SOME NOTES
MTH102A

1. A short proof of the uniqueness of row reduced echelon form

Theorem 1.1. *The row reduced echelon form of a matrix is unique.*

Proof. Let A be a matrix and suppose it has two row reduced echelon forms say B and C. That means applying a sequence of row operations to A we got B and applying another sequence of row operations we got C. We need to show that $B = C$.

Note that A, B, C are row equivalent to each other since row operation gives a row equivalent matrix. That means every row in A is a linear combination of rows of B and vice versa. Similarly every row in A is a linear combination of rows of C and vice versa.

On the contrary let us assume that B and C are not equal. Then select the leftmost column where they differ and also select all pivot columns (leading 1 columns) to the left of this column giving rise to two matrices say R and S. Since B and C were row rquivalent the matrices R and S are row equivalent since deletion of columns does not affect row equivalence.

Note that after interchanging some rows (if requird) the matrices R and S look like:

$$R = \left[\begin{array}{c|c} I_{r\times r} & \mathbf{r} \\ \hline \mathbf{0} & 0 \end{array}\right] S = \left[\begin{array}{c|c} I_{r\times r} & \mathbf{s} \\ \hline \mathbf{0} & 0 \end{array}\right]$$

It follows that R and S are row equivalent since deletion of columns (variables simultaneously) does not affect row equivalence, and that they are reduced but not equal. Now we treat these matrices as augmented matrices of two linear systems. The system for R has a unique solution $\mathbf{r}$ or is inconsistent, respectively. Similarly, the system for S has a unique solution $\mathbf{s}$ or is inconsistent, respectively. Since the systems are equivalent, $r = s$ or both systems are inconsistent. Either way we have $R = S$, a contradiction.

□

2. Even and odd permutations

Theorem 2.1. *The identity permutation is even.*

Proof. Let $id = t_1 t_2 \cdots t_{m-1} t_m$ where t_i's are transpositions. We need to show that m is even. Note that $m \neq 1$ as a single transposition is not the identity.

If $m = 2$ we are done.

We proceed by (strong) induction. Suppose that the theorem is true for any integer less than m, $m \geq 2$. We will show that it holds for m. Let $t_m = (a, b)$

The idea is that we will try to rewrite the permutation in such a way that we shift a as far left as possible until we eventually remove a from the permutation. The last pair of transpositions $t_{m-1}t_m$ must be one of these four cases:

$(ab)(ab), (bc)(ab), (ac)(ab), (cd)(ab)$.

If $t_{m-1}t_m = (ab)(ab) = id$, we are left with $m-2$ transpositions and by induction $m-2$ is even and so m is even.

If $t_{m-1}t_m = (bc)(ab)$, then we can replace it by $(ac)(bc)$ since $(bc)(ab) = (ac)(bc)$.

If $t_{m-1}t_m = (ac)(ab)$, then we can replace it by $(ab)(bc)$ since $(ac)(ab) = (ab)(bc)$.

If $t_{m-1}t_m = (cd)(ab)$, then we can replace it by $(ab)(cd)$ since $(cd)(ab) = (ab)(cd)$.

So we have rewritten $t_{m-1}t_m$ in such a way that a no longer occurs in the last transposition.

Successively, we rewrite the pairs $t_{m-1}t_m$, then $t_{m-2}t_{m-1}$, $t_{m-2}t_{m-1}$, and so on. Eventually, we will reach the first case above, $(ab)(ab)$, where we can cancel out two transpositions. If we don?t, then the left most transposition t_1 will have the only occurrence of a. This would contradict the assumption that the permutation is the identity, because if only one transposition contains a, then the permutation does not fix a.

Once we cancel the two transpositions, then there are only $m-2$ transpositions in the permutation, and we can apply our induction hypothesis. □

Corollary 2.2. *Suppose a permutation σ can be written as a product of m number of transpositions and also as a product of n number of transpositions for some m and n. Then m and n are both even or both odd.*

Proof. If $\sigma = t_1t_2\cdots t_m = s_1s_2\cdots s_n$ where t_i and s_j are transpositions. Then $id = \sigma.\sigma^{-1} = t_1t_2\cdots t_m s_n s_{n-1}\cdots s_2 s_1$. Since the identity permutation is even $m+n$ is even. So m and n are both even or both odd.

