Miscellaneous Short Topics
The Essential Uniqueness of Determinants, Change of Basis, the Structure of Orthogonal Real Matrices, and the Orthogonal Group.

I. The essential uniqueness of determinants, and determinant of product = product of determinants.

Lemma (essential uniqueness): Suppose
D: $n \times n$ matrices, entries in a field $F \to F$
is a function satisfying

(1) antisymmetry: the interchange of two columns of a matrix A to give $\hat{A}$ satisfies
$D(\hat{A}) = -D(A)$

(2) D is linear in each column, i.e.

$$\begin{pmatrix} \mu_1 & | & | & | & | \\ \vdots & | & | & | & | \\ \mu_n & | & | & | & | \end{pmatrix}$$ is a linear function of the

column vector, the remaining columns being fixed [by (1) it is enough to assume this for the first column!]. Then $\exists\ \lambda_0 \in F$ such that for all A, $n \times n$ F-entry matrices,

$$D(A) = \lambda_0 \det(A)$$

where det is the determinant function defined earlier $\left(= \sum_{\pi} (-1)^{\pi} a_{1\pi(1)} \cdots a_{n\pi(n)} \right)$.

Proof: If A has linearly dependent columns then $\det(A)=0$. Also, in this case, $D(A)=0$. The reason is that by (1), we can suppose wolog that the first column of A = a linear combination of the 2nd through nth columns:

$$\text{col } 1 = \sum_{j=2}^{n} \alpha_j \text{ col}_j$$

Then $\overset{\text{by (2)}}{} D(A) = \sum_{j=2}^{n} \alpha_j D\left(\text{col}_j \quad \text{other columns as in } A \text{ itself}\right)$

$$\underset{\text{from (1)}}{=} \sum \alpha_j \cdot 0 = 0$$

It remains to consider the case that A has linearly independent columns ($\Longleftrightarrow$ A is invertible or "nonsingular"). Since $D(A)$ and $\det(A)$ both reverse sign on interchange of columns, we can assume wolog that $A_{11} \neq 0$. Now subtracting a multiple of column 1 from the jth column, $j>1$, does not alter $\det(A)$ nor $D(A)$. This follows easily from (1) and (2) ((1) implying as before that a repeated column $\Rightarrow D = 0$, $\det = 0$). Namely (e.g. for $j=2$)

$$D\left(\text{col } 1 \quad \text{col } 2 + \alpha \text{ col } 1, \text{ other columns same}\right) =$$

$$D(\text{col } 1, \text{col } 2, \text{others as before}) + \alpha\, D(\text{col } 1, \text{col } 1, \text{other columns same})$$

$$= D(A) + 0$$ since ↗ has a repeated column

Thus $D(A) = D\begin{pmatrix} \lambda_1 & 0 \; 0 \cdots 0 \; 0 \\ \vdots & \boxed{\quad} \end{pmatrix}$ and

$\det(A) = \det\begin{pmatrix} \lambda_1 & 0 \; 0 \cdots 0 \; 0 \\ \vdots & \boxed{\quad} \end{pmatrix}$ $\lambda_1 \neq 0$

Now the $n-1 \times n-1$ block is nonsingular (since A is). Again without loss of generality we can suppose its upper left entry $\neq 0$ and do column operations to get

$D(A) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ \vdots & \lambda_2 & 0 \cdots & 0 \\ \vdots & \vdots & \boxed{n-2 \times n-2} & \end{pmatrix}$ $\det(A) = \det\begin{pmatrix} \lambda_1 & 0 & 0 \cdots & 0 \\ \vdots & \lambda_2 & 0 \cdots & 0 \\ \vdots & \vdots & \boxed{n-2 \times n-2} & \end{pmatrix}$

where $\lambda_1 \neq 0, \lambda_2 \neq 0$

Continuing we get "triangular" matrices

$D(A) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ & \lambda_2 & & 0 \\ & & \ddots & 0 \\ & & & \lambda_n \end{pmatrix}$ $\det(A) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ & \ddots & & 0 \\ & & \ddots & 0 \\ & & & \lambda_n \end{pmatrix}$

nonzero λ's down the diagonal, 0's "above" the diagonal. Then

$D(A) = D\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} + D\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$

by column linearity. But

$$D\begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \lambda_2 & \cdots & 0 \\ & & & \lambda_n \end{pmatrix} = 0$$ because the columns are linearly dependent, no matter what lies below the upper left 0 (the vectors, n columns, are effectively in $\mathbb{R}^{n-1}$, hence linearly dependent).

