots of this polynomial.*

Moreover, if T *is cyclic one can augment this list of equivalent statements with the following:*

(d) *the characteristic polynomial of* T *is separable, and* L *contains a complete set of eigenvalues of* T.

By the "eigenvalues of T" we mean those of T in K together with those of T_L in $L \setminus K$.

Proof :

(a) $\Leftrightarrow$ (b) : Obvious.

(b) $\Rightarrow$ (c) : By Corollary 3.4 (with K in that statement replaced by L) the minimal polynomial of T_L factors into distinct monic linear polynomials in $L[x]$, and (b) follows easily.

(c) $\Rightarrow$ (b) : By Proposition 2.14 and Corollary 3.4 (with K in that statement again replaced by L).

(c) $\Leftrightarrow$ (d) : By Proposition 2.1(c).

q.e.d.

We offer a simple example to illustrate how this last corollary might be used in practice.

Example 3.6 : Let $K = \mathbb{Q}$, $V = \mathbb{Q}^4$, and let $T : V \to V$ be the $\mathbb{Q}$-linear operator with usual basis matrix $A := \begin{bmatrix} 0 & 0 & 0 & -9 \\ 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -10 \\ 0 & 0 & 1 & 6 \end{bmatrix}$. Since this matrix is in rational form we conclude from Proposition 2.1 that T is cyclic. However, the characteristic polynomial is $\mathrm{char}_T(x) = x^4 - 6x^3 + 10x^2 - 6x + 9$, which factors in $\mathbb{Q}[x]$ as $(x^2+1)(x-3)^2$, and is therefore not separable. It follows from Corollary 3.5 that $T_L : L \otimes_{\mathbb{Q}} V \to L \otimes_{\mathbb{Q}} V$ cannot be diagonalized for any choice of field extension $L \supset \mathbb{Q}$.

4. Galois Extensions

Throughout this section we assume T admits a cyclic vector.

We now offer a definition of a Galois extension for a linear operator in the spirit of our definition of a Picard-Vessiot extension for a differential structure (see the paragraph immediately after that surrounding (1.29)).

A field extension $L \supset K$ is a *Galois extension for* T if:

(I) the extension is normal;

(II) the L-space $L \otimes_K V$ admits a basis consisting of eigenvectors of T_L; and

(III) when $M \supset K$ is any other field extension satisfying (a) and (b) there is a field embedding $\phi : L \to M$ over K.

The presentation of examples is simplified by first working out an equivalent definition (see Theorem 4.3).

A linear operator $S : V \to V$ is *separable* if the characteristic polynomial $\mathrm{char}_{S,K}(x) \in K[x]$ has this property.

Proposition 4.1 : *Any separable linear operator on V is cyclic.*

The converse is false: see, e.g., Example 3.6.

Proof : When $S : V \to V$ be a linear operator any root of the characteristic polynomial $\mathrm{char}_{S,K}(x)$ is also a root of the minimal polynomial $\mathrm{min}_{S,K}(x)$ (see, e.g., [N, Chapter 7, §2, Theorem 7.6, p. 190]). Under the given hypothesis $\mathrm{char}_{S,K}(x)$ has precisely n roots, all distinct; since $\deg(\mathrm{min}_{S,K}(x)) \leq n$, this forces $\deg(\mathrm{min}_{S,K}(x)) = n$. The result is then immediate from Proposition 2.1(d). **q.e.d.**

Proposition 4.2 : *When* $L \supset K$ *is an extension of fields condition* (II) *in the definition of a Galois extension is equivalent to the following two assertions:*

(a) T *is separable; and*

(b) L *contains a complete set of eigenvalues of* T.

Moreover, (a) *and* (b) *are, in turn, equivalent to:*

(c) *there is a basis* $\mathbf{e}$ *of* $L \otimes_K V$ *such that the* $\mathbf{e}$*-matrix of* T_L *is diagonal, say*

$$\text{(i)} \qquad \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & & 0 \\ 0 & 0 & \lambda_3 & \ddots & \vdots \\ & \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix},$$

Proof : The equivalence of (a) and (b) with (II) was established in Corollary 3.5. The equivalence of (a) and (b) with (c) is obvious. **q.e.d**.

Theorem 4.3 : *Suppose* T *is separable and* $L \supset K$ *is a field extension. Then the following statements are equivalent.*

(a) $L \supset K$ *is a Galois extension for* T.

(b) $L \supset K$ *is a classical Galois extension for the characteristic polynomial of* T.

Proof :

(a) $\Rightarrow$ (b) : By Proposition 4.2 the field L contains a complete set of eigenvalues for T, and therefore a splitting field M for $p := \mathrm{char}_{T,K}(x)$. The extension $M \supset K$ is then a classical Galois extension for the separable polynomial p, and to establish (b) it suffices to prove that $M = L$. The field extension $M \supset K$ obviously satisfies (II) of the definition of a Galois extension, since it contains a full set of eigenvalues of T, and it satisfies (I) by definition. By (III) there must be a field embedding $\phi : L \to M$ over K. Since $L \supset K$ is normal, ϕ must be an automorphism, hence $L = \phi(L) \subset M \subset L$, and $M = L$ follows.

(b) $\Rightarrow$ (a) : Only (III) requires proof, so assume $M \supset K$ satisfies (I) and (II). Then, as above, $M \supset K$ must contain a classical Galois extension $N \supset K$ for p.

But any two classical Galois extensions for p are isomorphic over K (see, e.g., [Lang, Chapter V, §3, Theorem 3.1, p. 236]), and (III) follows. **q.e.d.**

Note that in the statement of Theorem 4.3 the characteristic polynomial $\mathrm{char}_{T,K}(x) \in K[x]$ of T is not assumed irreducible, e.g., some or all of the roots might already be in K. Of course all are in K if and only if $L = K$.

Corollary 4.4 : *Suppose T is separable and the characteristic polynomial $\mathrm{char}_{T,K}(x)$ factors in $K[x]$ in the form $q(x)\prod_{j=r+1}^{n}(x-k_j)$, with $q \in K[x]$ irreducible. Then any classical Galois extension $L \supset K$ for q is a Galois extension for T.*

Proof : Any such extension is a classical Galois extension for $\mathrm{char}_{T,K}(x)$, i.e., it is normal, separable, and generated by the roots of this polynomial. Theorem 4.3 therefore applies. **q.e.d.**

Examples 4.5 : We consider the $\mathbb{Q}$-linear operators $T : \mathbb{Q}^n \to \mathbb{Q}^n$ having the indicated usual basis matrices A. (There is nothing special about $\mathbb{Q}$, except that elementary examples of classical Galois group calculations tend to focus on that field.) The standard necessary and sufficient condition for the separability a polynomial p with coefficients in a field K is that the discriminant

$$\text{(i)} \qquad \Delta := \Big(\prod_{i<j}(\lambda_i - \lambda_j)\Big)^2$$

of that polynomial be non-zero, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the roots, and if this is the case for $p = \mathrm{char}_{T,\mathbb{Q}}(x) \in \mathbb{Q}[x]$ then $\mathbb{Q}(\lambda_1, \lambda_2, \ldots, \lambda_n) \supset \mathbb{Q}$ (which could be a trivial extension) will, by Theorem 4.3, be a Galois extension for T. We will therefore be content to list the characteristic polynomial and the discriminant of that polynomial in each of the cases considered.

(a) $(n = 2)$ $A = \begin{bmatrix} 0 & -c \\ 1 & -b \end{bmatrix}$, $b, c \in \mathbb{Q}$. Here

$$\text{(ii)} \qquad \mathrm{char}_{T,\mathbb{Q}}(x) = x^2 + bx + c,$$

and

$$\text{(iii)} \qquad \Delta = b^2 - 4c.$$

(b) $(n=3)$ $A = \begin{bmatrix} 0 & 0 & -d \\ 1 & 0 & -c \\ 0 & 1 & -b \end{bmatrix}$, $b, c, d \in \mathbb{Q}$. Here

(iv) $$\mathrm{char}_{T,\mathbb{Q}}(x) = x^3 + bx^2 + cx + d,$$

and

(v) $$\Delta = -27d^2 - 4c^2 - (4b^2d - bc^2 - 18cd) \cdot b.$$

(c) $(n=4)$ $A = \begin{bmatrix} 0 & 0 & 0 & -e \\ 1 & 0 & 0 & -d \\ 0 & 1 & 0 & -c \\ 0 & 0 & 1 & -b \end{bmatrix}$, $b, c, d, e \in \mathbb{Q}$. Here

(vi) $$\mathrm{char}_{T,\mathbb{Q}}(x) = x^4 + bx^3 + cx^2 + dx + e,$$

and

(vii) $$\begin{cases} \Delta = 16c^2e - 4c^3d^2 + 256e^3 - 27d^4 + 144cd^23 - 128c^2e^2 \\ + (18cd^3 - 80c^2de - 192de^2) \cdot b + (c^2d^2 - 4c^3e + 144ce^2 - 6d^2e) \cdot b^2 \\ + (-4d^3 + 18cde) \cdot b^3 - 27e^2 \cdot b^4. \end{cases}$$

5. The Galois Group

In this section $T : V \to$ *is a separable* (and therefore cyclic) *linear operator and* $L \supset V$ *is a Galois extension for* T. *We denote the* L*-space* $L \otimes_K V$ *by* V_L.

Define the *Galois group* of T to be the automorphism group over K of the Galois extension $L \supset K$ for T, i.e., the group of automorphisms of L which fix K pointwise. In view of Theorem 4.3 this automorphism group coincides with the classical Galois group of the characteristic polynomial of T. Denote the action of G on L (by evaluation) by $g \cdot \ell$, i.e., let $g \cdot \ell := g(\ell)$ for any $(g, \ell) \in G \times L$.

Define a representation $\rho : G \to \mathrm{GL}(V_L, L)$, which we often express as an action, by

$$\rho(g)(\ell \otimes v) = g \cdot (\ell \otimes v) := (g \cdot \ell) \otimes v, \qquad \ell \otimes v \in L \otimes_K V. \tag{5.1}$$

Now extend T to $T_L : V_L \to V_L$ as in (2.11), and recall that T_L is uniquely determined by the property

$$T_L(\ell \otimes v) := \ell \otimes Tv, \qquad \ell \otimes v \in L \otimes_K V. \tag{5.2}$$

The basic relation between the extension T_L and the representation ρ is that they "commute." Specifically, one has the following analogue of Proposition 1.31.

Theorem 5.3 : *For any* $g \in G$ *one has*

$$\text{(i)} \qquad \rho(g) \circ T_L = T_L \circ \rho(g).$$

In fancier language: T_L is equivariant w.r.t. the given G-action on V_L.

Proof : For any $\ell \otimes v \in V_L$ one has

$$\begin{aligned} (\rho(g) \circ T_L)(\ell \otimes v) &= g \cdot (\ell \otimes Tv) \\ &= (g \cdot \ell) \otimes Tv \\ &= T_L((g \cdot \ell) \otimes v) \\ &= (T_L \circ \rho(g))(\ell \otimes v). \end{aligned}$$

q.e.d.

We now list a few consequences of G being an automorphism group of a field.

Proposition 5.4 : *Suppose* $g \in G,\ \ell_1, \ell_2 \in L,\ v \in V$ *and* $v_L \in V_L$ *are arbitrary. Then*

(a) $g \cdot \big(\ell_1(\ell_2 \otimes v)\big) = (g \cdot \ell_1)((g \cdot \ell_2) \otimes v)$;

(b) $g \cdot \big(\ell_1(\ell_2 \otimes v)\big) = (g \cdot \ell_1)(g \cdot (\ell_2 \otimes v))$; *and*

(c) $g \cdot \ell_1 v_L = (g \cdot \ell_1)(g \cdot v_L)$.

Proof :

(a) From (5.1) we have

$$\begin{aligned} g \cdot \big(\ell_1(\ell_2 \otimes v)\big) &= g \cdot (\ell_1\ell_2 \otimes v) \\ &= (g \cdot (\ell_1\ell_2)) \otimes v \\ &= g(\ell_1\ell_2) \otimes v \\ &= g(\ell_1)g(\ell_2) \otimes v \qquad \text{(because } g \text{ is a field automorphism)} \\ &= g(\ell_1) \cdot (g \cdot \ell_2) \otimes v \\ &= (g \cdot \ell_1)((g \cdot \ell_2) \otimes v). \end{aligned}$$

(b) By (a) and (5.1).

(c) By (b) and additivity.

q.e.d.

Corollary 5.5 : *Suppose* $(\ell, v_L) \in L \times V_L$ *is an eigenpair for* T_L. *Then for any* $g \in G$ *the pair* $(g \cdot \ell,\ g \cdot v_L)$ *has the same property.*

Less formally: G permutes eigenpairs of T_L. This is the analogue for linear operators of Corollary 1.32.

Proof : One has

$$\begin{aligned} (g \cdot \ell)(g \cdot v_L) &= g \cdot \ell v_L \qquad \text{((by Proposition 5.4(c))} \\ &= g \cdot (T_L v_L) \quad \text{(because } T_L v_L = \ell v_L) \\ &= T_L(g \cdot v_L) \quad \text{(by Theorem 5.3).} \end{aligned}$$

q.e.d.

We will need two key facts about the action of G on L (by evaluation).

Proposition 5.6 : *The action of the Galois group on L permutes the eigenvalues of T. Moreover, when the characteristic polynomial* $\mathrm{char}_{T,K}(x) \in K[x]$ *is irreducible this action on the eigenvalues is transitive, i.e., for any two distinct eigenvalues* λ_1, λ_2 *of* T *there is a* (not necessarily unique) $g \in G$ *such that* $g \cdot \lambda_1 = \lambda_2$.

Proof : The first assertion is immediate from Corollary 5.5. The second is standard for the classical Galois group, e.g., see [Lang, Chapter V, §2, p. 233], and, as previously noted, the two groups coincide. **q.e.d.**

6. Fundamental Matrices

In this section $T : V \to V$ *is assumed separable unless specifically stated to the contrary. In addition,* $L \supset K$ *denotes a Galois extension for* T*,* G *is the corresponding Galois group, and* $\mathcal{D} = \mathcal{D}(n, L) \subset \mathrm{GL}(n, L)$ *is the* (multiplicative) *subgroup of invertible diagonal matrices. As in* §4 *we write* $L \otimes_K V$ *as* V_L*.*

Henceforth $\mathfrak{gl}(n, L)$ denotes the L-algebra[22] of $n \times n$ matrices with entries in L, and $\mathrm{diag}_L(n) \subset \mathfrak{gl}(n, L)$ denotes the L-subalgebra of diagonal matrices.

Select a basis $\mathbf{e}$ for V and for the remainder of this section let $A = (a_{ij}) \in \mathfrak{gl}(n, K)$ denote the $\mathbf{e}$-matrix of T. Let I, or I_n when confusion might otherwise result, denote the $n \times n$ identity matrix of $\mathfrak{gl}(n, L)$, which of course is also the $n \times n$ identity matrix of $\mathfrak{gl}(n, K)$.