□

3. A short proof of Cramer's rule

Theorem 3.1. *The system $AX = d$, where A is an $n \times n$ invertible matrix, has a unique solution, whose individual values for the unknowns are given by: $x_i = \frac{det(A_i)}{det(A)}$, $i = 1, 2, \cdots n$, where A_i is the matrix formed by replacing the i-th column of A by the column vector d.*

Proof. Let X_i be the matrix obtained from the identity matrix by replacing the i-th column by the column X. Then $AX_i = A_i$. Since $det(X_i) = x_i$ we have $x_i = \frac{det(A_i)}{det(A)}$.

□

4. Existence of a Basis

Example: *The set of real numbers $\mathbb{R}$ is a vector space over $\mathbb{Q}$. What could possibly be a basis ? The elements $\sqrt{(}2), \sqrt{(}3), \sqrt{(}5), \sqrt{(}6), \cdots$ can be shown to be linearly independent, but they certainly don't span $\mathbb{R}$ as we also need elements like $\pi, \pi^2, \pi^3, \cdots$ which also form a* *linearly independent set. In fact, because $\mathbb{Q}$ is countable, it is easy to show that the subspace of $\mathbb{R}$ generated by any countable subset of $\mathbb{R}$ must be countable. Because $\mathbb{R}$ itself is uncountable, no countable set can be a basis for $\mathbb{R}$ over $\mathbb{Q}$. This means that any basis for $\mathbb{R}$ over $\mathbb{Q}$, if one exists, is going to be difficult to describe.*

The above example makes it clear that even if we could show that every vector space has a basis, it is unlikely that a basis will be easy to find or to describe in general. To prove that every vector space has a basis, we need **Zorn's Lemma***.*

Zorn's Lemma: *Let $\mathcal{C}$ be a collection of subsets of some fixed set, and assume that $\mathcal{C}$ has the property that whenever there is a chain $S_1 \subset S_2 \subset S_3 \cdots$ of sets in $\mathcal{C}$ the union of this chain also belongs to $\mathcal{C}$ then $\mathcal{C}$ contains a maximal element.*

Theorem 4.1. *Every vector space has a basis.*

Proof. First, consider any linearly independent subset of a vector space V, for example, a set consisting of a single non-zero vector will do. Call this set S_1. If S_1 spans V it is a basis, and the proof is complete. If not, we can choose a vector of V not in S and the union $S_2 = S_1 \cup \{v\}$ is a larger linearly independent set. Either this set is a basis, or we can again enlarge it by choosing some vector of V not in the span. We can repeat this process over and over, and hope that it eventually ends. But it is easy to see that such a naive approach will not work in general unless V is finite dimensional. Indeed, starting from S_1 being a single element set, every S_i produced this way will be finite. On the other hand, using this idea, we either get a basis for V eventually or we get an increasing collection of linearly independent sets $S_1 \subset S_2 \subset S_3 \cdots$. The union S of all the S_i is a linearly independent set, since any finite linear combination of the elements of the union must involve elements from one of the sets S_i. If this set S spans V, it is a basis and we are done.

However, even if S does not span V, it is at least linearly independent, so we could again choose a vector v not in the span of S. By adding v to S, we again get a larger linearly independent set, and we can repeat the process. Does this process eventually terminate, producing for us a basis of V ? This is not at all clear.

Now let $\mathfrak{C}$ be the collection of all linearly independent subsets of a vector space V. Since the union of any increasing chain $S_1 \subset S_2 \subset S_3 \cdots$ of linearly independent sets is also a linearly independent set, Zorn's Lemma implies that there is a maximal linearly independent set say M. This maximal linearly independent set is a basis for V. Indeed, if it doesn't span V, we could choose a vector in V but not in the span of M, and by adding it to M, we could enlarge our supposedly maximal linearly independent set. This contradiction completes the proof that every vector space V has a basis.