Continuing, we get

$$D\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ | & \lambda_2 & 0 \cdots & 0 \\ | & | & \ddots & \\ & & & \lambda_n \end{pmatrix} = D\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = \lambda_1 \cdots \lambda_n \cdot D\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

Similarly

$$\det\begin{pmatrix} \lambda_1 & & 0 \\ | & \ddots & \\ | & | & \lambda_n \end{pmatrix} = \lambda_1 \cdots \lambda_n .$$

Thus $$D(A) = \lambda_1 \cdots \lambda_n D(I_n)$$
$$= D(I_n) \det(A).$$

Thus
$$D = \lambda_0 \det$$
where $\lambda_0 = D(I_n)$. $\square$

Theorem: If P and A are $n \times n$ F-entry matrices, then
$$\det(P \times A) = \det(P)\det(A)$$

Proof: Let $D(A)$ be the function $\det(P \times A)$. By the definition of matrix multiplication and the properties (1) and (2) of det (antisymmetry & column linearity), D has properties (1) and (2). So $\exists$ $\lambda_0 \in F$ such that

$$D(A) = \det(P \times A) = \lambda_0 \det(A).$$

for all A (λ_0 independent of A).
Put $A = I_n$ to get $\lambda_0 = \det(P)$. Thus

$$\det(P \times A) = \det(P).\det(A) \quad \square .^{1}$$

Change of basis

Theorem: Suppose $T : V \to V$ is a linear transformation of a finite-dimensional vector space V. Suppose $v_1^{old} \dots v_n^{old}$ and $v_1^{new}, \dots, v_n^{new}$ are two bases of V and let
A_{old} = the matrix of T relative to $v_1^{old} \dots v_n^{old}$
A_{new} = the matrix of T relative to $v_1^{new} \dots v_n^{new}$

(same basis at both ends in each case). Then

$$A_{new} = P \, A_{old} \, P^{-1}$$

where P is the $n \times n$ matrix the columns

[1] This formula can also be checked directly from the permutation definition of determinants using $\mathrm{sgn}(\pi \circ \sigma) = \mathrm{sgn}(\pi)\,\mathrm{sgn}(\sigma)$. Details of this are left as an exercise.

of which are the components in the new basis of old basis vectors $v_1^{old} \ldots v_n^{old}$, ie, the jth column of P is $\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ where $v_j^{old} = \sum_{\ell=1}^{n} \mu_\ell v_\ell^{new}$.

Proof: By definition $P \begin{pmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{pmatrix} \leftarrow$ 1 in jth slot is a column vector representing v_j^{old} in the v^{new} basis. So $A_{new} P$ has j^{th} column representing $T v_j^{old}$ in the v^{new} basis. But

$A_{old} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow$ 1 in jth slot gives the v^{old} components of Tv_j so the columns of PA_{old} give the new-basis components of Tv_j, j = number of column. Hence

$$A_{new} P = P A_{old}$$

Hence $A_{new} = P A_{old} P^{-1}$.
(P is invertible since it is the new-basis representation, column by column of the v_j^{old} and hence has independent columns). $\square$

Recall:

Definition: An orthogonal matrix (over $\mathbb{R}$) is a matrix A with $AA^t = I_n$.

Lemma: If $v_1^{old} \ldots v_n^{old}$ and $v_1^{new} \ldots v_n^{new}$

are orthonormal bases for a real finite-dimensional vector space V with inner product $\langle \, , \, \rangle$, then the matrix P of the Theorem above is orthogonal.