A matrix $\alpha \in \mathrm{GL}(n, L)$ is a[23] *fundamental* $\mathbf{e}$*-matrix* for T if[24]

$$D_\alpha := \alpha^{-1} A \alpha \in \mathrm{diag}_L(n). \tag{6.1}$$

Since D_α and A in (6.1) are similar the characteristic polynomials must be the same. In particular,

$$\det(D_\alpha) = \det(A) \qquad \text{and} \qquad \mathrm{trace}(D_\alpha) = \mathrm{trace}(A). \tag{6.2}$$

By a *fundamental matrix of* T we mean a fundamental $\mathbf{e}$-matrix of T for some basis $\mathbf{e}$ of V.

Before presenting examples of fundamental matrices we formulate two equivalent definitions.

Let $S : W \to W$ be a linear operator on a finite-dimensional vector space W over a field M. By an *S-eigenbasis* of W we mean a basis of W consisting of eigenvectors of S. (Such a basis need not exist.) By the *transition matrix*[25] between bases $\mathbf{e} = (e_j)_{j=1}^n$ and $\hat{\mathbf{e}} = (\hat{e}_j)_{j=1}^n$ of W we mean the matrix $P = (p_{ij}) \in \mathrm{GL}(n, M)$ defined by $e_j := \sum_i p_{ij} \hat{e}_i$ for $j = 1, 2, \ldots, n$.

[22] The notation reflects the fact that $\mathfrak{gl}(n, L)$ is actually a Lie algebra, but we will not make use of this added structure.

[23] The terminology is not standard. It is used by this author because this matrix is in many ways the analogue of a fundamental matrix solution of a first-order system of linear differential equations, and in that subject "fundamental matrix" is standard.

[24] In practice such α are easy to construct provided one has access to reasonable computer-algebra software. One simply asks for the Jordan form of A and the matrix that conjugates A to that form; the conjugating matrix will then be a fundamental $\mathbf{e}$-matrix for T.

[25] This concept is certainly familiar from elementary linear algebra. We recall the definition here to establish our notation.

Proposition 6.3 : *For any basis* $\mathbf{e} = (e_j)_{j=1}^n$ *of* V *and any* $n \times n$ *matrix* $\alpha \in \mathrm{GL}(n, L)$ *the following assertions are equivalent.*

(a) α *is a fundamental* $\mathbf{e}$*-matrix for* T*;*

(b) *for* $j = 1, 2, \ldots, n$ *the* j^{th}*-column of* α *is the* $\mathbf{e}$*-column of an eigenvector of* T*; and*

(c) α *is the transition matrix from a* T_L*-eigenbasis* $\hat{\mathbf{e}}$ *of* V_L *to the basis* $\mathbf{e}$ $(\simeq (1 \otimes e_j)_{j=1}^n)$ *of* V_L*.*

We will follow custom and refer to eigenvalues, eigenvectors, and eigenpairs of T_L as eigenvalues, eigenvectors and eigenpairs of T.

Proof :

(a) $\Rightarrow$ (b) : For $j = 1, 2, \ldots, n$ let α_j denote column j of α and write D_α in (6.1) as $\mathrm{diag}_L(\lambda_1, \lambda_2, \ldots, \lambda_j)$. Then the equivalent formulation $A\alpha = \alpha D_\alpha$ of (6.1) is in turn equivalent to

$$\text{(i)} \qquad A\alpha_j = \lambda_j \alpha_j \qquad \text{for} \qquad j = 1, 2, \ldots, n.$$

(b) $\Rightarrow$ (c) : If for $j = 1, 2, \ldots, n$ we set $\hat{e}_j := \sum_{ij} \alpha_{ij} e_i$, then α is (by definition) the the transition matrix from the basis $\hat{\mathbf{e}}$ to the basis $\mathbf{e}$, and $\hat{\mathbf{e}}$ is a T_L-eigenbasis of V_L since (i) is then equivalent to

$$\text{(ii)} \qquad T_L \hat{e}_j = \lambda_j \hat{e}_j \qquad \text{for} \qquad j = 1, 2, \ldots, n.$$

(c) $\Rightarrow$ (a) : If (c) holds then (ii) $\Leftrightarrow$ (i) $\Leftrightarrow$ $A\alpha = \alpha D_\alpha$ $\Leftrightarrow$ (6.1).

q.e.d.

Examples 6.4 :

(a) Let $T : \mathbb{Q}^3 \to \mathbb{Q}^3$ be the $\mathbb{R}$-linear mapping defined relative to the usual basis $\mathbf{e}$ by the matrix $A := \begin{bmatrix} 5 & 7 & 7 \\ -2 & 1 & -2 \\ 2 & -3 & 0 \end{bmatrix}$. The characteristic polynomial is $x^3 - 6x^2 - x - 30 = (x - 5)(x - 2)(x + 2)$, and T is therefore separable. For $\alpha := \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$ one verifies that $\alpha^{-1} A \alpha$ is diagonal, and α is therefore a fundamental $\mathbf{e}$-matrix for T.

To illustrate that fundamental matrices are not unique note that $\beta := \begin{bmatrix} 7 & 0 & -5 \\ -7 & 6 & 0 \\ 7 & -6 & 5 \end{bmatrix}$ is also a fundamental **e**-matrix for T. Also note that $D_\alpha := \alpha^{-1}A\alpha = \text{diag}[-2, 3, 5]$ and $D_\beta := \beta^{-1}A\beta = \text{diag}[5.3. - 2]$ are not the same diagonal matrix.

Finally, observe that $\alpha^{-1}\beta = \begin{bmatrix} 0 & 0 & 5 \\ 0 & 6 & 0 \\ 7 & 0 & 0 \end{bmatrix}$ is *not* a fundamental **e**-matrix for T; one has $(\alpha^{-1}\beta)^{-1}A(\alpha^{-1}\beta) = \begin{bmatrix} 0 & -\frac{18}{7} & \frac{10}{7} \\ -\frac{7}{3} & 1 & -\frac{5}{3} \\ \frac{49}{5} & \frac{42}{5} & 5 \end{bmatrix}$, which is certainly not diagonal. The analogue of Theorem 1.38(b) is therefore false for separable linear operators.

(b) When $K = \mathbb{R}$ or $\mathbb{C}$ the previous example generalizes as follows. Any $n \times n$ matrix A with entries in K is the usual basis matrix of a K-linear mapping $T : K^n \to K^n$, and[26] we can always produce an $n \times n$ matrix $\alpha \in \text{GL}(n, K)$ which conjugates A to Jordan form. However, if T is separable that Jordan form must be diagonal, and it is then obvious from (6.1) that α will be a fundamental usual basis matrix for T.

(c) When the **e**-matrix A is in rational form there is a straightforward way to write down a fundamental **e**-matrix for any separable T. To see the method recall the n "elementary symmetric functions" $s_j(z) = s_j(z_1, z_2, \ldots, z_n)$ in n-variables $z_1, z_2, \ldots, z_n$, i.e.,

(i)
$$\begin{cases} s_1(z) &:= z_1 + z_2 + \cdots + z_n \\ s_2(z) &:= z_1z_2 + \cdots + z_1z_n + z_2z_3 + \cdots + z_{n-1}z_n \\ &\vdots \\ s_k(z) &:= \text{the sum of all products of } k \text{ distinct } z_j \\ &\vdots \\ s_n(z) &:= z_1z_2 \ldots z_n. \end{cases}$$

[26]Recall Footnote 24.

These polynomials can also be defined, up to sign, as the coefficients of the various powers of x in the polynomial $\prod_{j=1}^{n}(x - z_j) \in \mathbb{Z}[z_1, x_2, \ldots, z_n][x]$. Indeed, one sees easily that

(ii)
$$\begin{aligned}\prod_{j=1}^{n}(x - z_j) &= x^n - (z_1 + z_2 + \cdots + z_n)x^{n-1} \\ &\quad + (z_1z_2 + z_1z_3 + \cdots + z_{n-1}z_n)x^{n-2} - \cdots + (-1)^n z_1z_2\cdots z_n \\ &= x^n - s_1(z)x^{n-1} + s_2(z)x^{n-2} - \cdots \\ &\quad + (-1)^k s_k(z)x^k + \cdots + (-1)^n s_n(z).\end{aligned}$$

From (ii) above, together with (ii) of Proposition 2.1 and the uniqueness of the rational form, one concludes that when T has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ the matrix A must have the form

(iii) $$A := \begin{bmatrix} 0 & 0 & \cdots & 0 & (-1)^{n+1}s_n(\lambda) \\ 1 & 0 & & 0 & (-1)^n s_{n-1}(\lambda) \\ 0 & 1 & & & \vdots \\ \vdots & & & & -s_2(\lambda) \\ 0 & & \cdots & 1 & s_1(\lambda) \end{bmatrix}, \text{ wherein } s_k(\lambda) := s_k(\lambda_1, \ldots, \lambda_n).$$

To write down a fundamental **e**-matrix for T we must first exhibit an eigenvector for each λ_j, and to this end the following observation proves useful. Suppose $1 \leq j, k \leq n$ and we set $z_j = 0$ in $s_k(z)$. The result, which we denote by $s_{k,j}(z)$, consists precisely of those terms in $s_k(z)$ which do not involve z_j. If we then multiply that result by z_j we obtain all those terms of $s_{k+1}(z)$ which do involve z_j, i.e., $s_{k+1}(z) - s_{k+1,j}(z)$, and we must therefore have

(iv) $$z_j \cdot s_{k,j}(z) = s_{k+1}(z) - s_{k+1,j}(z),$$

by which we mean

(v) $$z_j \cdot s_{n-1,j}(z) = s_n(z)$$

when $k = n - 1$.

We claim that

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & (-1)^{n+1}s_n(\lambda) \\ 1 & 0 & & 0 & (-1)^n s_{n-1}(\lambda) \\ 0 & 1 & & & \vdots \\ \vdots & & & & -s_2(\lambda) \\ 0 & & \cdots & 1 & s_1(\lambda) \end{bmatrix} \begin{bmatrix} (-1)^{n+1}s_{n-1,j}(\lambda) \\ (-1)^n s_{n-2,j}(\lambda) \\ \vdots \\ (-1)^{k+2}s_{k,j}(\lambda) \\ \vdots \\ (-1)^3 s_{1,j}(\lambda) \\ (-1)^2 \end{bmatrix} = \lambda_j \begin{bmatrix} (-1)^{n+1}s_{n-1,j}(\lambda) \\ (-1)^n s_{n-2,j}(\lambda) \\ \vdots \\ (-1)^{k+2}s_{k,j}(\lambda) \\ \vdots \\ (-1)^3 s_{1,j}(\lambda) \\ (-1)^2 \end{bmatrix},$$

hence that $\sum_{k=1}^{n-1}(-1)^{k+2}s_{k,j}(\lambda)e_k + e_n$ is an eigenvector of T with associated eigenvalue λ_j. Indeed, the top (column) entry in the matrix product on the left is $\lambda_k \cdot (-1)^{n+1}s_{n-1,j}(\lambda)$ by (v), the bottom entry is $-s_{1,j}(\lambda)+s_1(\lambda) = \lambda_j$, both precisely as needed. Intermediate entry k is given by $(-1)^k(s_{k,j}(\lambda) - s_k(\lambda))$, which by (iv) can be written $\lambda_j \cdot (-1)^{k+1}s_{k-1,j}(\lambda)$, and the claim is thereby established.

Since the eigenvalues are distinct (by separability) these eigenvectors must form a basis, and we conclude that

(vi) $$\alpha := \begin{bmatrix} (-1)^{n-1}s_{n-1,1}(\lambda) & (-1)^{n-1}s_{n-1,2}(\lambda) & \cdots & (-1)^{n-1}s_{n-1,n}(\lambda) \\ (-1)^{n-2}s_{n-2,1}(\lambda) & (-1)^{n-2}s_{n-2,2}(\lambda) & \cdots & (-1)^{n-2}s_{n-2,n}(\lambda) \\ \vdots & \vdots & & \vdots \\ (-1)^k s_{k,1}(\lambda) & (-1)^k s_{k,2}(\lambda) & \cdots & (-1)^k s_{k,n}(\lambda) \\ \vdots & \vdots & & \vdots \\ -s_{1,1}(\lambda) & -s_{1,2}(\lambda) & \cdots & -s_{1,n}(\lambda) \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

is a fundamental **e**-matrix for T. In more compact form:

(vii) $$\alpha = (\alpha_{ij}), \qquad \text{where} \qquad \alpha_{ij} := \begin{cases} (-1)^{n-i}s_{n-i,j}(\lambda) & \text{if } 1 \le i < n \\ 1 & \text{if } i = n. \end{cases}$$

For later reference we list the matrices (vii) corresponding to the case $n = 2, 3$

and 4:

(viii)
$$\begin{cases} \text{(a)} \quad (n=2) \qquad \alpha_2 = \begin{bmatrix} -\lambda_2 & -\lambda_1 \\ 1 & 1 \end{bmatrix}; \\ \\ \text{(b)} \quad (n=3) \qquad \alpha_3 = \begin{bmatrix} \lambda_2\lambda_3 & \lambda_1\lambda_3 & \lambda_1\lambda_2 \\ -\lambda_2-\lambda_3 & -\lambda_1-\lambda_3 & -\lambda_1-\lambda_2 \\ 1 & 1 & 1 \end{bmatrix}; \\ \\ \text{(c)} \quad (n=4) \end{cases}$$

$$\alpha_4 = \begin{bmatrix} -\lambda_2\lambda_3\lambda_4 & -\lambda_1\lambda_3\lambda_4 & -\lambda_1\lambda_2\lambda_4 & -\lambda_1\lambda_2\lambda_3 \\ \lambda_2\lambda_3+\lambda_2\lambda_4+\lambda_3\lambda_4 & \lambda_1\lambda_3+\lambda_1\lambda_4+\lambda_3\lambda_4 & \lambda_1\lambda_2+\lambda_1\lambda_4+\lambda_2\lambda_4 & \lambda_1\lambda_2+\lambda_1\lambda_3+\lambda_2\lambda_3 \\ -\lambda_2-\lambda_3-\lambda_4 & -\lambda_1-\lambda_3-\lambda_4 & -\lambda_1-\lambda_2-\lambda_4 & -\lambda_1-\lambda_2-\lambda_3 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

There is an alternate method for proving the non-singularity of α; it involves a bit more work, but one is rewarded by a more interesting result. Specifically, we claim that

(ix)
$$\det(\alpha) = \prod_{i<j}^{n}(\lambda_i - \lambda_j),$$

hence that the square of this determinant is the discriminant Δ of the characteristic polynomial of T.

