University of Leeds, School of Mathematics
MATH 2080 Further Linear Algebra, results and formulae (1)

Please let me know of any errors, omissions or obscurities.

Vector spaces. We are given a field F of scalars ($\mathbb{R}$, $\mathbb{C}$, or $\mathbb{F}_2$, which is the set $\{0, 1\}$ where we have the usual rules for arithmetical operations except that $1 + 1 = 0$) and a set V of vectors. Vectors can be added to each other and multiplied by scalars. Axioms: $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$, $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$, $\mathbf{u}+\mathbf{0} = \mathbf{0}+\mathbf{u} = \mathbf{u}$, $\mathbf{u}+(-\mathbf{u}) = \mathbf{0}$, $a(\mathbf{u}+\mathbf{v}) = a\mathbf{u}+a\mathbf{v}$, $(a+b)\mathbf{u} = a\mathbf{u}+b\mathbf{u}$, $(ab)\mathbf{u} = a(b\mathbf{u})$ and $1\mathbf{u} = \mathbf{u}$. For example, F^n.
Subspaces. A subset $W \subseteq V$ is a subspace if it is also a vector space. This means that $0 \in W$ and W is closed under addition and scalar multiplication. The intersection of subspaces is a subspace.
The smallest subspace containing a set S is the *linear span* $\mathrm{span}(S)$ or $\mathrm{lin}(S)$, consisting of all *linear combinations* $a_1\mathbf{s}_1 + \ldots + a_n\mathbf{s}_n$ of elements of S.
Sum of two subspaces. $U + W$ is the set of all $\mathbf{u} + \mathbf{w}$ where $\mathbf{u} \in U$ and $\mathbf{w} \in W$, the smallest subspace containing U and W. If $U \cap W = \{0\}$ then we write $U + W$ as $U \oplus W$, the *direct* sum and everything in it can be written as $\mathbf{u} + \mathbf{w}$ in a unique way.
Spanning and Independence. $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *span* V if every $\mathbf{v} \in V$ is a linear combination $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. They are *linearly independent* if the only linear combination $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ which is $\mathbf{0}$ is with $a_1 = \cdots = a_n = 0$.
Bases. A basis of V is a set which spans and is linearly independent, so that everything in V is a linear combination in a unique way. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an independent set and $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a spanning set, then $n \geq k$, and we can replace k elements of the second set by the elements from the first set to get a set which still spans.
Dimension. All bases of a space V have the same number of elements, the *dimension* of V, n, say. If a set has more than n elements it cannot be independent; if it has fewer than n elements, it cannot span. If it has exactly n elements and is independent, then it is automatically a basis.
Any independent set (e.g. a basis for a subspace) can be extended to a basis for the whole space by adding further independent vectors. So, if V is finite-dimensional and U is a subspace, then there is a *complement* W (generally not unique) with $U \oplus W = V$.
Any spanning set can be reduced to a basis by dropping vectors depending on earlier ones.
Linear transformations (or maps, or mappings). $T : U \to V$ is *linear* if it satisfies $T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$ and $T(a\mathbf{u}) = aT(\mathbf{u})$. Example: $T : F^n \to F^m$, left multiplication of a column vector by an $m \times n$ matrix A.
Rank and nullity. The *image* or *range* of T is $\mathrm{Im}\, T = \{T(\mathbf{u}) : \mathbf{u} \in U\}$, and is a subspace of V of dimension $r(T)$, the *rank*. The *kernel* or *null space* of T is $\mathrm{Ker}\, T = \{\mathbf{u} \in U : T(\mathbf{u}) = \mathbf{0}\}$, and is a subspace of U of dimension $n(T)$, the *nullity*. *Theorem:* $r(T) + n(T) = \dim U$. *Proof:* take a basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for $\mathrm{Ker}\, T$, extend to a basis for U by adding in $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$, and check that $\{T(\mathbf{u}_{k+1}), \ldots, T(\mathbf{u}_n)\}$ is a basis for $\mathrm{Im}\, T$. (*)