Proof: $P^t P$ has entries equal to the inner products of the $v_1^{old} \ldots v_n^{old}$ vectors, since the columns of P are the new-basis components of $v_1^{old} \ldots v_n^{old}$, and the new basis is orthonormal. Since the old basis is orthonormal, these inner products $= 1$ if lth row, l column entry of $P^t P$ considered $(l = 1, \ldots, n)$ but all other (offdiagonal) entries are 0. $\square$

Theorem:

If a matrix A can be "diagonalized" by an orthonormal change of basis, then A is symmetric.

(Hence we say that a matrix is

diagonalized by a change of basis if the associated P has (with $A_{old} = A$)

$P A_{old} P^{-1}$ diagonal. This agrees with our previous notion that the new basis has $T v_j^{new} = \lambda_j v_j$).

Proof: If $A_{new} = P A_{old} P^{-1}$ with

P from an orthonormal basis change, then $P^{-1} = P^t$ and $A_{new} = P A_{old} P^t$. Then, if A_{new} is diagonal

$$A_{new} = A_{new}^t = P^{tt} A_{old}^t P^t = P A_{old}^t P^t$$

So $P A_{old} P^t = P A_{old}^t P^t$. Multiplying

on the right by P and on the left by $P^t (= P^{-1})$ gives

$A_{old} = A_{old}^t$ so A_{old} was symmetric. $\square$

This means that when we proved that symmetric matrices could be diagonalized by an orthogonal change of basis, we were doing the best possible thing — only symmetric matrices could be diagonalized in such a way.

Exercise: Carry out these ideas for Hermitian inner products and unitary matrices ($A^{-1} = \overline{A^t}$) and diagonalizing Hermitian matrices (matrix = its own transposed conjugate).

Application of normal operator ideas to orthogonal matrices.

Suppose A is a real orthogonal $n \times n$ matrix. Then we can think of A as a complex $n \times n$ matrix and it is then unitary since $A^{-1} = A^t = \overline{A^t}$ (since A is real, $\overline{A^t} = A^t$!)

Thus we can find a basis $v_1 \dots v_n$ (over $\mathbb{C}$) for $\mathbb{C}^n$ which is orthonormal for the standard $\mathbb{C}^n$ inner product $\langle \, , \rangle$ and satisfies

$$T_A v_j = \lambda_j v_j \quad \text{for some } \lambda_j \in \mathbb{C},$$

$j = 1, \dots n$, where T_A = the linear transformation $\mathbb{C}^n \to$ determined by A relative to the standard basis.

Note that $T_A v_j = \lambda_j v_j$ implies that $A \overline{v}_j = \overline{\lambda}_j \overline{v}_j$ since A is real. Thus

we can "pair up" the eigenvalues and eigenvect[ors] namely we can arrange that the eigenvectors with complex, nonreal

$\langle v_1, \bar{v}_1 \rangle = 0$! since $\lambda_1 \neq \bar{\lambda}_1$

eigenvalues occur as $v_1, \bar{v}_1$ $\lambda_1, \bar{\lambda}_1$ eigenvalues, $v_2, \bar{v}_2$, $\lambda_2, \bar{\lambda}_2$ eigenvalues (note: λ_1, λ_2 need not be distinct: think about the multiple eigenvalues, eigenspaces with $\dim_{\mathbb{C}} > 1$ case!) while the real eigenvalues occur as v_l, $l \geq 2k+1$, (k = number of complex pairs) have eigenvalues $= \pm 1$. All this is happening with v's orthonormal relative to the Hermitian inner product.

Now consider a pair $v, \bar{v}$, $\lambda, \bar{\lambda}$ eigenvalues respectively. Note that

$$\frac{1}{\sqrt{2}}(v + \bar{v}) \quad \text{and} \quad \frac{1}{\sqrt{2}\,i}(v - \bar{v})$$

have $\langle\ ,\ \rangle = 1$ and are perpendicular — and they are real vectors (so their Hermitian $\mathbb{C}^n$-inner product = their $\mathbb{R}^n$ standard inner product). Moreover, if $\lambda = a + bi$

$$T_A\left(\frac{1}{\sqrt{2}}(v + \bar{v})\right) = \frac{1}{\sqrt{2}}\left((a+bi)v + (a-bi)\bar{v}\right)$$

$$= a\left(\frac{1}{\sqrt{2}}(v+\bar{v})\right) - b\,\frac{1}{\sqrt{2}\,i}(v - \bar{v}).$$ And similarly

$$T_A\left(\frac{1}{\sqrt{2}}(v - \bar{v})\right) = b\left(\frac{1}{\sqrt{2}}(v+\bar{v})\right) + a\,\frac{1}{\sqrt{2}\,i}(v - \bar{v})$$