More generally, we claim that for
(x)
$$\alpha(n, z(n)) := \begin{bmatrix} (-1)^{n-1}s_{n-1,1}(z) & (-1)^{n-1}s_{n-1,2}(z) & \cdots & (-1)^{n-1}s_{n-1,n}(z) \\ (-1)^{n-2}s_{n-2,1}(z) & (-1)^{n-2}s_{n-2,2}(z) & \cdots & (-1)^{n-2}s_{n-2,n}(z) \\ \vdots & \vdots & & \vdots \\ (-1)^{k}s_{k,1}(z) & (-1)^{k}s_{k,2}(z) & \cdots & (-1)^{k}s_{k,n}(z) \\ \vdots & \vdots & & \vdots \\ -s_{1,1}(z) & -s_{1,2}(z) & \cdots & -s_{1,n}(z) \\ 1 & 1 & \cdots & 1 \end{bmatrix},$$

where $n \geq 2$ is arbitrary and $z = z(n) = (z_1, z_2, \ldots, z_n)$, one has

(xi)
$$\det(\alpha(n, z(n))) = \prod_{i<j}^{n}(z_i - z_j),$$

We argue by induction on $n \geq 2$, immediately dispensing with the case $n = 2$ by direct calculation. We therefore assume $n > 2$, and that the result holds for all integers m satisfying $2 \leq m < n$.

Fix $1 \leq j < n$ and subtract column n from column j in (x). The result for the bottom entry will be 0, whereas for $1 \leq k < n$ the result for the entry with initial subscript k will be

$$(-1)^k\big(s_{k,j}(z(n)) - s_{k,n}(z(n))\big). \tag{xii}$$

The terms of $s_{k,j}(z(n))$ and $s_{k,n}(z(n))$ not involving either z_j or z_n will cancel in pairs. Since there are no terms in $s_{k,j}(z(n))$ involving z_j, and none in $s_{n,j}(z(n))$ involving z_n, the remaining terms can also be paired: if t is a term of $s_{k,j}(z(n))$ involving z_n, then $z_j \cdot t/z_n$ can be designated as the corresponding term of $s_{n,j}(z(n))$, and the procedure can obviously be reversed. Since t/z_n is a term of $s_{k-1,j}(z(n-1)))$ one sees that (xii) simplifies by means of these pairings to

$$(-1)^k(x_n - x_j)s_{k-1,j}(z(n-1)) = (z_j - z_n) \cdot (-1)^{k-1}s_{k-1,j}(z(n-1)),$$

and by factoring $(z_j - z_n)$ out of column j for $1 \leq j < n$ the determinant calculation in (x) is reduced to

$$\prod_{1\leq j<n}(z_j - z_n)\cdot\det\begin{bmatrix} (-1)^{n-2}s_{n-2,1}(z) & (-1)^{n-2}s_{n-2,2}(z) & \cdots & (-1)^{n-1}s_{n-1,n}(z) \\ (-1)^{n-3}s_{n-3,1}(z) & (-1)^{n-3}s_{n-3,2}(z) & \cdots & (-1)^{n-2}s_{n-2,n}(z) \\ \vdots & \vdots & & \vdots \\ (-1)^{k-1}s_{k-1,1}(z) & (-1)^{k-1}s_{k-1,2}(z) & \cdots & (-1)^k s_{k,n}(z) \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \cdots & -s_{1,n}(z) \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

where in all but (possibly) the final column $z = z(n-1) := (z_1, z_2, \ldots, z_{n-1})$. Expanding the final determinant along the bottom row results in a value of $(-1)^{2n}\det(\alpha(n-1, z(n-1))$. By induction we have $\det(\alpha(n-1, z(n-1)) = \prod_{1\leq i<j<n}(z_i - z_j)$, and the full determinant is therefore

$$\textstyle\prod_{1\leq j<n}(z_j - z_n) \cdot \prod_{1\leq i<j<n}(z_i - z_j) = \prod_{1\leq i<j\leq n}(z_i - z_j).$$

Equality (xi) is thereby established, and (ix) follows immediately.

One can significantly improve the results of Example 6.4(c) by choosing a cyclic vector for the dual operator T^* and using the dual of the resulting basis as the basis for V. One then obtains striking analogues of Proposition 1.39 and Corollary 1.41.

Proposition 6.5 : *The following statements are equivalent, even without the standing assumption in this section that T be separable.*

(a) *$T^* : V^* \to V^*$ admits a cyclic vector.*

(b) *There is a basis $\mathbf{e}$ of V such that the $\mathbf{e}$-matrix of T has the form*

(i)
$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix},$$

i.e., A is the transpose of the unique rational form of T^*.

Moreover, either (and therefore both) *of these conditions holds when the separability hypothesis on T is reimposed.*

Proof : Let $\mathbf{e}$ be a basis for V and let $\mathbf{e}^*$ be the dual basis of V^*. Since $(V^*)^* \simeq V$, we can (and do) regard $\mathbf{e}$ as the dual basis of $\mathbf{e}^*$. As we have seen previously, the $\mathbf{e}$-matrix for T is M if and only if the $\mathbf{e}^*$-matrix of T^* is M^τ. However, for A as in (i) we have

$$A^\tau = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & -a_{n-2} \\ 0 & \cdots & 0 & 0 & 1 & -a_{n-1} \end{bmatrix},$$

which is precisely of the form associated with (a basis generated by) a cyclic vector (recall Proposition 2.1). **q.e.d.**

We now re-institute the separability hypothesis on T.

Corollary 6.6 : *There is a*(n ordered) *basis* $\mathbf{e}$ *of* V *such that the* $\mathbf{e}$*-matrix of* T *has the form seen in* (i) *of* Proposition 6.5. *In fact one can take* $\mathbf{e} = (e_j)_{j=1}^n$, *where* $\hat{\mathbf{e}} = (e_{n+1-j})_{j=1}^n$ *is any basis for which the* $\hat{\mathbf{e}}$*-matrix of* T *is in rational form.*

In specific examples involving not-too-large matrices the standard computer algebra systems can easily produce such matrices.

Proof : For the initial assertion recall Proposition 2.3. The second assertion is a basis reformulation of the first proof of Proposition 2.3. **q.e.d**.

Proposition 6.7 : *Let* $\ell_1, \ell_2, \ldots, \ell_n \in L$ *be a complete set of eigenvalues of* T, *set* $\ell := (\ell_1, \ell_2, \ldots, \ell_n) \in L^n$, *and let* $\mathbf{e}$ *be a basis of* V *as in the statement of* Corollary 6.6, *i.e., such that the* $\mathbf{e}$*-matrix of* T *has the form seen in* (i) *of* Proposition 6.5. *Then an* $n \times n$ *matrix* $\alpha \in \mathrm{GL}(n, L)$ *is a fundamental* $\mathbf{e}$*-matrix of* T *if and only if* α *can be expressed as* $\alpha = \mathrm{vdmm}_{K[x],n}(\ell) \cdot D$, *where* $D \in \mathrm{GL}(n, L)$ *is a diagonal matrix.*

Proof : The condition for $\alpha = (\alpha_{ij}) \in \mathrm{G}(n, L)$ to be a fundamental matrix of T is (by definition) that $\alpha^{-1} A \alpha$ be diagonal, say $\mathrm{diag}[d_1, d_2, \ldots, d_n]$. For purposes of this proof it is more convenient to write this condition as

$$\text{(i)} \qquad A(\alpha_{ij}) = (\alpha_{ij}) \cdot \mathrm{diag}[d_1, d_2, \ldots, d_n]$$

or, in full matrix form, as

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2n} \\ \vdots & & & & \\ \alpha_{n-1,1} & \alpha_{n-1,2} & \alpha_{n-1,33} & \cdots & \alpha_{n-1,n} \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \cdots & \alpha_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2n} \\ \vdots & & & & \\ \alpha_{n-1,1} & \alpha_{n-1,2} & \alpha_{n-1,33} & \cdots & \alpha_{n-1,n} \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \cdots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & d_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & d_n \end{bmatrix}.$$

Fix $1 \leq j \leq n$. Equality (i) obviously holds if and only if j^{th}-columns of the upper and lower products displayed above are equal, and α is therefore a fundamental **e**-matrix for T if and only if

$$\begin{bmatrix} \alpha_{2j} \\ \alpha_{3j} \\ \vdots \\ \alpha_{nj} \\ -\sum_{k=0}^{n-1} a_k \alpha_{k+1,j} \end{bmatrix} = \begin{bmatrix} d_j\alpha_{1j} \\ d_j\alpha_{2j} \\ \vdots \\ d_j\alpha_{n-1,j} \\ d_j\alpha_{nj} \end{bmatrix} \qquad \text{for all} \qquad 1 \leq j \leq n.$$

These last conditions are, in turn, equivalent to

(ii)
$$\begin{cases} \text{(a)} \begin{cases} \alpha_{2j} = d_j\alpha_{1j} \\ \alpha_{3j} = d_j\alpha_{2j} = d_j^2\alpha_{1j} \\ \vdots \\ \alpha_{nj} = d_j\alpha_{n-1,j} = d_j^{n-1}\alpha_{1j} \end{cases} \\ \quad \text{together with} \\ \text{(b)} \quad d_j\alpha_{nj} + \sum_{k=0}^{n-1} a_k\alpha_{k+1,j} = 0. \end{cases}$$

Note that substituting the identities of (iia) into (iib) gives

(iii)
$$(d_j^n + \textstyle\sum_{k=0}^{n-1} a_k d_j^k) \cdot \alpha_{1j} = 0.$$

For α to be fundamental we must have $\det(\alpha) \neq 0$, hence $\alpha_{1j} \neq 0$ in (ii) and (iii), and we conclude that α is a fundamental **e**-matrix for T if and only if for each $1 \leq j \leq n$ the element $d_j \in L$ is a solution of $x^n + \sum_{k=0}^{n-1} a_k x^k = 0$, and that α has j^{th}-column $\alpha_{1j} \begin{bmatrix} 1 \\ d_j \\ d_j^2 \\ \vdots \\ d_j^{n-1} \end{bmatrix}$. But this is equivalent to each d_j being an eigenvalue

of T, and to α having the form

(iv) $\mathrm{vdmm}_{K[x],n}(k)D$, where $D = \begin{bmatrix} \alpha_{11} & 0 & 0 & \cdots & 0 \\ 0 & \alpha_{12} & 0 & \cdots & 0 \\ 0 & 0 & \alpha_{13} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \alpha_{1n} \end{bmatrix} \in \mathrm{Gl}(n, L)$.

The proof is therefore complete. **q.e.d.**

We can now give the analogue for operators of Corollary 1.41.

Corollary 6.8 : *Suppose $\ell_1, \ell_2, \ldots, \ell_n \in L$ are distinct and $\ell := (\ell_1, \ell_2, \ldots, \ell_n) \in L^n$. Then the following assertions are equivalent.*

(a) *The Vandermonde matrix $\mathrm{vdmm}_{K[x],n}(\ell)$ is a fundamental matrix of T.*

(b) *The elements $\ell_1, \ell_2, \ldots, \ell_n \in L$ constitute a complete set of eigenvalues of T.*

Since L contains all the roots of the characteristic polynomial of T, it follows that *one can always choose a Vandermonde matrix as a fundamental matrix.*

The proof is actually constructive: it shows that if one chooses a basis $\mathbf{e}$ such that the $\mathbf{e}$-matrix of T is the transpose of the the rational form of this operator, then $\mathrm{vdmm}_{K[x],n}(\ell)$ will be a fundamental $\mathbf{e}$-matrix of T. If one is in possession of a basis $\hat{\mathbf{e}}$ which exhibits this rational form, a basis $\mathbf{e}$ with the desired property can always be constructed as in Corollary 6.6.

Proof :

(a) $\Rightarrow$ (b) : Abbreviate $\mathrm{vdmm}_{K[x],n}(\ell)$ as v.

By assumption there is a basis $\mathbf{e}$ of V such that $v^{-1}Av = \mathrm{diag}[d_1, d_2, \ldots, d_n]$, where $d_j \in L$ for $j = 1, 2, \ldots, n$, and where A is the $\mathbf{e}$-matrix of T. It is immediate from similarity that $d_1, d_2, \ldots, d_n$ must be a full set of eigenvalues of A. Since the eigenvalues of A are those of T, (b) follows.

(b) $\Rightarrow$ (a) : Since T is assumed separable (in this section) T must be cyclic (Proposition 4.1), and the same therefore holds for T^* (Proposition 2.3). It then follows from Proposition 6.5 that there is a basis $\mathbf{e}$ of V such that the $\mathbf{e}$-matrix of T has the form seen in (i) of that result, i.e., it is the transpose of the rational form of T. Assertion (a) is now verified by choosing $D = I$ (the identity matrix) in the statement of Proposition 6.7.

q.e.d.

Example 6.9 : Let $T : \mathbb{Q}^4 \to \mathbb{Q}^4$ be the linear operator with usual basis representation

$$A = \begin{bmatrix} 6 & 18 & 62 & -131 \\ 35 & 310 & 1127 & -2209 \\ 0 & -14 & -52 & 99 \\ 5 & 37 & 134 & -264 \end{bmatrix}.$$

The characteristic polynomial $x^4 + 2x^2 - 8$ has discriminant $-165,888$ (according to the computer algebra package this author is using), from which we see that T is separable. The rational form of A is

$$\begin{bmatrix} 0 & 0 & 0 & 8 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

and transpose of this matrix will have the form needed for Proposition 6.7. When my computer algebra package computed a matrix converting that transpose to Jordan form the result was

$$\begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \\ -\frac{i}{3} & \frac{i}{3} & \frac{\sqrt{2}}{3} & -\frac{\sqrt{2}}{3} \\ -\frac{2}{3} & -\frac{2}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{4i}{3} & -\frac{4i}{3} & \frac{2\sqrt{2}}{3} & -\frac{2\sqrt{2}}{3} \end{bmatrix},$$

Expressing this matrix in the form

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -2i & 2i & \sqrt{2} & -\sqrt{2} \\ -4 & -4 & 2 & 2 \\ 8i & -8i & 2\sqrt{2} & -2\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$= \mathrm{vdmm}_{\mathbb{Q}[x],4}(-2i, 2i, \sqrt{2}, -\sqrt{2}) \cdot \begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{bmatrix}$$

then gives a concrete illustration of Proposition 6.7.

7. Preliminaries on Permutation Matrices

S_n denotes the symmetric group on n letters; the identity permutation will be denoted e. A permutation $\sigma \in S_n$ will, when convenient, be expressed as

$$\sigma = \begin{bmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{bmatrix}.$$

Unless specifically indicated to the contrary, e.g., by column vector notation, all matrices are assumed $n \times n$. δ_{ij} denotes the Kronecker delta, and a superscript τ indicates matrix transposition.

By a *permutation matrix* we mean any matrix p of the form $(\delta_{i\sigma(j)})$, where $\sigma \in S_n$. In other words, p must result from the identity matrix I by a permutation of the columns. (One could also use rows, although much of what follows would need to be changed accordingly.) To indicate the relationship to σ we write p as p_σ, hence

$$p_\sigma := (\delta_{i\sigma(j)}). \tag{7.1}$$

The collection of permutation matrices, which constitutes a subset of $\mathrm{GL}(n, \mathbb{Z})$, is denoted by $\mathcal{P}_n$.

The first result of the section suggests why permutation matrices might be of interest in our context. It is the analogue for separable operators of Theorem 1.38(a) for differential structures.

