Math 304-6 Feingold   Linear Algebra   Spring 2020   Topics Summary

These notes will summarize some topics and results needed for this course.
This version was created on 3-14-2020.

Notations: $\mathbf{R}^m_n$ is the set of all $m \times n$ real matrices, $\mathbf{R}^m = \mathbf{R}^m_1$ and $\mathbf{0} \in \mathbf{R}^m_n$ is the zero matrix of the appropriate size, depending on context. The transpose of matrix A is denoted by A^T. For $A = [a_{ij}] \in \mathbf{R}^m_n$, let $Row_i(A) = [a_{i1} \cdots a_{in}] \in \mathbf{R}_n$ be the i^{th} row of A and let $Col_j(A) = [a_{1j} \cdots a_{mj}]^T \in \mathbf{R}^m$ be the j^{th} column of A. Then we write $Row(A)$ (the row space of A) for the span of the rows of A, and we write $Col(A)$ (the column space of A) for the span of the columns of A. For vector spaces, we write $U \leq V$ when U is a subspace of V.

Topic 1: Linear map $L_A : \mathbf{R}^n \to \mathbf{R}^m$ determined by $A \in \mathbf{R}^m_n$.

For $A \in \mathbf{R}^m_n$ define the linear function (transformation or map) $L_A : \mathbf{R}^n \to \mathbf{R}^m$ by

$$L_A(X) = AX = \sum_{j=1}^{n} x_j Col_j(A)$$

so $Range(L_A) = Col(A)$ is the span of the columns of A. For $1 \leq j \leq n$ let $\mathbf{e}_j \in \mathbf{R}^n$ be the matrix with 1 in row j and 0 in all other rows. Then $L_A(\mathbf{e}_j) = Col_j(A)$ and $Range(L_A) = Col(A) = \langle L_A(\mathbf{e}_1), \cdots, L_A(\mathbf{e}_n)\rangle$.

We also defined $Nul(A) = Ker(L_A) = \{X \in \mathbf{R}^n \mid L_A(X) = 0\}$, and we know how to find this subspace in terms of free variables by row reducing $[A|\mathbf{0}]$ to $[C|\mathbf{0}]$ where C is in Reduced Row Echelon Form (RREF) and has $r = Rank(A)$ leading 1's in certain "pivot columns". Then the interpretation of $[C|\mathbf{0}]$ gives the values of all variables, $x_1, \cdots, x_n$ in terms of $n - r$ free variables if $n - r > 0$, or gives only the trivial solution if $n = r$. If $n-r > 0$ and that expression for X in $Nul(A)$ is written with the free variables separated, it is of the form

$$X = f_1K_1 + f_2K_2 + \cdots + f_{n-r}K_{n-r}$$

where free variable $f_i = x_{k_i}$ and each vector $K_i \in Nul(A)$ is specifically determined with constant entries. Those free variables correspond to non-pivot columns numbered $k_1, \cdots, k_{n-r}$. The list of vectors $\{K_1, \cdots, K_{n-r}\}$ is independent, so it is a basis for $Nul(A) = Ker(L_A)$ and $dim(Nul(A)) = dim(Ker(L_A)) = n - r$.

But $K_i \in Nul(A)$ means that $AK_i = \mathbf{0}$ is a certain linear combination of columns of A with the column numbered k_i having coefficient 1, and the other columns with non-zero coefficients having subscripts less than k_i. It means $AK_i = \mathbf{0}$ is a dependence relation among the columns of A showing that (non-pivot column) $Col_{k_i}(A)$ is redundant because it is a linear combination of previous (pivot) columns of A. Doing this for all $n - r$ free variables, we find that all non-pivot columns of A are redundant, and the set consisting of only the pivot columns is an independent spanning set for $Range(L_A) = Col(A)$, that is, a basis for $Range(L_A) = Col(A)$. Then $dim(Range(L_A)) = dim(Col(A)) = r = Rank(A)$. This argument establishes the following result.