□

Remark: *There is a major drawback to this proof that every vector space has a basis: unless the dimension is finite, or at least countable, it doesn't give us any idea how to actually find a basis. In fact, this is a serious problem with the concept of a basis for infinite dimensional spaces in general. Although Zorn's Lemma tells us a basis exists, in practice, this fact may be useless if we do not have a procedure for finding one.*

Remark: *Zorn's Lemma is logically equivalent to the "axiom of choice". The axiom of choice says that given any collection $\mathcal{C}$ of sets, we can choose an element x from each set S of $\mathcal{C}$. This may seem "obvious"- or does it ? There is of course no problem if there are finitely many sets in the collection, but what if there are infinitely many, may be even uncountably many ? The axiom of choice and Zorn's Lemma bothered many mathematicians (and still bothers some!) for various reasons. For example, using the axiom of choice, one can prove that a ball the size of the sun can be cut into finitely many pieces and then reassembled into a ball the size of a pinhead. So if we accept the axiom of choice (and equivalently, Zorn's Lemma), we must accept such statements.*

Remark: *There is no way to prove the axiom of choice: one either accepts it as an axiom or one doesn't. The axiom of choice (and so the equivalent formulation Zorn's Lemma) is logically independent from the other axioms of set theory, a fact proven by Paul Cohen in 1963. In other words, we derive no contradictions if we assume it is true, and we derive no contradictions if we assume it is false. The axiom of choice is no longer as controversial as it once was. It is accepted by most mathematicians these days, but the degree to which it is used without comment depends on the branch of mathematics.*

5. Minimal Polynomial and diagonalizability

Theorem 5.1. *A $n\times n$ matrix A is diagonalizable over $\mathbb{F}$ if and only if its minimal polynomial is a product of distinct linear factors over $\mathbb{F}$.*

Proof. If A is diagonalizable, then there exists an invertible matrix P such that $P^{-1}AP$ is diagonal. Since A and $P^{-1}AP$ have the same minimal polynomial, we can compute the minimal polynomial of A using the diagonal matrix, and then it is clear that we just need one linear factor for each of the distinct entries along the diagonal. That means if $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the distinct eigen values, then $m(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_k)$ is the minimal polynomial of A. (Note that $m(A) = (A-\lambda_1 I)(A-\lambda_2 I)\cdots(A-\lambda_k I) = \mathbf{0}$ and it is the smallest degree monic polynomial having this property).

Conversely, assume that the minimal polynomial $m(x)$ for A is a product of distinct linear factors. We need to show that A is diagonalizable, that means we need to find an invertible matrix P such that $P^{-1}AP$ is diagonal. Note that in this case the columns of A have to be the eigen vectors of A and the invertibility of P ensures that the columns of P form a basis for $\mathbb{F}^n$.

Let W be the subspace of $\mathbb{F}^n$ spanned by all eigenvectors of A. So in order to show that A is diagonalizable we need to show that $W = \mathbb{F}^n$. Assume on the contrary that $W \neq V$.

Claim: There exists $v \in \mathbb{F}^n \setminus W$ such that $(A-\lambda I).v \in W$ for some eigen value λ.

Proof of the claim: Let $z \in V \setminus W$. Since $m(A) = \mathbf{0}$ we have $0 = m(A).z \in W$. Let $g(x)$ be a monic polynomial of minimal degree with $g(A).z \in W$. We show that $g(x)$ divides $m(x)$. By division algorithm write $m(x) = q(x)g(x) + r(x)$ where either $r(x) = 0$ or $deg(r(x)) < deg(g(x))$. Then $0 = m(A).z = q(A)g(A).z + r(A).z$.

Note that since W is spanned by all eigenvectors of A, we have $A.w \in W$ for any $w \in W$ and so $q(A).g(A).z \in W$ as $w = g(A).z \in W$. So from the above equation we have $r(A).z \in$

W. Since $deg(r(x)) < deg(g(x))$ and $g(x)$ is a monic polynomial of minimal degree with $g(A).z \in W$ we conclude that $r(A).z \in W$ implies $r(x) = 0$. So $g(x)$ divides $m(x)$. So by the assumption $g(x)$ is a product of distinct linear factors.