Theorem 7.2 : *Suppose $T : V \to V$ is a separable linear operator on an n-dimensional K-space, $\mathbf{e}$ is a basis for V, and A is the $\mathbf{e}$-matrix of T. Suppose α and β are fundamental $\mathbf{e}$-matrices for T with entries in some extension field $L \supset K$. Then $\alpha^{-1}\beta$ has the form dp_σ, where d is a diagonal matrix with entries in L and p_σ is a permutation matrix.*

Proof : By hypothesis we have both

$$\text{(i)} \qquad \alpha^{-1}A\alpha = \ell^\alpha := \mathrm{diag}[\ell_1^\alpha, \ell_2^\alpha, \ldots, \ell_n^\alpha] = (\ell_i^\alpha \delta_{ij})$$

and

$$\text{(ii)} \qquad \beta^{-1}A\beta = \ell^\beta := \mathrm{diag}[\ell_1^\beta, \ell_2^\beta, \ldots, \ell_n^\beta] = (\delta_{ij}\ell_j^\beta)$$

(recall (6.1)), which for the purposes of this proof we express as

(iii) $$\alpha\ell^\alpha = A\alpha \qquad \text{and} \qquad \beta\ell^\beta = A\beta.$$

If we write $\alpha^{-1}\beta$ as m we see from (iii) and (i) that

$$\begin{aligned} m\ell^\beta &= \alpha^{-1}\beta\ell^\beta \\ &= \alpha^{-1}A\beta \\ &= \alpha^{-1}A\alpha\alpha^{-1}\beta \\ &= \ell^\alpha m, \end{aligned}$$

i.e., that

(iv) $$m\ell^\beta = \ell^\alpha m.$$

Since both collections ℓ_i^α and ℓ_j^β constitute full sets of eigenvalues of T in L there must be a permutation $\sigma \in S_n$ such that

(v) $$\ell^\alpha_{\sigma(j)} = \ell^\beta_j \qquad \text{for} \qquad j = 1, 2, \ldots, n.$$

If we write m as (m_{ij}) we see from (ii) and (v) that the ij term on the left in (iv) is

$$\textstyle\sum_k m_{ik}\delta_{kj}\ell^\beta_j = m_{ij}\ell^\beta_j = m_{ij}\ell^\alpha_{\sigma(j)} = m_{ij}\ell^\alpha_{\sigma(j)}m_{ij},$$

from (i) that the corresponding term on the right in (iv) is

$$\textstyle\sum_k \ell^\alpha_i\delta_{ik}m_{kj} = \ell^\alpha_i m_{ij},$$

and we therefore have

(vi) $$(\ell^\alpha_i - \ell^\alpha_{\sigma(j)})m_{ij} = 0 \qquad \text{for all} \qquad 1 \le i, j \le n.$$

From the separability assumption the ℓ_i^α are pairwise distinct, whereupon from (vi) and $\det(m) \neq 0$ we conclude that $m_{ij} = 0$ if and only if $i \neq \sigma(j)$, hence that $m = (m_{\sigma(j)j}\delta_{i\sigma(j)}) = \mathrm{diag}[m_{\sigma(1)1}, m_{\sigma(2)2}, \ldots, m_{\sigma(n)n}] \cdot p_\sigma$. **q.e.d.**

By choosing σ in (7.1) to be the identity permutation we see that

(7.3) $$p_e = I \in \mathcal{P}_n.$$

We claim that

(7.4) $$p_\sigma, p_\nu \in \mathcal{P}_n \qquad \Rightarrow \qquad p_\sigma p_\nu = p_{\sigma\circ\nu} \in \mathcal{P}_n,$$

hence that $\mathcal{P}_n$ is closed under (matrix) multiplication. Indeed, we have

$$p_\sigma p_\nu = \left(\sum_k \delta_{i\sigma(k)}\delta_{k,\nu(j)}\right), \tag{7.5}$$

and for any fixed $1 \leq k \leq n$ we see that

$$\delta_{i\sigma(k)}\delta_{k,\nu(j)} = \begin{cases} 1 & \Leftrightarrow \quad \sigma(k) = i \ \text{ and } \ \nu(j) = k \quad \Leftrightarrow \quad (\sigma \circ \nu)(j) = i \ \text{ and } \ k = \sigma^{-1}(i) \\ 0 & \text{otherwise.} \end{cases}$$

Equality (7.5) thereby reduces to $p_\sigma p_\nu = (\delta_{i,(\sigma\circ\nu)(j)}) = p_{\sigma\circ\nu}$, which establishes our claim.

By choosing $\nu = \sigma^{-1}$ in (7.4) and recalling (7.3) we see that

$$p_\sigma \in \mathcal{P}_n \qquad \Rightarrow \qquad p_\sigma^{-1} = p_{\sigma^{-1}} \in \mathcal{P}_n. \tag{7.6}$$

It now follows from (7.3) and (7.4) that $\mathcal{P}_n$ is a subgroup of $\mathrm{GL}(n,\mathbb{Z})$, which we refer to as the *group of permutation matrices*.

We next observe that for p_σ as in (7.1) we have

$$p_\sigma^\tau = (\delta_{\sigma(i)j}). \tag{7.7}$$

To see this write $p = (p_{ij})$ and $p^\tau = (q_{ij})$. Then

$$q_{ij} = p_{ji} = \delta_{j\sigma(i)} = \delta_{\sigma(i)j},$$

and (7.7) follows.

Proposition 7.8 : *For any $p_\sigma \in \mathcal{P}_n$ one has*

$$\text{(i)} \qquad p_\sigma^\tau \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix} = \begin{bmatrix} \sigma(1) \\ \sigma(2) \\ \vdots \\ \sigma(n) \end{bmatrix}.$$

Moreover, the permutation matrix p_σ is uniquely determined by this identity, i.e., if $p \in \mathcal{P}_n$ satisfies

$$\text{(ii)} \qquad p^\tau \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix} = \begin{bmatrix} \sigma(1) \\ \sigma(2) \\ \vdots \\ \sigma(n) \end{bmatrix},$$

then $p = p_\sigma$.

For example, suppose

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{bmatrix},$$

i.e., that $\sigma(1) = 4,\ \sigma(2) = 2,\ \sigma(3) = 1$ and $\sigma(4) = 3$. Then from the two columns appearing below one determines by inspection that

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 1 \\ 3 \end{bmatrix},$$

hence that

$$p_\sigma = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Many authors prefer to replace (7.1) by $p_\sigma := (\delta_{\sigma(i)j})$, i.e., they view permutation matrices in terms of row permutations rather than column permutations. When that choice is made equality (ii) would more likely be expressed in the row vector form

$$[1\ 2\ \cdots\ n]\, p_\alpha^\tau = [\sigma(1), \sigma(2), \ldots, \sigma(n)]\,.$$

Proof : By (7.7) we see that the $i1$-entry of the matrix product on the left in (i) is

$$\textstyle\sum_k \delta_{\sigma(i)k} k = \sigma(i),$$

and this establishes (i).

To complete the proof note from the hypothesis $p \in \mathcal{P}_n$ that p must have the form $p_\nu = (\delta_{i\nu(j)})$ for some permutation $\nu \in S_n$. By replacing σ with ν in the previous paragraph (ii) is seen to reduce to

$$\begin{bmatrix} \nu(1) \\ \nu(2) \\ \vdots \\ \nu(n) \end{bmatrix} = \begin{bmatrix} \sigma(1) \\ \sigma(2) \\ \vdots \\ \sigma(n) \end{bmatrix},$$

hence $\nu = \sigma$, and $p = p_\nu = p_\sigma$ follows. **q.e.d.**

Proposition 7.9 :

(a) *Every permutation matrix is orthogonal, i.e.,* $\mathcal{P}_n \subset O(n, Z)$.

(b) *The mapping* $\theta : \sigma \in S_n \mapsto p_\sigma \in \mathcal{P}_n$ *is a group isomorphism.*

Proof :

(a) First note that for any $1 \leq i, j \leq n$ we have $\sigma(i) = j \Leftrightarrow i = \sigma^{-1}(j)$, from which one immediately sees that $\delta_{\sigma(i),j} = \delta_{i,\sigma^{-1}(j)}$. From (7.7), (7.1) and (7.6) we conclude that

$$p_\sigma^\tau = (\delta_{\sigma(i),j}) = (\delta_{i,\sigma^{-1}(j)}) = p_{\sigma^{-1}} = p_\sigma^{-1},$$

and (a) follows.

(b) The mapping θ is a group homomorphism by (7.3) and (7.4). If $p_\sigma = I = p_e$ then $\sigma = e$ by the assertion surrounding (ii) of Proposition 7.8(b), and injectivity follows. Surjectivity is evident from the definition of a permutation matrix (see the discussion leading to (7.1)).

q.e.d.

Proposition 7.10 : *Let* A *be any* $n \times n$ *matrix with coefficients in a* (commutative) *ring* R (with unity). *Then for any* $\sigma \in S_n$ *the* j^{th}*-column of the matrix* Ap_σ *is the* $\sigma(j)^{th}$*-column of* A.

In other words, when we write $A = [A_1\ A_2\ \cdots\ A_n]$ in terms of columns the corresponding expression for AP_σ is $[A_{\sigma(1)}\ A_{\sigma(2)}\ \cdots\ A_{\sigma(n)}]$. Less formally: right multiplication by p_σ is equivalent to permuting the columns of A by σ.

Proof : If $A = (a_{ij})$ then from (7.1) we see that ij-entry of Ap_σ is $\sum_k a_{ik}\delta_{k\sigma(j)} = a_{i\sigma(j)}$. **q.e.d.**

Let R be as in Proposition 7.10. We claim that a left action of S_n on the polynomial algebra $R[x] = R[x_1, x_2, \ldots, x_n]$ is defined by[27]

$$(7.11)\quad \sigma \cdot p = (\sigma \cdot p)(x_1, x_2, \ldots, x_n) = p(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}),\quad (\sigma, p) \in S_n \times R[x].$$

Indeed, that $e \cdot p = p$, where $e \in S_n$ is the identity permutation and $p \in R[x]$ is arbitrary, is obvious. Now suppose $\nu, \sigma \in S_n$ and $p, q \in R[x]$ are arbitrary. Then

[27] In this definition we are following [Lang, Chapter I, §5, p. 30]. One also sees the subscripts $\sigma(j)$ replaced by $\sigma^{-1}(j)$, e.g., as in [E, Chapter 1, §3, Example 1.1, p. 25].

- $\sigma \cdot p$ results from p by replacing each x_i with $x_{\sigma(i)}$ or, equivalently, each $x_{\sigma^{-1}(i)}$ with x_i, and
- $\nu \cdot q$ results from q by replacing each x_j with $x_{\nu(j)}$.

By choosing $q = \sigma \cdot p$ and $j = i$ in the last bulleted item it follows that $\nu \cdot (\sigma \cdot p)$ results from p by replacing each $x_{\sigma^{-1}(i)}$ with $x_{\nu(i)}$ or, equivalently, by replacing each x_i with $x_{\nu(\sigma(i))} = x_{(\nu \circ \sigma)(i)}$, which by definition is $(\nu \circ \sigma) \cdot p$. We conclude that $\nu \cdot (\sigma \cdot p) = (\nu\sigma) \cdot p$, and this establishes our claim.

Example 7.12 : Take $R = \mathbb{Z}$, $n = 4$, $p = x_1 + x_1x_2 + x_1x_2x_3$, $\sigma = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}$ and $\nu = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}$. From (7.11) we obtain

$$\begin{aligned} \sigma \cdot p &= x_3 + x_3x_1 + x_3x_1x_2 \\ &= x_3 + x_1x_3 + x_1x_2x_3, \end{aligned}$$

and by a second appeal to that formula, now with p replaced by $\sigma \cdot p$, we obtain

$$\begin{aligned} \nu \cdot (\sigma \cdot p) &= x_2 + x_1x_2 + x_1x_3x_2 \\ &= x_2 + x_1x_2 + x_1x_2x_3. \end{aligned}$$

On the other hand, we see from $\nu\sigma \ (= \nu \circ \sigma) = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}$ and a third appeal to (7.11), this time with σ replaced by $\nu\sigma$, that

$$\begin{aligned} \nu\sigma \cdot p &= x_2 + x_1x_2 + x_2x_1x_3 \\ &= x_2 + x_1x_2 + x_1x_2x_3, \end{aligned}$$

and we have thereby confirmed that $\nu \cdot (\sigma \cdot p) = \nu\sigma \cdot p$ through straightforward verification.

For any subgroup $G \subset S_n$ one obtains an action of G on $R[x]$ by restricting to $\sigma = g \in G$ in (7.11). A polynomial $p \in R[x]$ is an *invariant* (of this action) of G, or is (a) *G-invariant*, if

$$g \cdot p = p \qquad \text{for all} \qquad g \in G. \tag{7.13}$$

For our immediate purposes the most important example of a G-invariant, for any subgroup $G \subset S_n$ (including $G = S_n$), is the polynomial

$$p = \Big(\prod_{i<j}(x_i - x_j) \Big)^2 \in \mathbb{Z}[x] = \mathbb{Z}[x_1, x_2, \ldots, x_n]. \tag{7.14}$$

Indeed, (7.13) is obvious from (7.11). More generally, from (7.11) we see that the polynomial

$$\sqrt{p} := \prod_{i<j}(x_i - x_j) \in \mathbb{Z}[x] \tag{7.15}$$

satisfies

$$g \cdot \sqrt{p} = \pm\sqrt{p} \qquad \text{for any} \qquad g \in S_n, \tag{7.16}$$

from which (7.13) immediately follows. ($\sqrt{p}$ is an example of a "semi-invariant.")

One defines the *sign* of a permutation $g \in S_n$ by

$$\mathrm{sgn}(g) := \begin{cases} +1 & \text{if} \quad g \cdot \sqrt{p} \;=\; \sqrt{p} \\ -1 & \text{if} \quad g \cdot \sqrt{p} \;=\; -\sqrt{p}, \end{cases} \tag{7.17}$$

which immediately gives

$$g \cdot \sqrt{p} = \mathrm{sgn}(g)\sqrt{p}. \tag{7.18}$$

Permutations g satisfying $\mathrm{sgn}(g) = 1$ are *even*; those which satisfy $\mathrm{sgn}(g) = -1$ are *odd*.

Readers are assumed familiar with the notion of the sign of a permutation, but perhaps not with the approach we have taken. The following standard fact, and the consequent definition, are also assumed familiar.

Proposition 7.19 : *The function* $\mathrm{sgn} : g \in S_n \mapsto \mathrm{sgn}(g) \in \{1, -1\}$ *is a group homomorphism from* S_n *onto the group of units* $\{1, -1\}$ *of* $\mathbb{Z}$. *The even permutations form the kernel, and this collection is therefore a normal subgroup of* S_n.

The kernel is called the *alternating group* (*on* n *letters*), and is denoted A_n.

Corollary 7.20 : *When G is a subgroup of S_n the following statements are equivalent:*

(a) *the polynomial $\sqrt{p}$ is a G-invariant ; and*

(b) $G \subset A_n$.