So relative to the real orthonormal pair $w_1 = \frac{1}{\sqrt{2}}(v+\bar{v})$ and $w_2 = \frac{1}{\sqrt{2}i}(v - \bar{v})$, T_A has matrix $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$.

Now, since $A = A^t$, T_A has an Hermitian matrix relative to any bases that is orthonormal relative to the (standard) Hermitian inner product. In particular it follows that when T_A is diagonalized relative to the $v_1, \bar{v}_1, v_2, \bar{v}_2 \ldots v_{2k}, v_{2k+1} \ldots v_n$

basis, that $\lambda_1 \bar{\lambda}_1 = 1$ etc.
because for this diagonal matrix, the transposed conjugate is just the same diagonal matrix except with the diagonal elements conjugated.
In particular, the $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ blocks have $a^2 + b^2 = 1$: they are "rotation blocks" of the form $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$

for some $\theta \in \mathbb{R}$. Thus we get that A has the block form

$$\begin{pmatrix} \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{pmatrix} & & & & & \text{0's elsewhere} \\ & \ddots & & & & \\ & & \begin{pmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{pmatrix} & & & \\ & & & \pm 1 & & \\ & & & & \ddots & \\ & & & & & \pm 1 \end{pmatrix}$$

Note that $\det A = (-1)^{\text{number of } -1\text{'s that occur as eigenvalues}}$.

Corollary: If $\det A = 1$, then there is a continuous curve of orthogonal matrices $A(t)$ with $A(1) = A$, $A(0) = I_n$.

Proof: The rotation blocks not of the form $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ nor $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ can be deformed to $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ by changing angle θ_j to $t\theta_j$. The -1 eigenvalues ~~even~~, even in number, can be grouped to form $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ blocks and be deformed to $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ by

$$\begin{pmatrix} \cos \pi t & \sin \pi t \\ -\sin \pi t & \cos \pi t \end{pmatrix} \quad t \in [0,1]. \quad \square$$

Exponentiation of Matrices

Let A be an $n \times n$ $\mathbb{R}$-entry or $\mathbb{C}$-entry matrix. Set

$$\exp(A) = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots$$

(note: $A^n = A \times \dots \times A$ n times, matrix multiplication)

Lemma: The series converges in operator norm [operator norm was discussed earlier, in inverting $I+T$]

Proof: $\|A^n\| \leq \|A\|^n$ from which the conclusion follows by the usual arguments about the convergence of $1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots$ □

exp is a differentiable map from $\mathbb{R}^{n^2}$ to $\mathbb{R}^{n^2}$. Its differential at $0 \in \mathbb{R}^{n^2}$ (goes to I_n in $\mathbb{R}^{n^2}$) is Identity map since

$$\exp(0 + tA) = I + tA + \underbrace{O(t^2)}_{o(t)} .$$

Note $\exp(n \times n \text{ matrices}) \subset Gl(n, \mathbb{R})$ (or $Gl(n, \mathbb{C})$)

because $\exp(A)\exp(-A) = \exp(0) = I$ since $A, -A$ commute so series cancels out as in $e^x e^{-x} = e^0$. (exercise).

As it happens, exp is not onto $Gl(n, \mathbb{R})$. For one thing $\exp(A)$ has to have $\det \exp(A) > 0$ because $\exp(tA)$ $t \in [0,1]$ is never 0 (since $\exp(tA)$ is invertible) and, since $\exp(0A)$ has det $= +1$, $\det \exp(tA)$ is positive for all $t \in [0,1]$ — a sign change would force a zero!