Write $g(x) = (x - \lambda)h(x)$ where $deg(h(x)) < deg(g(x))$ and λ is an eigen value. By the definition of $g(x)$ we have $h(A).z \notin W$. Then $(A - \lambda I).h(A).z = g(A).z \in W$. So we showed that **the vector** $v = h(A).z \notin W$ **but** $(A - \lambda I).v \in W$.

Since $g(x)$ divides $m(x)$, write $m(x) = (x - \lambda)q(x)$ for some $q(x)$.

Now $0 = m(A).z = (A - \lambda I)(q(A).v)$. Then $q(A).v$ is an eigen vector for the eigen value λ. So $q(A).v \in W$.

Again write $q(x) = t(x)(x - \lambda) + q(\lambda)$ for some polynomial $t(x)$. Then $q(\lambda).v = q(A).v - t(A)(A - \lambda I).v$. Now $q(A).v \in W$ because it is an eigenvector. By the claim we have $(A - \lambda I).v \in W$ and so $t(A)(A - \lambda I).v \in W$ (since W is spanned by the eigen vectors). So $q(\lambda).v \in W$. Since $v \notin W$we have $q(\lambda) = 0$. This says that λ is a multiple root of $m(x)$, a contradiction. □

Remark: Note that the above claim holds for any proper subspace $W \subset \mathbb{F}^n$ with the property that $A.w \in W$ for all $w \in W$.

Triangulizable Matrix: A matrix A is said to be triangularizable (or triangulable) if there exists an invertible matrix P such that $P^{-1}AP$ is upper-triangular.

Theorem 5.2. *A $n \times n$ matrix A is triangularizable over $\mathbb{F}$ if and only if its characteristic (or minimal) polynomial is a product of linear factors over $\mathbb{F}$.*

Proof. Note that the characteristic polynomial is a product of linear factors over $\mathbb{F}$ if and only if the minimal polynomial is a product of linear factors over $\mathbb{F}$. The matrix A defines a linear map $T : \mathbb{F}^n \to \mathbb{F}^n$ defined by $X \mapsto AX$. Then A is triangularizable iff T is triangularizable, i.e., there exists a basis B of $\mathbb{F}^n$ for which $[T]_B$ is upper -triangular.

Assume that the minimal polynomial is a product of linear factors over $\mathbb{F}$. Let u_1 be an eigen vector of A with some eigen value say λ_1. Let $W_1 = span\{u_1\}$. So W_1 is a proper subspace of $\mathbb{F}^n$. So by the remark (and proof of the claim) there exists $u_2 \in V \setminus W_1$ such that $(T - \lambda_2 I)(u_2) = (A - \lambda_2 I).u_2 \in W_1$ for some eigen value λ_2. So $T(u_2) - \lambda_2 u_2 \in W_1$. Hence $T(u_2) = \lambda_2 u_2 + w_1$ where $w_1 \in W_1$. Next let W_2 be the subspace spanned $\{u_1, u_2\}$. Since T maps the basis for W_2 back into W_2. So by the remark again there exists $u_3 \in V \setminus W_2$ such that $(T - \lambda_3 I)(u_3) = (A - \lambda_3 I).u_3 \in W_2$. So $T(u_3) = \lambda_3 u_3 + w_2$ where $w_2 \in W_2$ and we can repeat the argument. In this process we get a basis $B = \{u_1, u_2, \cdots, u_n\}$ such that $T(u_i) = \lambda_i u_i + w_{i-1}$ for an eigen value λ_i and w_{i-1} belongs to the subspace spanned by $\{u_1, u_2, \cdots, u_{i-1}\}$. t follows that the matrix for T relative to the basis B which has been chosen is an upper triangular matrix. □

Corollary 5.3. *Any matrix is triangulizable over $\mathbb{C}$.*

==

6. Singular Value Decomposition

We need the following two lemmas for the proof of singular value decomposition.