Theorem (Rank-Nullity Theorem):

For $A \in \mathbf{R}^m_n$ and associated linear map $L_A : \mathbf{R}^n \to \mathbf{R}^m$, we have

$$n = dim(\mathbf{R}^n) = dim(Ker(L_A)) + dim(Range(L_A))$$

equivalently, $n = dim(Nul(A)) + dim(Col(A))$. With the notation $Nullity(A) = dim(Nul(A))$, since $Rank(A) = dim(Col(A))$, this theorem says

$$n = Nullity(A) + Rank(A).$$

Example: On Exam 1, in problem (1), recall the matrix $A = [a_{ij}] \in \mathbf{R}^4_5$ with $a_{ij} = i + j - 1$, and let $S = \{v_1, v_2, v_3, v_4, v_5\}$ where $v_j = Col_j(A) \in \mathbf{R}^4$. Ker($L_A$) was found by solving the linear system $AX = \mathbf{0}$ by row reducing $[A|\mathbf{0}]$ to $[C|\mathbf{0}]$, C in RREF:

$$\left[\begin{array}{ccccc|c} 1 & 2 & 3 & 4 & 5 & 0 \\ 2 & 3 & 4 & 5 & 6 & 0 \\ 3 & 4 & 5 & 6 & 7 & 0 \\ 4 & 5 & 6 & 7 & 8 & 0 \end{array}\right] \text{ to } \left[\begin{array}{ccccc|c} 1 & 0 & -1 & -2 & -3 & 0 \\ 0 & 1 & 2 & 3 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \text{ so } \begin{array}{r} x_1 = r + 2s + 3t \\ x_2 = -2r - 3s - 4t \\ x_3 = r \in \mathbf{R} \\ x_4 = s \in \mathbf{R} \\ x_5 = t \in \mathbf{R} \end{array}$$

$$\text{and } X \in Ker(L_A) \text{ when } X = \begin{bmatrix} r + 2s + 3t \\ -2r - 3s - 4t \\ r \\ s \\ t \end{bmatrix} = r \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 3 \\ -4 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

In the previous notation, the free variables in the solution are:

$$r = f_1 = x_3 \qquad s = f_2 = x_4 \qquad t = f_3 = x_5$$

and $Ker(L_A) = Nul(A)$ has basis

$$\left\{ K_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, K_2 = \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, K_3 = \begin{bmatrix} 3 \\ -4 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Since $AX = \mathbf{0}$ means $\sum_{j=1}^5 x_j Col_j(A) = \mathbf{0}$, solutions give all dependence relations on S:

$$(r + 2s + 3t)v_1 + (-2r - 3s - 4t)v_2 + rv_3 + sv_4 + tv_5 = \theta \quad \text{for any } r, s, t \in \mathbf{R}.$$

Using each K_i separately gives three separate dependence relations among the columns of A (and notice that these relations also hold among the columns of C):

$1v_1 - 2v_2 + 1v_3 = \theta$ so $v_3 = -v_1 + 2v_2$.
$2v_1 - 3v_2 + 1v_4 = \theta$ so $v_4 = -2v_1 + 3v_2$.
$3v_1 - 4v_2 + 1v_5 = \theta$ so $v_5 = -3v_1 + 4v_2$.

This shows that the columns $v_3, v_4, v_5 \in \langle v_1, v_2\rangle$ so the last three vectors are redundant in S and the subset $T = \{v_1, v_2\}$ has the same span as S but is independent, so T is a basis for $Range(L_A) = Col(A)$. But, of course, $Col(C)$ is completely different from $Col(A)$, and the span of the first two columns of C would be all vectors in $\mathbf{R}^4$ whose last two coordinates are zero. None of the columns of A satisfies that condition. $Nullity(A) = 3$ and $Rank(A) = 2$ and $5 = dim(\mathbf{R}^5) = 3 + 2 = Nullity(A) + Rank(A)$.