Coordinates. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is a basis of V and $\mathbf{v} = a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m$, then $a_1, \ldots, a_m$ are the *coordinates* of $\mathbf{v}$ with respect to this basis.
Fixing bases $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and $\mathbf{v}_1, \ldots, \mathbf{v}_m$ for U and V means that $T : U \to V$ corresponds to a unique $m \times n$ matrix $A = (a_{ij})$, where $T(\mathbf{u}_i) = \sum_{j=1}^{m} a_{ji}\mathbf{v}_j$, giving the ith column of A. If $\mathbf{u} \in U$ has coordinates $x_1, \ldots, x_n$ then $T(\mathbf{u}) \in V$ has coordinates $y_1, \ldots, y_m$, where $y_i = \sum_{j=1}^{n} a_{ij}x_j$, which is the same as $\mathbf{y} = A\mathbf{x}$, where $\mathbf{x}$ and $\mathbf{y}$ are written as column vectors.
Canonical form under equivalence. We can find bases such that the linear map has the block matrix form $\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}$, where r is the rank and I_r is an identity matrix of size r. Indeed we take $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n, \mathbf{w}_1, \ldots, \mathbf{w}_k\}$ as a basis for U (as in (*) above). For V we take $\{T(\mathbf{u}_{k+1}), \ldots, T(\mathbf{u}_n)\}$ and extend it to a basis for V.
Composition of mappings. Given $T : U \to V$, $S : V \to W$ linear then $ST : U \to W$ is also linear; if the same bases are used both times, then the matrix of ST is the product of the matrices for S and T.
Isomorphisms. These are maps $T : U \to V$ of vector spaces for which there is an inverse $T^{-1} : V \to U$ with $T^{-1}T = I_U$ and $TT^{-1} = I_V$, identity mappings. The spaces are then *isomorphic*. This only happens if $\dim U = \dim V$. If T has the matrix A with respect to some basis, then the matrix of T^{-1} (with the same bases) will be A^{-1}.
Change of basis. If A represents T with respect to some given bases, then by changing both bases we get $B = Q^{-1}AP$, where P is the matrix of the identity mapping on U, i.e., expresses the new basis in terms of the old one. Similarly for Q with V. (**)
Equivalence. We call two $m \times n$ matrices A and B *equivalent*, if $B = Q^{-1}AP$ for some nonsingular P and Q, and write $A \equiv B$. It's an *equivalence relation*, i.e., (i) $A \equiv A$, (ii) $A \equiv B \Rightarrow B \equiv A$, and (iii) $A \equiv B$, $B \equiv C \Rightarrow A \equiv C$. As a result the set of matrices is divided up into disjoint subsets (equivalence classes), and $A \equiv B$ if and only if they are in the same class. Each class contains exactly one matrix in canonical form under equivalence.
Mappings from a space to itself. Let $T : V \to V$ be represented by a matrix A using one basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. If instead we use the basis $\{\mathbf{v}'_1, \ldots, \mathbf{v}'_n\}$, then T is represented by $B = P^{-1}AP$ for some P, (since $Q = P$ in the theorem (**)) and also $\mathbf{v}'_j = \sum_{i=1}^{n} p_{ij}\mathbf{v}_i$.
Similar matrices. Two $n \times n$ matrices A and B are *similar* or *conjugate* if they represent the same linear map w.r.t. possibly different bases, i.e., $B = P^{-1}AP$. We write this $A \sim B$. Similarity is also an equivalence relation.
Eigenvectors and eigenvalues. For $T : V \to V$ linear, if we have $T(\mathbf{v}) = \lambda\mathbf{v}$, where $\mathbf{v} \neq \mathbf{0}$, then $\mathbf{v} \in V$ is an eigenvector and $\lambda \in F$ is an eigenvalue. Similarly for matrices, with $A\mathbf{x} = \lambda\mathbf{x}$. Eigenvalues λ are roots of the *characteristic equation* $\chi(t) = 0$ where $\chi(t) = \det(A - tI)$, a polynomial of degree n. Similar matrices have the same characteristic equation, and so the same eigenvalues. If A represents T using some basis, then A and T have the same eigenvalues.

J. R. Partington, revised 8.12.10

University of Leeds, School of Mathematics
MATH 2080 Further Linear Algebra, results and formulae (2)

Diagonalisability. The linear mapping $T : V \to V$ is *diagonalisable* if we can find a basis of eigenvectors, so that T is represented by a diagonal matrix using this basis. For a matrix A this means that there exists P s.t. $P^{-1}AP = B$, diagonal. Problems arise if we work with real scalars and some eigenvalues are not real. Even over $\mathbb{C}$ there are non-diagonalisable matrices, such as $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. (Why?)

Distinct eigenvalues. Eigenvectors corresponding to distinct eigenvalues are linearly independent, so if there are no repeated eigenvalues (and they are all in F) then the transformation/matrix is diagonalisable.

Polynomials. The *Cayley–Hamilton theorem* says that $\chi(A) = 0$. The *minimum* or *minimal polynomial* is the unique monic polynomial $\mu(t)$ of minimal degree s.t. $\mu(A) = 0$. All other such polynomials are multiples of it, so $\mu(t)$ divides $\chi(t)$ exactly. Every eigenvalue is a root of μ. Similar matrices have the same minimum polynomial, and A is diagonalisable if and only if μ has no repeated roots.

Jordan canonical (or normal) form. First, a Jordan block matrix is a square matrix like (λ), $\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, $\begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$, $\begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}$, etc., with λs on the diagonal and 1s just above it. If T is represented by a Jordan block matrix using a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then $(T-\lambda I)\mathbf{v}_k = \mathbf{v}_{k-1}$ for $k = 1, \ldots, n$, where we write $\mathbf{v}_0 = \mathbf{0}$. So $\mathbf{v}_1$ is an eigenvector, and we have a chain of *generalized eigenvectors*, satisfying $(T - \lambda I)^k \mathbf{v}_k = \mathbf{0}$ for each $k \geq 1$. Similar definitions hold for $n \times n$ matrices.

Now a square matrix A is in JCF if it consists of Jordan block matrices strung out along the diagonal, otherwise zero. The diagonal entries are the eigenvalues.

Theorem. If A is a complex $n \times n$ matrix, it is similar to a matrix in JCF, unique up to re-ordering of the blocks. If it has real entries the same is true (over the real field) provided that the eigenvalues are all real.

From JCF we can deduce Cayley–Hamilton, since if A is a block matrix with blocks B_i and p is a polynomial, then $p(A)$ has blocks $p(B_i)$. But for a Jordan block B_i of size $k \times k$ with eigenvalue λ we have $(B - \lambda I)^k = O$. So we get the result for matrices in JCF and hence for all square matrices.

Finally, if no eigenvalue has multiplicity greater than 3, then we can deduce the JCF knowing the characteristic and minimum polynomials: for we have the following possibilities for blocks: 1, 2, 1+1, 3, 2+1, 1+1+1) and χ has degree in λ equal to the total sizes of the blocks, whereas μ has degree equal to the size of the biggest block.

The number of blocks corresponding to λ is the dimension of the eigenspace $\mathrm{Ker}(T-\lambda I)$.

J. R. Partington, revised 8.12.10