Proof : For $g \in G$ we have $g \cdot \sqrt{p} = \sqrt{p} \Leftrightarrow \operatorname{sgn}(g) = 1 \Leftrightarrow g \in A_n$. **q.e.d.**

Proposition 7.21 : *Let $\theta : \sigma \in S_n \mapsto p_\sigma \in \mathcal{P}_n$ be the group isomorphism introduced in the statement of* Proposition 7.9(b). *Then the diagram*

(i)
$$\begin{array}{ccc} S_n & \xrightarrow{\theta} & \mathcal{P}_n \\ & {\scriptstyle \operatorname{sgn}} \searrow \quad \swarrow {\scriptstyle \det} & \\ & \{1, -1\} & \end{array}$$

of group homomorphisms is commutative. In particular, for all $\sigma \in S_n$ one has

(ii)
$$\operatorname{sgn}(\sigma) = \det(p_\sigma).$$

Proof : Recall that a permutation $\tau \in S_n$ is a *transposition* if τ interchanges two elements of $\{1, 2, \ldots, n\}$ and leaves all others fixed. We assume readers are (or were at sometime) familiar with the standard fact that S_n is generated by transpositions, i.e., that any element of $\sigma \in S_n$ can be expressed (but not uniquely) as a finite product $\prod_{j=1}^{m} \tau_j$ in which the τ_j are transpositions. Since θ is an isomorphism, it follows that $p_\sigma = \prod_{j=1}^{m} p_{\sigma_j}$. However, since sgn and det are homomorphisms we have

$$\operatorname{sgn}(\sigma) = \textstyle\prod_j \operatorname{sgn}(\tau_j) \qquad \text{and} \qquad \det(p_\sigma) = \textstyle\prod_j \det(p_{\tau_j}),$$

and it therefore suffices to prove (ii) when $\sigma = \tau$ is a transposition. But in that case the result is obvious: the sign of any transposition τ is obviously -1, and the corresponding permutation matrix p_τ is obtained from the identity matrix by interchanging two columns, hence has determinant -1. **q.e.d.**

Examples 7.22 : Although the identification $\theta : S_n \to \mathcal{P}_n$ of Proposition 7.9(b) is completely standard, it is generally not exploited (and often not even mentioned) in introductory treatments of the symmetric group. The following two examples (particularly the second) indicate the flavor of a matrix approach to S_n.

(a) $n = 2$: We first list all (two!) permutations σ and the corresponding permutation matrices $\beta(\sigma)$.

the permutation σ	the corresponding permutation matrix p_σ	$\operatorname{sgn}(\sigma) = \det(p_\sigma)$
$\sigma_1 := e = \mathrm{id} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$	$p_{\sigma_1} = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$+1$
$\sigma_2 := \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$	$p_{\sigma_2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	-1

It follows immediately from the second column that $|\mathcal{P}_2| = |S_2| = 2$, hence that $\mathcal{P}_2$ is cyclic, with generator p_{σ_2}. From the third column and (ii) of Proposition 7.21 we see that the alternating group A_2 can be identified with the subgroup $\{I\} \subset \mathcal{P}_2$. and is therefore trivial.

(b) $n = 3$

the permutation σ	the corresponding permutation matrix p_σ	$\operatorname{sgn}(\sigma) = \det(p_\sigma)$
$\sigma_1 = e = \mathrm{id} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$	$p_{\sigma_1} = I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$+1$
$\sigma_2 := \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$	$p_{\sigma_2} := \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$	$+1$
$\sigma_3 := \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}$	$p_{\sigma_3} := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$	$+1$
$\sigma_4 := \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}$	$p_{\sigma_4} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	-1
$\sigma_5 := \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$	$p_{\sigma_5} := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$	-1
$\sigma_6 := \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}$	$p_{\sigma_6} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	-1

Here we see from the third column that $|\theta(A_3)| = |A_3| = 3$, and $\theta(A_3)$ is therefore cyclic, generated by each of p_{σ_2} and p_{σ_3}. Moreover, we see that $A_3 \subset S_3$ is the unique subgroup of order 3. Note that each of p_{σ_4}, p_{σ_5} and p_{σ_6} is of order 2, hence that $\mathcal{P}_3$, and therefore S_3, has three distinct subgroups of order 2. Also note, e.g., from $p_{\sigma_2} p_{\sigma_4} = p_{\sigma_6} \neq p_{\sigma_5} = p_{\sigma_4} p_{\sigma_2}$, that $\mathcal{P}_3$, and therefore S_3, is not abelian.

When $n \geq 2$ the group $\mathcal{P}_n$ has an interesting faithful representation.

Proposition 7.23 : *Assume $n \geq 2$, let $\zeta \in \mathbb{C}$ be a primitive n^{th}-root of unity, and let $V = V_n$ denote the $n \times n$ matrix with ij-entry $\zeta^{(i-1)(j-1)}$. Then the following assertions hold.*

(a) $V = (\mathrm{vdm}_{\mathbb{C}[x],n}(\zeta, \zeta^2, \ldots, \zeta^{n-1}))^\tau$. *In particular, V is the transpose of a non-singular Vandermonde matrix.*

(b) V^{-1} *is matrix $\frac{1}{n} \cdot W$, where W is the matrix obtained from V by replacing ζ with ζ^{n-1}.*

(c) *For any $p \in \mathcal{P}_n$ the matrix $V^{-1}pV$ is block diagonal with at least two blocks, the first* (i.e., that in the upper left corner) *being the 1×1 identify matrix.*

(d) *Let $\nu : \mathfrak{gl}(n, \mathbb{C}) \to \mathfrak{gl}(n-1, \mathbb{C})$ denote the mapping that removes the first row and column from any $n \times n$ matrix $M = (m_{ij})$* (i.e., $\nu(M)$ is the $(n-1) \times (n-1)$ matrix with ij-entry $m_{i+1,j+1}$). *Then the function $\kappa : p \in \mathcal{P}_n \mapsto \nu(V^{-1}pV) \in \mathrm{GL}(n-1, \mathbb{C})$ is a faithful determinant-preserving representation of $\mathcal{P}_n$.*

(e) *Let $\theta : S_n \to \mathcal{P}_n$ be the group isomorphism given in the statement of* Proposition 7.9(b). *Then the composition $\kappa \circ \theta : S_n \to \mathrm{GL}(n-1, \mathbb{C})$ is a faithful representation of S_n, and for any $\sigma \in S_n$ one has*

(i) $$\mathrm{sgn}(\sigma) = \det((\kappa \circ \theta)(\sigma)).$$

Notice that the mapping $\nu : \mathfrak{gl}(n, \mathbb{C}) \to \mathfrak{gl}(n-1, \mathbb{C})$ introduced in (d) is *not* a group homomorphism.

Proof :

(a) The initial assertion is trivial to check. The non-singularity follows from Proposition 1.9(c) and the assumption that ζ is primitive.

(b) From the assumption that ζ is a primitive root of $x^n - 1 = (x-1)(x^{n-1} + x^{n-1} + \cdots + x + 1) = 0$ we see that ζ^r, for every integer r, satisfies

(i) $$\sum_{k=0}^{n-1} (\zeta^r)^k = \begin{cases} 0 & \text{if } r \text{ is not divisible by } n, \text{ and} \\ n & \text{otherwise.} \end{cases}$$

The ij-entry of the product VM is therefore

$$\begin{aligned}\textstyle\sum_{k=1}^{n} \zeta^{(i-1)(j-1)}(\zeta^{n-1})^{(k-1)(j-1)} &= \textstyle\sum_k (\zeta^{k-1})^{[i-1+(n-1)(j-1)]} \\ &= \textstyle\sum_k (\zeta^{[(i-j)+n(j-1)]})^k \\ &= \begin{cases} 0 & \text{if } i \neq j \text{ and} \\ n & \text{otherwise.} \end{cases}\end{aligned}$$

(c) For $p = p_\sigma = (\delta_{i\sigma(j)})$ this ij-entry of pV is $\sum_k \delta_{i\sigma(k)}\zeta^{(\sigma(i)-1)(j-1)} = \zeta^{(i-1)(j-1)}$, and from (b) and (i) above we conclude that the ij-entry of $V^{-1}pV$ is

$$\begin{aligned}\tfrac{1}{n} \cdot \textstyle\sum_k (\zeta^{n-1})^{(i-1)(\sigma(k)-1)}\zeta^{(\sigma(k)-1)(j-1)} &= \tfrac{1}{n} \cdot \big(\textstyle\sum_k (\zeta^{n(i-1)+(j-i)})^{\sigma(k)}\big) \cdot \zeta^{n(i-1)+(j-i)} \\ &= \tfrac{1}{n} \cdot \big(\textstyle\sum_k (\zeta^{n(i-1)+(j-i)})^{k}\big) \cdot \zeta^{n(i-1)+(j-i)} \\ &= \begin{cases} 1 & \text{if } i = j = 1, \text{ and} \\ 0 & \text{if precisely one of } i \text{ and } j \text{ has value } 1. \end{cases}\end{aligned}$$

(d) Immediate from (c).

(e) By (d) and Proposition 7.21.

q.e.d.

Examples 7.24 : We continue with Examples 7.22, using the same ordering and notation.

(a) $n = 2$: Here $V = V_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and the only two permutation matrices are $p_{\sigma_1} = I$ and $p_{\sigma_2} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. One obviously has $V^{-1}p_{\sigma_1}V = p_{\sigma_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, and trivially verifies that $V^{-1}p_{\sigma_2}V = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. The faithful determinant-preserving representation $\nu : \mathcal{P}_2 \to \mathfrak{gl}(1, \mathbb{C}) \simeq \mathbb{C}$ is therefore given (not surprisingly!) by $p_{\sigma_1} \mapsto 1$ and $p_{\sigma_2} \mapsto -1$.

(b) $n = 3$: In this case we take $V = V^3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \lambda & \lambda^2 \\ 1 & \lambda^2 & \lambda^4 \end{bmatrix}$, where $\lambda := \frac{1}{2} \cdot (-1 + \sqrt{3}i)$. (One could also take λ to be the conjugate of this first choice, since both

are primitive cube roots of unity.) The following table is thereby obtained.

the permutation matrix p_σ	$V^{-1}p_\sigma V$	$\nu(p_\sigma) = (\kappa \circ \theta)(p_\sigma)$
$p_{\sigma_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$p_{\sigma_2} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \overline{\lambda} & 0 \\ 0 & 0 & \lambda \end{bmatrix}$	$\begin{bmatrix} \overline{\lambda} & 0 \\ 0 & \lambda \end{bmatrix}$
$p_{\sigma_3} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \overline{\lambda} \end{bmatrix}$	$\begin{bmatrix} \lambda & 0 \\ 0 & \overline{\lambda} \end{bmatrix}$
$p_{\sigma_4} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
$p_{\sigma_5} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & \lambda \\ 0 & \overline{\lambda} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & \lambda \\ \overline{\lambda} & 0 \end{bmatrix}$
$p_{\sigma_6} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & \overline{\lambda} \\ 0 & \lambda & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & \overline{\lambda} \\ \lambda & 0 \end{bmatrix}$

For this author the advantage of the final column over the others is that it makes it much easier to determine the nature of individual p_{α_j}. For example, from $\lambda^2 = \overline{\lambda}$, $\overline{\lambda}^2 = \lambda$ and $\lambda^3 = \overline{\lambda}^3 = 1$ one sees that both σ_2 and σ_3 generate the cyclic order-three subgroup $A_3 \subset S_3$, and that p_{σ_4}, p_{σ_5} and p_{σ_6} each have order two. Of course the advantages and disadvantages of particular viewpoints depend on which is more comfortable for a given individual: others might well prefer column one or two, or dropping representations completely and working directly with the original permutations.

8. Preliminaries on Semi-Direct Products

In this section groups are written multiplicatively, and the identity element of a group G will be written as e, or as e_G when confusion might otherwise result.

Let N and H be groups and let $\rho : H \to \mathrm{Aut}(N)$ be a representation. The action of H on N induced by ρ will be expressed by writing $\rho(h)(n)$ as $h \cdot n$. As the reader can easily check, a group structure on the Cartesian product $N \times H$, with identity (e_N, e_H), is given by

$$(n_1, h_1)(n_2, h_2) := (n_1(h_1 \cdot n_2), h_1 h_2), \qquad n_1, n_2 \in N, \qquad h_1, h_2 \in H. \tag{8.1}$$

When this structure is assumed $N \times H$ is called the *semi-direct product* of N and H, and is written $N \rtimes H$. Verification of the following properties of this construct are straightforward, and are left to the reader:

- The mapping $\iota : n \in N \mapsto (n, e_H) \in N \rtimes H$ is a group embedding.
- The same is true of the mapping $\sigma : h \in H \mapsto (e_N, h) \in N \rtimes H$.
- The projection mapping $\pi : (n, h) \in N \rtimes H \mapsto h \in H$ is a group homomorphism.
- the sequence

$$e_N \to N \xrightarrow{\iota} N \rtimes H \xrightarrow{\pi} H \longrightarrow e_H \tag{8.2}$$

  is split-exact: specifically, $\pi \circ \sigma = \mathrm{id}_H$.

The practical consequence of the first two bulleted items is that we can (and do) regard N and H as subgroups of $N \rtimes H$. A practical consequence of the final items is that $N \lhd N \rtimes H$, i.e., that N (when identified with $\iota(N)$) is a normal subgroup of this semi-direct product.

Examples 8.3 : In these examples K denotes a field.

(a) Let $N = K^\times$ be the multiplicative group of K and let H be the additive group of $\mathbb{Z}/2\mathbb{Z} = \{[0], [1]\}$. Then H acts on N by inversion, i.e., for $[n] \in \mathbb{Z}/2\mathbb{Z}$ and $k \in K^\times$ an action is well-defined by

(i) $$[n] \cdot k := k^{(-1)^n}.$$

The product of elements (k_1, g_1), $(k_1, g_2) \in K^\times \rtimes \mathbb{Z}/2\mathbb{Z}$, as defined in (8.1), is then given by

(ii) $\qquad (k_1, [n_1])(k_2, [n_2]) = (k_1([n_1] \cdot k_2), [n_1 + n_2]) = (k_1 k_2^{(-1)^{n_1}}, [n_1 + n_2]).$

with identity $e := (1, [n])$.