Lemma 6.1. *For a nonzero $m \times n$ matrix the non-zero eigen values of A^*A and AA^* are same.*

Proof. Let $\lambda \neq 0$ be an eigenvalue of A^*A, i.e. $A^*Ax = \lambda x$ for some $x \neq 0$. Multiplying A from the left we get $AA^*(Ax) = \lambda(Ax)$. If $Ax = 0$ then $\lambda x = 0$ which is not possible as both λ and x are non-zero. So $Ax \neq 0$ and hence Ax is an eigen vector of AA^* with eigen value λ. □

Definition: We say A is a positive definite (resp. positive semi-definite) matrix if $x^*Ax > 0$ (resp. $x^*Ax \geq 0$) for all nonzero vectors x.

Lemma 6.2. *All the eigen values of a positive definite matrix are strictly positive.*

Proof. Let λ be an eigen value of A. If $\lambda = 0$, then there is some eigenvector x so that $Ax = 0$. But then $x^*Ax = 0$, and so A is not positive definite.

If $\lambda < 0$, then there is some eigenvector x so that $Ax = \lambda x$. But then $x^*Ax = \lambda\|x\|^2$, which is negative since $\|x\| > 0$ and $\lambda < 0$. Thus A is not positive definite.

So if A is positive definite, it only has positive eigenvalues. □

Remark: All the eigen values of a positive semi-definite matrix are strictly positive.

Singular Value Decomposition: The singular value decomposition (SVD) is a matrix factorization. If A is an $m \times n$ matrix, then we may write A as a product of three factors:

$$A = U\Sigma V^* \ , \tag{6.1}$$

where U is an unitary $m \times m$ matrix, V is an unitary $n \times n$ matrix, V^* is the conjugate transpose of V, and Σ is an $m \times n$ matrix that has all zeros except for its diagonal entries, which are nonnegative real numbers. If σ_{ij} is the i, j entry of Σ, then $\sigma_{ij} = 0$ unless $i = j$ and $\sigma_{ii} = \sigma_i \geq 0$. The σ_i's are called the "singular values" and the columns of U and V are respectively called the left and right singular vectors. A common convention is to list the singular values in descending order.

$$\sigma_1 \geq \sigma_2 \geq \cdots \ .$$

$$A = U\Sigma V^* = \begin{bmatrix}\underbrace{\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_r}_{\operatorname{Col} A} & \underbrace{\mathbf{u}_{r+1} \quad \dots \quad \mathbf{u}_m}_{\operatorname{Nul} A^*}\end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \dots & & & & & & \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \dots & & & & & & \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^* \\ \mathbf{v}_2^* \\ \dots \\ \mathbf{v}_r^* \\ \mathbf{v}_{r+1}^* \\ \dots \\ \mathbf{v}_n^* \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \\ a \end{matrix}}\right\} \operatorname{Row} A \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \operatorname{Nul} A \end{matrix}$$

In the above picture $ColA, RowA, NulA, NulA^*$ denote the four fundamental subspaces associated to the matrix A, namely column space of A, row space of A, null space of A and null space of A^* respectively.

Proof of Existence: Let A be an $m \times n$ complex matrix. Since A^*A is Hermitian, it is unitarily diagonalizable. So there exists an $n\times n$ unitary matrix V such that $V^*A^*AV = \mathbf{D} = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ where D is a diagonal matrix of size $r\times r$. ($r = rank(A^*A) = rank(A) \leq min\{m, n\}$ and if $r = n$ then there are no zeros in the diagonal of $\mathbf{D}$). Again since A^*A is positive semi-definite the diagonal entries of D (the eigen values) are non-negative real numbers. Note that by definition the columns of V are the eigen vectors of A^*A. Write $V = [V_1\ V_2]$ where V_1 is a $n \times r$ matrix whose columns are the eigen vectors corresponding to the non-zero eigen values of A^*A whereas V_2 is a $n \times (n - r)$ matrix whose columns are the eigen vectors corresponding to the eigen value 0.