There is a generalization of the Rank-Nullity Theorem which works for any linear map $L : V \to W$ when $dim(V)$ is finite.

Theorem: If $L : V \to W$ is a linear map and $dim(V)$ is finite then

$$dim(V) = dim(Ker(L)) + dim(Range(L)).$$

Here are several useful applications of this theorem.

Theorem: If $L : V \to W$ is a linear map and $dim(V) = n$ and $dim(W) = m$ are both finite, then:

(a) If $n > m$ then $n - m \leq dim(Ker(L)) \leq dim(V)$.
(b) If $n < m$ then $0 \leq dim(Range(L)) \leq dim(V)$.
(c) If $n = m$ then L is **injective** iff L is **surjective**.

The proofs of these results use a few other little results about dimensions of subspaces which you should know.

Theorem: If $U \leq V$ then $dim(U) \leq dim(V)$.

Theorem: If $U \leq V$ and $dim(V)$ is finite and $dim(U) = dim(V)$ then $U = V$.

Topic 2: Linear independence/dependence, spanning, basis and dimension.

Let $S = \{v_1, \cdots, v_k\} \subset V$ be a subset (list) in a vector space V and let $W = \langle S\rangle$ be the span of S. We say S is **independent** when the vector equation

$$\sum_{i=1}^{k} x_i v_i = \theta$$

has only the trivial solution $x_i = 0$ for all $1 \leq i \leq k$. Otherwise, we say S is **dependent**, and such an equation with not all coefficients zero is a **dependence relation** on S.

In such a dependence relation there must be a term with largest subscript having a non-zero coefficient, say it is the term with $i = m$. Then, since $x_m \neq 0$, we can solve for vector v_m as

$$v_m = -\sum_{i=1}^{m-1} \frac{x_i}{x_m} v_i$$

which means v_m is in the span of the previous vectors $\{v_1, \cdots, v_{m-1}\}$ so $\langle S\rangle = \langle S - \{v_m\}\rangle$. This is what we mean when we say v_m is **redundant** in S. If the reduced set $S_1 = S - \{v_m\}$ is independent, then S_1 is a basis for the subspace $\langle S\rangle = \langle S_1\rangle$. But if S_1 is dependent, there must be another dependence relation on S_1 giving a redundant vector in it which can be removed to produce an even smaller set S_2. This process continues and in ℓ steps reduces S to an independent subset $T = S_\ell$ such that $\langle S\rangle = \langle T\rangle$ and T is a basis for $W = \langle S\rangle = \langle T\rangle$. This is called **reducing a spanning set to a basis**.

Theorem: If S is any subset of a vector space V and $W = \langle S\rangle$ is the span of S, then S contains a subset T such that T is a basis for W, that is, T is independent and spans W.

On the other hand, suppose we are given an independent set S in a vector space V. How can we get a basis for V containing S, that is, how can we extend S to a basis for V? The answer is in the following result.

Theorem: Let $S = \{v_1, \cdots, v_k\}$ be independent in V and let $v_{k+1} \in V$. Then $S \cup \{v_{k+1}\} = \{v_1, \cdots, v_k, v_{k+1}\}$ is independent if and only if $v_{k+1} \notin \langle S\rangle$.

This gives the following procedure to **extend an independent set to a basis**. If S already spans V, then we are done since S itself is a basis for V. Otherwise, $\langle S\rangle$ is a proper subspace of V and there must be a vector, $v_{k+1} \notin \langle S\rangle$ so adjoining it to S gives the larger independent set $S_1 = \{v_1, \cdots, v_k, v_{k+1}\}$ whose span is more than $\langle S\rangle$. If $\langle S_1\rangle = V$ then we are done, but otherwise $\langle S_1\rangle$ is still a proper subspace of V and we can find another vector, $v_{k+2} \notin \langle S_1\rangle$ so adjoining it to S_1 gives the even larger independent set $S_2 = \{v_1, \cdots, v_k, v_{k+1}, v_{k+2}\}$ whose span is more than $\langle S_1\rangle$. If $\langle S_2\rangle = V$ then we are done, but otherwise it is still proper and we can repeat this extension process as long as necessary to eventually get an independent spanning set for V, that is, a basis for V.