When $K = \mathbb{C}$ there is a simple way to formulate this semi-direct product in matrix terms[28]. Specifically, we claim that the mapping

(iii) $$\eta : (k, [n]) \in \mathbb{C}^\times \rtimes \mathbb{Z}/2\mathbb{Z} \mapsto \begin{bmatrix} k & 0 \\ 0 & k^{-1} \end{bmatrix} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^n \in \mathrm{SL}(2, \mathbb{C})$$

is a group embedding. Indeed, that $\eta(e) = I = I_2 \in \mathrm{SL}(2, \mathbb{C})$ is obvious, and the required multiplicative property can be seen from the identity

(iv) $$\begin{bmatrix} k & 0 \\ 0 & k^{-1} \end{bmatrix} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^n = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^n \begin{bmatrix} k^{(-1)^n} & 0 \\ 0 & k^{(-1)^{n+1}} \end{bmatrix},$$

which is is easily established by induction for all integers $n \geq 0$. Specifically, from (ii) and (iii) one has

$$\begin{aligned}
\eta((k_1, [n_1])(k_2, [n_2])) &= \eta((k_1 k_2^{(-1)^{n_1}}, [n_1 + n_2])) \\
&= \begin{bmatrix} k_1 k_2^{(-1)^{n_2}} & 0 \\ 0 & k_1^{-1} k_2^{(-1)^{n_2+1}} \end{bmatrix} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^{n_1+n_2} \\
&= \begin{bmatrix} k_1 & 0 \\ 0 & k_1^{-1} \end{bmatrix} \begin{bmatrix} k_2^{(-1)^{n_1}} & 0 \\ 0 & k_2^{(-1)^{n_1+1}} \end{bmatrix} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^{n_1} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^{n_2} \\
&= \begin{bmatrix} k_1 & 0 \\ 0 & k_1^{-1} \end{bmatrix} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^{n_1} \begin{bmatrix} k_2^{(-1)^{2n_1}} & 0 \\ 0 & k_2^{(-1)^{2n_1+1}} \end{bmatrix} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^{n_2} \\
&= \begin{bmatrix} k_1 & 0 \\ 0 & k_1^{-1} \end{bmatrix} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^{n_1} \begin{bmatrix} k_2 & 0 \\ 0 & k_2^{-1} \end{bmatrix} \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}^{n_2} \\
&= \eta((k_1, [n_1]))\eta((k_2, [n_2])).
\end{aligned}$$

[28] In fact one can replace $\mathbb{C}$ in this discussion by any field containing a square root of -1, e.g., $\mathbb{Q}(i)$

Note that (iii) can also be expressed as

$$\eta((k,[n]) = \begin{cases} \begin{bmatrix} k & 0 \\ 0 & k^{-1} \end{bmatrix} & \text{if } \ [n] = [0] \in \mathbb{Z}/2\mathbb{Z}, \\ \begin{bmatrix} 0 & ki \\ -k^{-1}i & 0 \end{bmatrix} & \text{if } \ [n] = [1] \in \mathbb{Z}/2\mathbb{Z}. \end{cases}$$

It is then immediate that η is an embedding, having as image the subgroup

$$\text{(v)} \qquad \left\{ \begin{bmatrix} k & 0 \\ 0 & k^{-1} \end{bmatrix} \right\}_{k \in \mathbb{C}^\times} \bigcup \left\{ \begin{bmatrix} 0 & \ell \\ -\ell^{-1} & 0 \end{bmatrix} \right\}_{\ell \in \mathbb{C}^\times} \subset \mathrm{SL}(2,\mathbb{C}).$$

(b) Let N and H be groups, and let H act on N trivially, i.e., define $h \cdot n = n$ for all $h \in H$ and all $n \in N$. Then (8.1) reduces to

$$(n_1, h_1)(n_2, h_2) = (n_1 n_2, h_1 h_2),$$

and the semi-direct product $N \rtimes H$ therefore coincides with the usual direct product $N \times H$ of N and H. In particular, the notion of a semi-direct product can be viewed as generalization of that of a direct product.

(c) Let $n \geq 1$ be an integer, let $\mathcal{D} = \mathcal{D}(n,K) \subset \mathrm{GL}(n,K)$ denote the subgroup of invertible diagonal matrices, and let $\mathcal{P}_n \subset \mathrm{GL}(n,K)$ denote the group of permutation matrices. Then $\mathcal{P}_n$ acts on $\mathcal{D}$ by conjugation, i.e., by $p \cdot d = pdp^{-1}$, and the groups therefore admit a semi-direct product. Specifically, the product of (d_1, p_1) and $(d_2, p_2) \in \mathcal{D} \rtimes \mathcal{P}_n$ as defined in (8.1) is here given by

(i)
$$(d_1, p_1)(d_2, p_2) = (d_1 p_1 d_2 p_1^{-1}, p_1 p_2) = (d_1 \tilde{d}_2, p_1 p_2), \quad \text{where} \quad \tilde{d}_2 := p_1 d_2 p_1^{-1} \in \mathcal{D},$$

and the identity element e is (I, I).

There is an obvious (set-theoretic) mapping from $\mathcal{D} \rtimes \mathcal{P}_n$ into $\mathrm{GL}(n,K)$, i.e.,

$$\text{(ii)} \qquad \eta : (d,p) \in \mathcal{D} \rtimes \mathcal{P}_n \mapsto dp \in \mathrm{GL}(n,K).$$

We claim that η is a group embedding. Indeed, that $\eta(e) = \eta((I,I)) = I \in \mathrm{GL}(n,K)$ is immediate from (ii), and for $(d_1,p_1),(d_2,p_2) \in \mathcal{D} \rtimes \mathcal{P}_n$ and $\tilde{d}_2 \in$

$\mathcal{D}$ as in (i) we have

$$\begin{aligned}\eta((d_1,p_1)(d_2,p_2)) &= \eta(d_1\tilde{d}_2, p_1p_2)\\ &= d_1\tilde{d}_2p_1p_2\\ &= d_1p_1d_2p_1^{-1}p_1p_2\\ &= d_1p_1d_2p_2\\ &= \eta((d_1,p_1))\eta((d_2,p_2)),\end{aligned}$$

thereby establishing the required group homomorphism property, and as a consequence the fact that the set of all matrices of the form dp is a subgroup of $\mathrm{GL}(n,K)$. To prove η is an embedding simply note that for $(d_1,p_1),(d_2,p_2) \in \mathcal{D} \rtimes \mathcal{P}_n$ one has

$$\eta((d_1,p_1)) = \eta((d_2,p_2)) \quad \Leftrightarrow \quad d_1p_1 = d_2p_2 \quad \Leftrightarrow \quad d_2^{-1}d_1 = p_2p_1^{-1}.$$

The desired equality $(d_1,p_1) = (d_2,p_2)$ is then immediate from $d_2^{-1}d_1 \in \mathcal{D}$, $p_2p_1^{-1} \in \mathcal{P}_n$ and $\mathcal{D} \cap \mathcal{P}_n = \{I\}$. In summary: $\mathcal{D} \rtimes \mathcal{P}_n$ can be identified with the subgroup

(iii) $$\mathcal{DP}_n := \{\, m \in \mathrm{GL}(n,K) \,:\, m = dp,\ d \in \mathcal{D} \text{ and } p \in \mathcal{P}_n \,\}$$

of $\mathrm{GL}(n,K)$. When $m \in \mathcal{DP}_n$ and $m = dp$ as in (iii) we refer to dp as the $\mathcal{DP}_n$-*decomposition* of m. This decomposition is unique (because η is injective). The $\mathcal{DP}_n$-decomposition of a product $d_1p_1 \cdot d_2p_2 \in \mathcal{DP}_n$ is achieved using the identity

(iv) $$d_1p_2d_2p_2 = d_1(p_1d_2p_1^{-1})p_1p_2.$$

We note that since the projection mapping $\pi : \mathcal{D} \rtimes \mathcal{P}_n \to \mathcal{P}_n$ of (8.2) is a group homomorphism, the same must be true of

(v) $$\gamma := \pi \circ \eta^{-1} : pd \in \mathcal{DP}_n \mapsto p \in \mathcal{P}_n.$$

This simple observation plays a crucial role in the next section.

When $n = 2$ and $K = \mathbb{C}$ the group $\mathcal{DP}_n$ is quite similar to that seen in (iv) of Example (a), except that union and the ambient set are now replaced by

$$\left\{\begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}\right\}_{k_1,k_2\in\mathbb{C}^\times} \bigcup \left\{\begin{bmatrix} 0 & \ell_1 \\ \ell_2 & 0 \end{bmatrix}\right\}_{\ell_1,\ell_2\in\mathbb{C}^\times} \subset \mathrm{GL}(2,\mathbb{C}).$$

(d) In the previous example one can replace $\mathcal{D}$ by any subgroup thereof which is stable under the action of $\mathcal{P}_n$. In particular, one can replace $\mathcal{D}$ by the matrix group $\mathcal{I} := \{I, -I\}$. In that case one sees that the image of the restriction to $\mathcal{I} \rtimes \mathcal{P}_n$ of the embedding η defined in (ii) of the previous example is the collection of all $n \times n$ matrices Q_n such that either $Q_n \in \mathcal{P}_n$ or $-Q_n \in \mathcal{P}_n$, and one thinks of the elements of this semi-direct product in terms of these images. Assuming this identification, note that each element of $\mathcal{I} \rtimes \mathcal{P}_n$ has determinant ± 1.

(e) The multiplicative order two matrix group

$$H := \left\{ \begin{bmatrix} -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \right\}$$

acts on $\mathrm{SL}(n, K)$, for any field K, by conjugation. (In fact one could generalize to [commutative] rings, e.g., one could take $K := \mathbb{Z}$.) As should now be evident from the preceding examples, one can identity the semi-direct product $\mathrm{SL}(n, K) \rtimes H$ with the collection of all matrices $m \in \mathrm{GL}(n, K)$ such that $\det(m) = \pm 1$. The group embedding $\kappa \circ \theta : S_n \to \mathrm{GL}(n-1, \mathbb{C})$ of Proposition 7.23(e) can then be described as a faithful matrix representation of S_n in $\mathrm{SL}(n-1, \mathbb{C}) \rtimes H$.

Of course H isomorphic to the multiplicative group $\{-1, 1\}$ and to the additive group $\mathbb{Z}/2\mathbb{Z}$. One might therefore express the faithful representation $\kappa \circ \theta$ as $\kappa \circ \theta : S_n \to \mathrm{SL}(2, K) \rtimes \{-1, 1\}$ or as $\kappa \circ \theta : S_n \to \mathrm{SL}(2, K) \rtimes \mathbb{Z}/2\mathbb{Z}$.

9. Faithful Matrix Representations of the Galois Group

We continue with the notation of §6. *In particular, we assume the linear operator* T *is separable. In addition, we fix a basis* $\mathbf{e}$ *for* V *and a fundamental* $\mathbf{e}$*-matrix* α *for* T.

The action of G on L induces an action of G on the collection $\mathfrak{gl}(n, L)$ of $n \times n$ matrices with entries in L: for any such matrix $m = (m_{ij})$ and any $g \in G$ define

$$g \cdot m := (g \cdot m_{ij}). \tag{9.1}$$

Since g is a field automorphism we have

$$\det(g \cdot m) = g \cdot \det(m). \tag{9.2}$$

Consequence: $\mathrm{GL}(n, L)$ and the subgroup $\mathcal{D} = \mathcal{D}(n, K)$ of invertible diagonal matrices are stable under this action. Since G fixes K pointwise, $\mathfrak{gl}(n, K)$ $(\subset \mathfrak{gl}(n, L))$ is also fixed pointwise by this action.

Let $\mathrm{diag}_L(n) \subset \mathfrak{gl}(n, L)$ denote the L-subalgebra of diagonal matrices. This is also stable under the action defined in (9.1).

Proposition 9.3 : *For any* $g \in G$ *the matrix* $g \cdot \alpha$ *is a fundamental matrix for* T. *Moreover,*

$$\text{(i)} \qquad g \cdot D_\alpha = D_{g \cdot \alpha}.$$

In this statement $D_{g\cdot\alpha}$ is the analogue of the diagonal matrix D_α defined in (6.1), i.e., it is given by

$$D_{g\cdot\alpha} := (g \cdot \alpha)^{-1} A (g \cdot \alpha).$$

However, the suggestive inclusion of "D" in this notation is not really justified until we know that $(g \cdot \alpha)^{-1} A (g \cdot \alpha)$ is diagonal; that is the point of the initial assertion of the proposition.

Proof : The calculation

$$\begin{aligned} A(g \cdot \alpha) &= (g \cdot A)(g \cdot \alpha) && \text{(because } A \in \mathfrak{gl}(n, K) \text{ and } g \text{ fixes } K) \\ &= g \cdot (A\alpha) && \text{(because } g \text{ is a field automorphism)} \\ &= g \cdot (\alpha D_\alpha) && \text{(by (6.1))} \\ &= (g \cdot \alpha)(g \cdot D_\alpha) && \text{(because } g \text{ is a field automorphism),} \end{aligned}$$

gives

$$g \cdot D_\alpha = (g \cdot \alpha)^{-1} A (g \cdot \alpha).$$

Since D_α is diagonal we see from (9.1) that the same holds for $g \cdot D_\alpha$, and the proposition follows. **q.e.d.**

The non-singular matrices[29]

$$\hat{P}_g := \alpha^{-1}(g \cdot \alpha), \qquad g \in G, \tag{9.4}$$

will play an important role in the sequel. One sees from

$$g \cdot \alpha = \alpha \hat{P}_g \tag{9.5}$$

that their introduction enables one to describe the action of G in terms of matrix multiplication.

Proposition 9.6 : *The matrices* $\hat{P}_g$ *defined in* (9.4) *have the following properties:*

(a) $\hat{P}_e = I$;

(b) $\hat{P}_g D_{g\cdot\alpha} = D_\alpha \hat{P}_g$;

(c) $(g \cdot \alpha)^{-1} = \hat{P}_g^{-1} \alpha^{-1}$;

(d) $g \cdot \alpha^{-1} = (g \cdot \alpha)^{-1}$;

(e) $g \cdot \alpha^{-1} = \hat{P}_g^{-1} \alpha^{-1}$;

(f) $\hat{P}_{gh} = \hat{P}_g (g \cdot \hat{P}_h)$;

(g) $\hat{P}_g^{-1} = g \cdot \hat{P}_{g^{-1}}$;

(h) *when* $\hat{P}_h$ *is fixed by* g *one has* $\hat{P}_{gh} = \hat{P}_g \hat{P}_h$;

(i) *when* $\hat{P}_h \in \mathrm{GL}(n, K)$ *one has* $\hat{P}_{gh} = \hat{P}_g \hat{P}_h$; *and*

(j) $\det(\hat{P}_g) = \dfrac{g \cdot \det(\alpha)}{\det(\alpha)}$.

[29] $\alpha^{-1}(g \cdot \alpha)$ denotes the product of the matrices α^{-1} and $g \cdot \alpha$.

Proof :

(a) Obvious.

(b) By (6.1) one has

$$\begin{aligned} P_g D_{g\cdot\alpha} &= \alpha^{-1}(g\cdot\alpha)D_{g\cdot\alpha} \\ &= \alpha^{-1}A\,(g\cdot\alpha) \\ &= \alpha^{-1}A\,\alpha\,\alpha^{-1}\,(g\cdot\alpha) \\ &= D_\alpha\,\alpha^{-1}(g\cdot\alpha) \\ &= D_\alpha P_g. \end{aligned}$$

(c) Immediate from (9.5).

(d) $I = g\cdot I = g\cdot(\alpha^{-1}\alpha) = (g\cdot\alpha^{-1})(g\cdot\alpha)$ and, similarly, $I = (g\cdot\alpha)(g\cdot\alpha^{-1})$. The result follows.

(e) By (c) and (b).