Then $\begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = V^*A^*AV = \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} A^*A[V_1\ V_2] = \begin{bmatrix} V_1^*A^*AV_1 & V_1^*A^*AV_2 \\ V_2^*A^*AV_1 & V_2^*A^*AV_2 \end{bmatrix}$.

Equating we get $V_1^*A^*AV_1 = D$ and $V_2^*A^*AV_2 = \mathbf{0}$. This implies $(AV_1)^*AV_1 = D$ and $(AV_2)^*AV_2 = \mathbf{0}$.

The 2nd equation shows that $AV_2 = \mathbf{0}$. (For a matrix B, if $B^*B = \mathbf{0}$ then $B = \mathbf{0}$).

Since $V = [V_1, V_2]$ is unitary we have $V^*V = VV^* = I$. From here we get $V_1^*V_1 = I, V_2^*V_2 = I$ and $V_1V_1^* + V_2V_2^* = I$. (The identity matrices on the right hand side of the equations are of different sizes, the 1st one is of size $r \times r$, the 2nd one is of size $n - r \times n - r$ and the 3rd one is of size $n \times n$).

We define $U_1 = AV_1D^{-\frac{1}{2}}$. It is a matrix of size $m \times r$. The entries of $D^{-\frac{1}{2}}$ are same as the entries of D except the positive diagonal entries of D are replaced by their $-\frac{1}{2}$-th power.

Then $U_1D^{\frac{1}{2}}V_1^* = AV_1D^{-\frac{1}{2}}D^{\frac{1}{2}}V_1^* = AV_1V_1^* = A(I - V_2V_2^*) = A - AV_2V_2^* = A$ (We used the fact that $V_1V_1^* + V_2V_2^* = I$ and $AV_2 = 0$.)

We also have $U_1^*U_1 = D^{-\frac{1}{2}}V_1^*A^*AV_1D^{-\frac{1}{2}} = D^{-\frac{1}{2}}DD^{-\frac{1}{2}} = I$. So the columns of U_1 form an orthonormal set in $\mathbb{C}^m$ and can be extended to form an orthonormal basis for $\mathbb{C}^m$.

Let $U = [U_1\ U_2]$ where the columns of U_2 are the extended part of the above orthonormal basis of $\mathbb{C}^m$. Then U is unitary.

Define $\Sigma = \begin{bmatrix} D^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ (if necessary extra zero rows may be added or removed to make the number of zero rows equal the number of columns of U_2 and hence the overall size of Σ is $m \times n$).

Then $U\Sigma V^* = [U_1\ U_2] \begin{bmatrix} D^{\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [V_1\ V_2]^* = U_1D^{\frac{1}{2}}V_1^* = A$.

Remark: We are aiming for $A = U\Sigma V^*$. We have $A^*A = V\Sigma^*U^*U\Sigma V^* = V\Sigma^*\Sigma V^*$. This implies $A^*AV = V\Sigma^*\Sigma$. So the columns of V must be the eigen vectors for A^*A with respect to the eigen values placed in the diagonal of $\Sigma^*\Sigma$. Similarly the column vectors of

U are the eigen vectors of the matrix AA^* with respect to the same (nonzero) eigen values. (A^*A and AA^* have same nonzero eigen values).

Remark: The diagonal entries σ_i of Σ are known as the singular values of A. A common convention is to list the singular values in descending order. In this case, the diagonal matrix, Σ, is uniquely determined by A but not the matrices U and V.

Algorithm: Given a $m \times n$ matrix A.

(1) Compute the eigen values of A^*A.

(2) Compute the eigen vectors corresponding to the eigen values of A^*A.

(3) Construct the matrix V by placing the eigen vectors of A^*A as columns.

(4) Construct the diagonal matrix Σ by putting the square roots of the (positive) eigen values of A^*A in decreasing order on the main diagonal and adding some zero rows at the end if necessary to make it an $m \times n$ matrix.

(5) Solve the linear system $AV^* = U\Sigma$ for U. (The matrix U can be constructed by finding the eigen vectors of AA^* with respect to the eigen values of A^*A found in step (1)).