But actually, $\exp(\mathbb{R}^{n^2})$ is not all of $Gl^+(n, \mathbb{R})$ (= invertible matrices with det > 0) either. The exponential map for $Gl(n, \mathbb{C})$ is, however, surjective (more on these points later).

For the moment, we content ourselves with observing that

(1) $\exp(A) \in SO(n)$ if $A = -A^t$ (A is "skew symmetric")

Proof: $I = \exp(A)\exp(-A) = \exp(A)\exp(A^t)$ if A is skew symmetric. But $\exp(A^t) = [\exp(A)]^t$ so $\exp(A)$ is orthogonal. $\square$

(2) If $\exp(tA)$ is orthogonal for all small t positive, then A is skew sym.

Proof: $\exp(tA)[\exp(tA)]^t = \exp(tA)\exp(tA^t)$
$= I + t(A + A^t) + o(t)$

If this $= I$ for all (small) t, then $A + A^t = 0$ and $A = -A^t$

(3) All orthogonal matrices near I are $\exp(A)$, some skew sym A near 0. More precisely $\exists$ a nbhd U of 0 such that $\exp|\{A \in U : A = A^t\}$ maps 1-1 onto (a neighborhood V of I) $\cap\ SO(n, \mathbb{R})$

Proof: See Inverse Function Theorem argument.

that follows:

Lemma: exp is a "local diffeomorphism" of $\mathbb{R}^{n^2}$ to $\mathbb{R}^{n^2}$ in a neighborhood of 0 (to a neighborhood of I_n)

Proof: $\text{differential}\big|_0 = \text{identity}$ (as above)

Apply inverse function theorem. $\square$

Note: Matrix logs (inverse of exp) can be computed near I by the series used for $\log(1-x)$, namely
$\frac{1}{1-x} = 1 + x + x^2 \cdots$ so integrating gives
$-\ln(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} \cdots$ for $|x|<1$
This works for matrices, too: If $\|A\|<1$ (operator norm) then

$$\exp\left(-A - \frac{A^2}{2} - \frac{A^3}{3} \cdots\right) = I - A$$

The reason is that the identity $\exp\left(-x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \cdots\right) = 1 - x$ must be true — since it is valid for $|x|<1$ — on the level of substitution into power series, i.e.

$$1 + \left(-x - \frac{x^2}{2} - \frac{x^3}{3} \cdots\right) + \frac{1}{2}\left(-x - \frac{x^2}{2} \cdots\right)^2$$

$$+ \frac{1}{3!}\left(-x - \frac{x^2}{2} - \frac{x^3}{3} \cdots\right)^3 + \cdots = 1 - x$$

in the series manipulation sense. So putting $x = A$, it still holds as long as everything is absolutely (operator norm) convergent.

Finally we note that

$$e^{t\begin{pmatrix}0&1\\-1&0\end{pmatrix}} = I + t\begin{pmatrix}0&1\\-1&0\end{pmatrix} + \frac{t^2}{2}\begin{pmatrix}0&1\\-1&0\end{pmatrix}^2 \cdots$$

$$= \begin{pmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{pmatrix}$$ by direct calculation

(using $\begin{pmatrix}0&1\\-1&0\end{pmatrix}^2 = \begin{pmatrix}-1&0\\0&-1\end{pmatrix}$ $\begin{pmatrix}0&1\\-1&0\end{pmatrix}^3 = \begin{pmatrix}0&-1\\1&0\end{pmatrix}$

$\begin{pmatrix}0&1\\-1&0\end{pmatrix}^4 = \begin{pmatrix}1&0\\0&1\end{pmatrix}$)

From this, one gets a "matrix log" for any 2×2 block $\begin{pmatrix}a&b\\-b&a\end{pmatrix}$ $\ni$ $a^2+b^2=1$

From this, it follows easily that A orthogonal, $\det A = +1$ $= \exp(B)$ for some ~~[illegible]~~ B, namely there are in the canonical form for A: 2×2 blocks (perhaps), which can be handled as indicated, some -1's — but these are even in number and hence can be treated as 2×2 blocks when grouped in pairs; and some diagonal 1's, which are just $\exp(0)$. (The details are left as an exercise).