Theorem: Every vector space has a basis.

Theorem: If $S = \{v_1, \cdots, v_m\}$ and $T = \{w_1, \cdots, w_n\}$ are both bases of the same vector space V, then $m = n$.

The last theorem means that the common number of vectors in any basis of V is an intrinsic property of V, which we call the **dimension** of V, denoted by $dim(V)$. For example, the standard basis of $\mathbf{R}^m_n$ consists of the mn matrices $E^{rs} = [e^{rs}_{ij}] = [\delta_{ri}\delta_{sj}]$, each having all zero entries except for a 1 in one entry, so $dim(\mathbf{R}^m_n) = mn$.

Example: The standard basis of $\mathbf{R}^2_3$ is the set (list) of matrices

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

$= \{E^{11}, E^{12}, E^{13}, E^{21}, E^{22}, E^{23}\}$. The standard basis of the vector space of polynomials of at most degree k,

$$Poly_k = \{\sum_{i=0}^{k} c_i t^i \mid c_i \in \mathbf{R}\}$$

is $\{1 = t^0, t = t^1, t^2, \cdots, t^k\}$, which contains $k+1$ vectors (polynomials), so $dim(Poly_k) = k + 1$.

There are special kinds of matrices, depending on the pattern of entries, especially for square matrices. We are very familiar with RREF as a special kind of $m \times n$ matrix, but here are some special kinds of square matrices you should know about.

Definitions: A square matrix $A = [a_{ij}] \in \mathbf{R}^n_n$ is called: (1) **Lower triangular** when $a_{ij} = 0$ if $i < j$, (2) **Upper triangular** when $a_{ij} = 0$ if $i > j$, (3) **Diagonal** when $a_{ij} = 0$ if $i \neq j$, (4) **Scalar** when it is diagonal and all diagonal entries are the same number, so $a_{ij} = c$ if $i = j$ and $a_{ij} = 0$ if $i \neq j$, that is, $A = cI_n$ where $I_n = [\delta_{ij}]$ is the identity matrix in $\mathbf{R}^n_n$, (5) **Symmetric** when $a_{ij} = a_{ji}$ for all $1 \leq i, j \leq n$, that is, when $A = A^T$, (6) **Anti-symmetric** when $a_{ij} = -a_{ji}$ for all $1 \leq i, j \leq n$, that is, when $A = -A^T$, so, in particular, $a_{ii} = -a_{ii}$, giving that $a_{ii} = 0$.

It is easy to check that the set of matrices of each of these types is a subspace of $\mathbf{R}^n_n$, and you should be able to find a basis for each of these subspaces. For example, a basis for the subspace of symmetric matrices in $\mathbf{R}^2_2$ is

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

and basis for the subspace of anti-symmetric matrices in $\mathbf{R}^2_2$ is

$$\left\{ \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}.$$

There are several important general theorems about independent and dependent sets of vectors.

Theorem: Let $S \subseteq T \subseteq V$. Then we have:
(a) If T is independent then S is independent.
(b) If S is dependent then T is dependent.

These two statements are logically equivalent since they are contrapositives of each other, that is, one is of the form P implies Q, and the other is of the form (not Q) implies (not P). So it is enough to just check one of them, say (b). If S is dependent then there is a dependence relation among the vectors in S, but that is also a dependence relation among the vectors in T since S is part of T, so T is dependent.

Theorem: A set $S = \{v_1\}$ consisting of just one vector is independent iff $v_1 \neq \theta$.

Theorem: A set $S = \{v_1, v_2\}$ consisting of just two vectors is independent iff neither vector is a scalar multiple of the other one.