(f) One has

$$\begin{aligned} \alpha\hat{P}_{gh} &= gh\cdot\alpha \qquad\qquad \text{(by (9.5))} \\ &= g\cdot(h\cdot\alpha) \\ &= g\cdot(\alpha\hat{P}_h) \\ &= (g\cdot\alpha)(g\cdot\hat{P}_h) \\ &= (\alpha\hat{P}_g)(g\cdot\hat{P}_h), \end{aligned}$$

and the result follows.

(g) By choosing $h = g^{-1}$ in (f) we see from (a) that

$$I = \hat{P}_e = \hat{P}_{gg^{-1}} = \hat{P}_g(g\cdot\hat{P}_{g^{-1}}),$$

and the result follows.

(h) Immediate from (f).

(i) Since $\mathrm{GL}(n,K)$ is pointwise fixed by g, this is a special case of (h).

(j) By (9.2) we have $g\cdot\det(\alpha) = \det(g\cdot\alpha) = \det(\alpha\hat{P}_g) = \det(\alpha)\det(\hat{P}_g)$, and $\alpha \neq 0$ since α is invertible. The equality follows.

q.e.d.

For the remainder of the section we let $\mathcal{D}$, $\mathcal{P}_n$ *and* $\mathcal{DP}_n$ *denote the subgroups of* $\mathrm{GL}(n,K)$ *introduced in* Example 8.3(c), *and we let* $\eta : (d,p) \in \mathcal{D} \rtimes \mathcal{P}_n \mapsto dp \in \mathcal{DP}_n$ *denote the group isomorphism defined in* (ii) *of that example.*

Proposition 9.7 : *For each* $g \in G$ *let* $\hat{P}_g$ *be the matrix defined in* (9.4), *i.e., the unique matrix satisfying*

(i) $$\alpha \hat{P}_g = g \cdot \alpha.$$

Then the following assertions hold.

(a) *Each column of* $g \cdot \alpha$ *is a scalar multiple of some column of* α.

(b) $\hat{P}_g \in \mathcal{DP}_n$.

(c) $\hat{P}_g \in \mathcal{D}$ *if and only if* $g = e\ (= e_G)$.

(d) $\hat{P}_g = \hat{P}_h$ *if and only if* $g = h$.

Proof :

(a) By Proposition 6.3(b) all columns in question are **e**-columns of eigenvectors of T, and by the separability assumption on T the corresponding eigenspaces are one-dimensional.

(b) By (a) there is a permutation matrix $\hat{p}$ such that column j of $(g\cdot\alpha)\hat{p}$, for any $1 \le j \le n$, is of the form $d_j\alpha_j$, where α_j denotes the j^{th}-column of α. By defining $d := \mathrm{diag}_L(d_1, d_2, \ldots, d_n) \in \mathcal{D}$ we can express this in matrix form as $(g \cdot \alpha)\hat{p} = \alpha d$, and we conclude that for $p := \hat{p}^{-1} \in \mathcal{P}_n$ we have $\hat{P}_g = \alpha^{-1}(g \cdot \alpha) = dp \in \mathcal{DP}_n$.

(c) $\Rightarrow$: If $g \neq e$ then at least one eigenvalue of T_L must be moved (i.e., is not fixed) by g (because L is generated over K by these eigenvalues, and K is fixed by G). It then follows from Corollary 5.5 and Proposition 6.3(b) that g must move the corresponding column of α, and this contradicts the diagonal hypothesis on $\hat{P}_g$.

$\Leftarrow$: It is obvious from (9.4) that $\hat{P}_e = I \in \mathcal{D}$.

(d) One has

$$\begin{aligned} \hat{P}_g = \hat{P}_h &\Leftrightarrow \alpha^{-1}(g \cdot \alpha) = \alpha^{-1}(h \cdot \alpha) \\ &\Leftrightarrow g \cdot \alpha = h \cdot \alpha \\ &\Leftrightarrow (h^{-1}g) \cdot \alpha = \alpha. \end{aligned}$$

By (a) and separability (which guarantees one-dimensional eigenspaces) this last equality holds if and only if the element $h^{-1}g \in G$ does not move any eigenvectors of T. By Corollary 5.5 this is equivalent to $h^{-1}g = e$, hence to $g = h$.

q.e.d.

By Proposition 9.7(d) the mapping

$$\hat{\rho} = \hat{\rho}_\alpha : g \in G \mapsto \hat{P}_g \in \mathcal{DP}_n \tag{9.8}$$

is an injection. Given the title of this section, one might expect $\hat{\rho}$ to be the relevant matrix representation. Unfortunately, one sees from Proposition 9.6(f) that it need not be a group homomorphism. One can also see this last assertion in a more concrete way.

Example 9.9 : Consider the linear operator $T : \mathbb{Q}^2 \to \mathbb{Q}^2$ with usual basis matrix $A = \begin{bmatrix} 0 & -4 \\ 1 & 0 \end{bmatrix}$. Here

$$\alpha := \begin{bmatrix} 2(1+i) & 2(1+i) \\ -1+i & 1-i \end{bmatrix}$$

is a fundamental matrix of A, as the reader can easily check. The splitting field of the characteristic polynomial $x^2 + 4 \in \mathbb{Q}[x]$ is $\mathbb{Q}(i) \subset \mathbb{C}$, from which we see that the classical Galois group is of order 2 and is generated by the restriction to $\mathbb{Q}(i)$ of complex conjugation. Denoting that particular automorphism by g one sees that

$$g \cdot \alpha = \begin{bmatrix} 2(1-i) & 2(1-i) \\ -1-i & 1+i \end{bmatrix},$$

and then that

$$\hat{P}_g = \alpha^{-1}(g \cdot \alpha) = \begin{bmatrix} \frac{1}{8}(1-i) & -\frac{1}{4}(1+i) \\ \frac{1}{8}(1-i) & \frac{1}{4}(1+i) \end{bmatrix} \cdot \begin{bmatrix} 2(1-i) & 2(1-i) \\ -1-i & 1+i \end{bmatrix} = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}.$$

But this last matrix has order 4, whereas if $\hat{\rho} : g \in G \to \hat{P}_g \in \mathcal{DP}_n$ were a group homomorphism the order would be at most 2, i.e., at most that of g.

As we will now see, the problem with $\hat{\rho} : G \to \mathcal{DP}_n$ not being a group homomorphism is easily overcome by combining the group identification $\eta : (d, p) \in \mathcal{D} \rtimes \mathcal{P}_n \mapsto dp \in \mathcal{DP}_n$ introduced in (ii) of Example 8.3(c) with the group homomorphism

$$\gamma = \pi \circ \eta^{-1} : dp \in \mathcal{DP}_n \mapsto p \in \mathcal{P}_n \tag{9.10}$$

defined in (v) of that example.

Example 9.11 : In Example 9.9 we encountered the matrix $\hat{P}_g = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}$. By (b) of Proposition 9.7 this matrix must belong to $\mathcal{DP}_n$, and therefore must be uniquely expressible in the form dp, where $d \in \mathcal{D}$ and $p \in \mathcal{P}_n$. In fact, as one sees by inspection, that unique decomposition is

$$\text{(i)} \qquad \hat{P}_g = \begin{bmatrix} -i & 0 \\ 0 & -i \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

from which we see that

$$\text{(ii)} \qquad \gamma(\begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

More generally, the result of evaluating γ on a matrix in $\mathcal{DP}_n$ can be computed simply by replacing each non-zero entry of that matrix with 1: equality (ii) of Example 9.11 is typical.

To tie the work in this section together define $\rho : G \to \mathcal{P}_n$ by

$$\rho = \rho_\alpha = \gamma \circ \hat{\rho}. \tag{9.12}$$

(One should keep in mind that $\hat{\rho} = \hat{\rho}_\alpha$, i.e., that $\hat{\rho}$ depends on α.) From the preceding comment we see that, for each $g \in G$, $P_g := \rho(g)$ can be computed directly from $\hat{P}_g = \hat{\rho}(g)$ simply by replacing each non-zero entry with 1. With more formality: express $\hat{P}_g = \hat{\rho}(g)$ in the (unique) $\mathcal{DP}_n$-decomposition form $D_g P_g$, where $D_g \in \mathcal{D}$ and $P_g \in \mathcal{P}_n$, and assign g to P_g. Either approach achieves the same result:

$$\rho : g \in G \mapsto P_g \in \mathcal{P}_n\,. \tag{9.13}$$

Example 9.14 : For g as in Example 9.11 we see from (i) of that example that $\begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix} = \hat{P}_g = D_g P_g$, where $D_g = \begin{bmatrix} -i & 0 \\ 0 & -i \end{bmatrix} \in \mathcal{D}$ and $P_g = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \in \mathcal{P}_n$. Thus $\rho(g) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Example 9.14 suggests that one will need to compute the $\mathcal{DP}_n$-decomposition of each $\hat{P}_g$ in every application. As we now show, this can be avoided by a judicious choice of the fundamental matrix.

Theorem 9.15 : *Assume the context of* Corollary 6.8(c)*; in particular, choose* α *as in* (vii) *of that example. Then for any* $g \in G$ *the following assertions hold.*

(a) *The* $\mathcal{DP}_n$*-decomposition* $\hat{P}_g = D_g P_g$ *of* $\hat{P}_g \in \mathcal{DP}_n$ *is given by*

(i) $$\mathcal{D}_g = I \qquad \text{and} \qquad P_g = \hat{P}_g.$$

(b) *If* $\sigma = \sigma(g) \in S_n$ *is the permutation defined by*

(ii) $$g \cdot \lambda_j = \lambda_{\sigma(j)}, \qquad j = 1, 2, \ldots, n,$$

then

(iii) $$P_g = \theta(\sigma) = p_\sigma,$$

where $\theta : \sigma \in S_n \mapsto p_\sigma \in \mathcal{P}_n$ *is the group isomorphism of* Proposition 7.9(b). *In particular, one can recover the permutation* σ*, and as a consequence the permutation of the eigenvalues of* T *induced by* g*, from*

(iv) $$P_g^\tau \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix} = \begin{bmatrix} \sigma(1) \\ \sigma(2) \\ \vdots \\ \sigma(n) \end{bmatrix}.$$

The conditions given in (i) are essentially equivalent to the assertion that $\rho = \hat{\rho}$. More precisely, they guarantee that one has a commutative diagram

(v) $$\begin{array}{ccc} & & \mathcal{DP}_n \\ & \overset{\hat{\rho}}{\nearrow} & \\ G & & \uparrow \\ & \underset{\rho}{\searrow} & \\ & & \mathcal{P}_n \end{array}$$

in which the vertical arrow represents the mapping $p \in \mathcal{P}_n \mapsto Ip \in \mathcal{DP}_n$.

Proof : By Corollary 5.5 the Galois group permutes eigenpairs of T and, by definition, fixes the ground field K. Since by Proposition 6.3 the columns α_j of α

are the **e**-columns of a T_L-eigenbasis, one sees from the structure of α that g will permute these columns according to

$$g \cdot \alpha_j = \alpha_{\sigma(j)}, \qquad j = 1, 2, \ldots, n.$$

We claim that right multiplication of α by p_σ achieves the same result. Indeed, from (7.1) we have

$$\alpha p_\sigma = (\textstyle\sum_k \alpha_{ik}\delta_{k\sigma(j)}) = (\alpha_{i\sigma(j)}),$$

and the claim follows. This gives $g \cdot \alpha = \alpha p_\sigma$, whereupon from (9.4) we see that $\hat{P}_g = \alpha^{-1}(g \cdot \alpha) = p_\sigma = Ip_\sigma$. In view of the uniqueness of $\mathcal{DP}_n$-decompositions, this establishes both (i) and (iii). For the final assertion recall Proposition 7.8. **q.e.d**.

We can now present our main result. It is the analogue for separable linear operators of Theorem 1.38(c) for differential structures.

Theorem 9.16 : *The mapping* $\rho : G \to \mathcal{P}_n$ *defined in* (9.12) *is a faithful matrix representation.*

Proof : Let $e \in G$ denote the identity automorphism and let I denote the $n \times n$ identity matrix. The equalities $\hat{\rho}(e) = (I, I)$ and $\pi(I, I) = I$ then give $\rho : e \mapsto I$ as required. Now let $g_1, g_2 \in G$ and write $\hat{P}_{g_j} = \mathcal{D}_{g_j}P_{g_j} \simeq (\mathcal{D}_{g_j}, P_{g_j}) \in \mathcal{D} \rtimes \mathcal{P}_n$, $j = 1, 2$. Then

$$g_1 \cdot \hat{P}_{g_2} = g_1 \cdot (\mathcal{D}_{g_2}P_{g_2}) = (g_1 \cdot \mathcal{D}_{g_2})(g_1 \cdot P_{g_2}) = (g_1 \cdot \mathcal{D}_{g_2})P_{g_2},$$

the final equality from Proposition 9.6(i) and the fact that $P_{g_2} \in \mathrm{GL}(n, K)$. Since $\ker(\pi) = \mathcal{D}$, and since $g_1 \cdot \mathcal{D}_{g_2} \in \mathcal{D}$ (by (9.1)), it follows that

$$\text{(i)} \qquad \gamma(g_1 \cdot \hat{P}_{g_2}) = P_{g_2}.$$

Using the fact that γ is a homomorphism we conclude from Proposition 9.6(f) that

$$\begin{aligned}
\rho(g_1g_2) &= \gamma(\hat{\rho}(g_1g_2)) \\
&= \gamma(\hat{P}_{g_1g_2}) \\
&= \gamma(\hat{P}_{g_1}(g_1 \cdot \hat{P}_{g_2})) \\
&= \gamma(\hat{P}_{g_1}) \cdot \gamma(g_1 \cdot \hat{P}_{g_2}) \\
&= P_{g_1} \cdot P_{g_2} \quad \text{(by (i))} \\
&= \rho(g_1)\rho(g_2),
\end{aligned}$$

and the homomorphism structure of ρ is thereby verified.

To prove injectivity suppose $\rho(g) = P_g = I$. Since $\hat{P}_g = \mathcal{D}_g P_g$, this gives $\hat{P}_g = \mathcal{D}_g \in \mathcal{D}$, and we conclude from Proposition 9.7(c) that $g = e$. **q.e.d.**

In the following statement $\theta : S_n \to \mathcal{P}_n$ denotes the isomorphism introduced in Proposition 7.9(b).

Corollary 9.17 : *The composition* $\beta := \theta^{-1} \circ \rho : G \to S_n$ *is a group embedding.*

In the next result the function $\kappa : \mathrm{GL}(n, \mathbb{C}) \to \mathrm{GL}(n-1, \mathbb{C})$ is that introduced in Proposition 7.23(d), and H is the order two matrix group introduced in Example 8.3(e).

Corollary 9.18 : *The mapping* $\kappa \circ \rho : G \to \mathrm{SL}(n-1, \mathbb{C}) \rtimes H$ *is a faithful matrix representation.*

Proof : View the mapping $\kappa \circ \rho$ as the sequence

$$G \xrightarrow{\rho} \mathcal{P}_n \xrightarrow{\theta^{-1}} S_n \xrightarrow{\kappa \circ \theta} \mathrm{SL}(n-1, \mathbb{C}) \rtimes H$$

and recall from Example 8.3(e) that the mapping $\kappa \circ \theta$ is a group embedding. **q.e.d.**

In Corollary 9.17 we have a concrete illustration of Cayley's Theorem: *every finite group can be embedded into* S_n *for some positive integer* n. Because the symmetric groups S_n become so complicated with increasing n Cayley's result is regarded for the most part as a curiosity. In contrast, we will make use of the embedding β.

Proposition 9.19 : *For any* $g \in G$ *one has* $\beta(g) = \sigma$, *where* $\sigma \in S_n$ *is defined by*

(i) $$g \cdot \lambda_j = \lambda_{\sigma(j)}, \qquad j = 1, 2, \ldots, n.$$

Moreover,

(ii) $$P_g = p_\sigma \, .$$

Proof : Equality (ii) is a restatement of (iii) of Theorem 9.15. To establish (i) choose any $g \in G$ and any $1 \leq j \leq n$. Then

$$\begin{aligned} \beta(g)(j) &= \theta^{-1}(\rho(g))(j) \\ &= \theta^{-1}(P_g)(j) \\ &= \theta^{-1}(p_\sigma)(j) \qquad \text{(by (ii))} \\ &= \sigma(j), \end{aligned}$$

and (i) is thereby established. **q.e.d.**

Corollary 9.20 : *Let* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *denote the eigenvalues of* T *and let* V *denote the matrix*

(i) $$V = (v_{ij}) = (\lambda_j^{n-i}).$$

Then for any $g \in G$ *one has*

(ii) $$g \cdot V = V p_{\beta(g)}$$

and, as a consequence

(iii) $$g \cdot \det(V) = \operatorname{sgn}(\beta(g)) \det(V).$$

Proof : By Proposition 9.19 one has

$$\begin{aligned} g \cdot V &= (g \cdot \lambda_j^{n-i}) \\ &= ((g \cdot \lambda_j)^{n-i}) \\ &= ((\lambda_{\sigma(j)})^{n-i}) \\ &= (\lambda_j^{n-i}) \\ &= (\textstyle\sum_k \lambda_k^{n-i} \delta_{k\sigma(j)}) \\ &= V p_\sigma \\ &= V p_{\beta(g)}, \end{aligned}$$

and this gives (ii).

Equality (iii) is a restatement of (7.18). **q.e.d.**

We are now in a position to make some quite general, but sometimes useful, remarks about the Galois group G of T. In the statement we view G as a subgroup of S_n by means of the embedding of Corollary 9.17.

Corollary 9.21 :

(a) $|G| \big| n!$.

(b) *Let* $\lambda_1, \lambda_2, \ldots, \lambda_n \in L$ *denote the eigenvalues of* T *and let*

(i) $$\sqrt{\Delta} := \textstyle\prod_{i<j} (\lambda_i - \lambda_j) \in L.$$

Then $G \subset A_n$ *if and only if* $\sqrt{\Delta} \in K$.

(c) *$G \subset A_n$ if and only if the discriminant Δ of $\mathrm{char}_{T,K}(x) \in K[x]$ is a perfect square in K, i.e., if and only if there is an element $k \in K$ such that $\Delta = k^2$.*

(d) *When the characteristic polynomial $\mathrm{char}_{T,K}(x) \in K[x]$ is irreducible it must be the case that $n \big| |G|$.*

(e) *When $\mathrm{char}_{T,K}(x)$ is irreducible one has both $|G| \big| n!$ and $n \big| |G|$.*

(f) *When $n = 2$ and $\mathrm{char}_{T,K}(x)$ is irreducible the Galois group of T is of order 2 and*

$$\rho(G) = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}.$$

(g) *When $n = 3$ and $\mathrm{char}_{T,K}(x) \in K[x]$ is irreducible the Galois group of T is S_3 if and only if Δ is not a perfect square in K, and is otherwise A_3, in which case $\rho(G) \subset \mathcal{P}_n$ is cyclic with generator*

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

Keep in mind that T, and therefore $\mathrm{char}_{T,K}(x)$, is always assumed separable in this section[30].

Recall from (i) of Example 4.5 that the discriminant Δ of $\mathrm{char}_{T,K}(x)$ is given by

(ii) $$\Delta := \Big(\textstyle\prod_{i<j} (\lambda_i - \lambda_j) \Big)^2,$$

which explains the notation introduced in (i) above.

Although (b) and (c) give equivalent conditions for $G \subset A_n$, there are practical difficulties with that given in (b) since one cannot compute $\sqrt{\Delta}$ (as defined) without knowing the roots of $\mathrm{char}_{T,K}(x)$. However, one can compute Δ without knowing these roots explicitly, so this difficulty is eliminated.

[30] Assertion (g) is usually stated under the assumption that $\mathrm{char}(K) \neq 2, 3$, (e.g., see [Lang, Chapter VI, §2, Example 2, p. 270]), but that is only to ensure the separability hypothesis.

Proof :

(a) Since $|S_n| = n!$, this is immediate from Lagrange's theorem and Corollary 9.17.

(b) From (iii) of Corollary 9.20 we see that

$$\text{(iii)} \qquad g \cdot \sqrt{\Delta} = \operatorname{sgn}(\beta(g))\, \sqrt{\Delta},$$

hence $G\ (\simeq\ \beta(G))\ \subset A_n$ if and only if $g \cdot \sqrt{\Delta} = \sqrt{\Delta}$ for all $g \in G$. Since K is the fixed field of G, this condition is equivalent to $\sqrt{\Delta} \in K$.

(c) If we define $\ell := \sqrt{\Delta} \in L$ then the roots of $x^2 - \Delta \in K[x]$ are given by $\pm\ell$, and both are in K if and only if either one is in K. But $\ell \in K$ if and only if $g \cdot \ell = \ell$ for all $g \in G$, which by (iii) is equivalent to $\operatorname{sgn}(\beta(g)) = 1$ for all such g. Assertion (c) follows.

(d) By hypothesis $\operatorname{char}_{T,K}(x)$ must be the irreducible polynomial of any particular eigenvalue λ of T, and $L \supset K(\lambda) \supset K$ is therefore a tower of fields. Assertion (d) is then immediate from (the well-known fact that) $[K(\lambda); K] = n$ and the multiplicative property $[L; K] = [L; K(\lambda)][K(\lambda); K]$ of indices.

(e) By (a) and (d).

(f) By (e) we have $|G| = 2$, and the indicated subgroup of $\mathcal{P}_2$ is the only subgroup having that order.

(g) By (e) we have $|G| \big| 6$ and $3 \big| |G|$, hence $|G| = 6$ or 3, and since A_3 is the only subgroup of S_3 of order 3 it follows that $G = S_3$ or $G = A_3$. From (c) we then see that $G = S_3$ if and only if Δ is not a perfect square in K, and that G is otherwise A_3.

q.e.d.

Examples 9.22 : In all these examples $T : \mathbb{Q}^n \to \mathbb{Q}^n$ is the linear operator having the indicated matrix A as usual basis matrix.

(a) $(n = 2)$: $A = \begin{bmatrix} 0 & -4 \\ 1 & 0 \end{bmatrix}$. In Example 9.9 we found that

$$\alpha := \begin{bmatrix} 2(1+i) & 2(1+i) \\ -1+i & 1-i \end{bmatrix}$$

is a fundamental matrix for T. that the order two Galois group is obtained by restricting complex conjugation to the splitting field $\mathbb{Q}(i)$ of the $\operatorname{char}_{T,\mathbb{Q}}(x)$,

and if g represents the generator of this group then

$$\hat{P}_g = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}.$$

One now sees from Example 9.11 that

(i) $$\rho(g) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The full faithful representation $\rho : G \to \mathcal{P}_n$ is therefore given by (i) and

(ii) $$\rho(e) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

(b) $(n = 2)$: We can both generalize and simplify the previous example by using Corollary 6.8 (which we take to include the comments following that statement). Specifically, we begin with the transpose $A = \begin{bmatrix} 0 & 1 \\ -c & -b \end{bmatrix}$ of a 2×2 matrix in rational form and assume the discriminant $\Delta := b^2 - 4c$ of $\mathrm{char}_{T,\mathbb{Q}}(x) = x^2 + bx + c \in \mathbb{Q}[x]$ is not a perfect square in $\mathbb{Q}$. The hypothesis guarantees that the two roots λ_1, λ_2 generating the splitting field $L \supset \mathbb{Q}$ are distinct and not in $\mathbb{Q}$; hence that the index $[L; \mathbb{Q}]$ and the order of the Galois group G are 2. Let g be the generator of G. For our fundamental matrix we (know from Corollary 6.8 that we may) choose the Vandermonde matrix

$$\alpha := \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix}$$

as our fundamental matrix. Since g permutes the roots we must have

$$g \cdot \alpha = \begin{bmatrix} 1 & 1 \\ \lambda_2 & \lambda_1 \end{bmatrix},$$

and from Theorem 9.15(a) we conclude that

$$\hat{P}_g = P_g = \alpha^{-1}(g \cdot \alpha) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The full faithful matrix representation can therefore be expressed exactly as in (i) and (ii) of the previous example (which is hardly a surprise).

(c) $(n = 3)$: We continue with the ideas surrounding Corollary 6.8, now taking

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -d & -c & -b \end{bmatrix},$$

and assuming that the discriminant $\Delta = -27d^2 - 4c^3 + (bc^2 + 18dc - 4b^2d) \cdot b$ is not zero. This restriction on Δ guarantees only that the three roots λ_1, λ_2 and λ_3 of the characteristic polynomial $\mathrm{char}_{T,\mathbb{Q}}(x) = x^3 + bx^2 + cx + d \in \mathbb{Q}[x]$ are distinct: it does not guarantee that all lie outside $\mathbb{Q}$. (We could guarantee the last condition by requiring $\mathrm{char}_{T,\mathbb{Q}}(x) \in \mathbb{Q}[x]$ to be irreducible, but at this point there is no reason to do so.) For our fundamental matrix we choose the Vandermonde matrix

$$\alpha := \begin{bmatrix} 1 & 1 & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 \end{bmatrix}.$$

We now focus on the structure of the associated Galois group G, using Theorem 9.16 to identify G with a subgroup of S_3. For any $g \in G$ we have, in the notation seen in (i) of Proposition 9.19,

$$g \cdot \alpha := \begin{bmatrix} 1 & 1 & 1 \\ \lambda_{\sigma(1)} & \lambda_{\sigma(2)} & \lambda_{\sigma(3)} \\ \lambda_{\sigma(1)}^2 & \lambda_{\sigma(2)}^2 & \lambda_{\sigma(3)}^2 \end{bmatrix}.$$

With a bit of work one can then show that

$$\alpha^{-1}(g \cdot \alpha) = \left((-1)^{i+1} \frac{\prod_{k \neq i}(\lambda_k - \lambda_{\sigma(j)})}{\prod_{k \neq i}(\lambda_k - \lambda_i)} \right).$$

and that

$$\det(\alpha^{-1}(g \cdot \alpha)) = \mathrm{sgn}(g).$$

We have therefore created a list containing all possibilities for the Galois group, but, as we see from Corollary 9.21(e), some of these possibilities might not occur.

Thus far we have only computed Galois groups for separable operators on spaces of dimensions two and three. The following observation suggests how one might try

to tackle higher dimensions, i.e., determine the invariants. Indeed, it has been long known that the invariants of finite groups determine these groups (up to isomorphism), and this is one of the (many) reasons why the search for invariants was a major mathematical industry in the late nineteenth century.

Proposition 9.23 : *Let* $\lambda_1, \lambda_2, \ldots, \lambda_n \in L$ *be a complete set of eigenvalues of* T *and set* $\lambda := (\lambda_1, \lambda_2, \ldots, \lambda_n) \in L^n$. *Then the following statements are equivalent:*

(a) $\Delta := \left(\prod_{i>j}(\lambda_i - \lambda_j)\right)^2$ *is a perfect square in* K;

(b) $\sqrt{\Delta} := \prod_{i>j}(\lambda_i - \lambda_j) \in K$;

(c) $\mathrm{vdmd}_{K[x],n}(\lambda) \in K$;

(d) *the Galois group* G *of* T *is contained in* A_n (i.e., can be identified via ρ with a subgroup of A_n); *and*

(e) *the polynomial* $\sqrt{p} := \prod_{i>j}(x_i - x_j) \in K[x_1, x_2, \ldots, x_n]$ *is* G*-invariant under the action defined in* (7.11).

Proof :

(a) $\Leftrightarrow$ (b) : Obvious.

(b) $\Leftrightarrow$ (c) : By Proposition 1.9(a).

(b) $\Leftrightarrow$ (d) : This is a restatement of Corollary 9.21(b).

(d) $\Leftrightarrow$ (e) : This is a restatement of Corollary 7.20.

q.e.d.

Notes and Comments

The best (and most complete) reference for differential Galois theory is [vdP-S]. Reference [Poole] was my basic source for information on linear differential operators.

The attribution appearing in Proposition 1.18(g) to Abel and Liouville is from [Poole, Chapter I, §5, p. 13].

Chapter 7 of [H-K] is a good reference for cyclic vectors, but only over the real and complex fields. I found [M-B] particularly useful in connection with this topic.

Acknowledgment

These notes were originally prepared for a lecture in the Algebra Seminar of the Department of Mathematics at the University of Calgary in July of 2010. I thank Dr. Peter Zvengrowski of that department for organizing that talk.

References

[B-M] G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, Third Edition, MacMillan, New York, 1965.

[C] R.C. Churchill, *A Geometric Approach to Linear Ordinary Differential Equations*, posted on the website of the Kolchin Seminar on Differential Algebra (http://www.sci.ccny.cuny.edu), New York, 2006.

[C-K] R.C. Churchill and J. Kovacic. Cyclic Vectors, in *Differential Algebra and Related Topics*, Li Guo, P. Cassidy, W. Keigher and W. Sit, eds., World Scientific, Singapore, 2002.

[E] D. Eisenbud, *Commutative Algebra, with a view toward Algebraic Geometry*, GTM 150, Springer-Verlag, New York, 1995.

[H-K] K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1961.

[J] N. Jacobson, *Lectures in Abstract Algebra, Volume II - Linear Algebra*, Van Nostrand Reinhold, New York, 1953.

[Lev] A.H.M. Levelt, Differential Galois Theory and Tensor Products, *Indag. Math.,* N.S. **1** (4) (1990), 439-450.

[M-B] S. MacLane and G. Birkhoff, *Algebra*, Macmillan, New York, 1967.

[N] K. Nomizu, *Fundamentals of Linear Algebra*, McGraw-Hill, New York, 1966.

[Lang] S. Lang, *Algebra*, Revised Third Edition, GTM 211, Springer, New York, 2002.

[Poole] E.G.C. Poole, *Introduction to the Theory of Linear Differential Equations*, Dover, New York, 1960.

[vdP-S] M. van der Put and M.F. Singer, *Galois theory of Linear Differential Equations*, Springer, Berlin, 2003.

R. Churchill
Department of Mathematics and Statistics
Hunter College and the Graduate Center,
City University of New York, and the
University of Calgary
e-mail rchurchi@hunter.cuny.edu
August, 2010