Theorem: A set $S = \{v_1, v_2, ..., v_k\}$ consisting of at least two vectors is independent iff no vector in S is a linear combination of the previous vectors on the list.

Topic 3: Composition of linear maps in relation to matrix multiplication.

For matrices $A \in \mathbf{R}^m_n$, $B \in \mathbf{R}^n_p$, and $C \in \mathbf{R}^m_p$ and the associated linear maps (functions) $L_A : \mathbf{R}^n \to \mathbf{R}^m$, $L_B : \mathbf{R}^p \to \mathbf{R}^n$, and $L_C : \mathbf{R}^p \to \mathbf{R}^m$ we have defined the matrix product $C = AB$ uniquely by the condition that $L_A \circ L_B = L_C$, where $\circ$ means composition of functions.

This gave us many properties of matrix multiplication, for example, that it is associative, as well as the formula for the entries of the product $C = [c_{ik}] = AB$ in terms of the entries of $A = [a_{ij}]$ and $B = [b_{jk}]$:

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

as well as the formula for AB in terms of columns of B:

$$C = AB = [ACol_1(B)|ACol_2(B)|\cdots|ACol_p(B)].$$

Topic 4: Getting elementary row operations through matrix multiplication.

There is a way to achieve elementary row operations through matrix multiplication on the left which is very useful.

Definition: A matrix $E \in \mathbf{R}^n_n$ is called an **elementary matrix** associated with an elementary row operation when E is obtained by doing that elementary row operation to the identity matrix I_n.

For example, the following are the 3×3 elementary matrices corresponding to the three elementary row operations (1) Switch Row 2 and Row 3, (2) Multiply row 2 by $0 \neq c \in \mathbf{R}$, (3) Replace Row 3 by c Row 1 + Row 3:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c & 0 & 1 \end{bmatrix}.$$

Theorem: For any $A \in \mathbf{R}^m_n$ if $E \in \mathbf{R}^m_m$ is the elementary matrix corresponding to an elementary row operation, then EA is the matrix obtained by doing that elementary row operation to A.

This means that a sequence of elementary row operations can be achieved by a sequence of left multiplications by elementary matrices. Suppose $E_1, E_2, \cdots, E_t$ are elementary matrices corresponding to elementary row operations $Op_1, Op_2, \cdots, Op_t$. Then $B = E_t \cdots E_2 E_1 A$ is the matrix obtained from A by doing those elementary row operations to A in that order. That is what happens when A and B are "row equivalent", and in particular, when A row reduces to B in RREF.

We know that a square $n \times n$ matrix A is invertible iff $Rank(A) = n$ iff A row reduces to I_n, that is, $[A|I_n] \to [I_n|A^{-1}]$. We can achieve the elementary row operations that do this reduction by using elementary matrix multiplications:

$$[A|I_n] \to [E_1 A|E_1 I_n] \to [E_2 E_1 A|E_2 E_1 I_n] \to \cdots \to [E_t \cdots E_2 E_1 A|E_t \cdots E_2 E_1 I_n] = [I_n|A^{-1}]$$

which says $E_t \cdots E_2 E_1 A = I_n$ and $E_t \cdots E_2 E_1 = E_t \cdots E_2 E_1 I_n = A^{-1}$.

Theorem: If $A, B \in \mathbf{R}^n_n$ are both invertible then so is their product, AB, and $(AB)^{-1} = B^{-1} A^{-1}$.

Theorem: If $A_1, A_2, \cdots, A_k \in \mathbf{R}_n^n$ are all invertible then so is their product, $A_1 A_2 \cdots A_k$ and $(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} \cdots A_2^{-1} A_1^{-1}$, the inverse of a product of invertible matrices is the product of the inverses in the opposite order.

Topic 5: The standard dot product in $\mathbf{R}^n$.

For $X = [x_j], Y = [y_j] \in \mathbf{R}^n$ define the standard dot product

$$X \cdot Y = \sum_{j=1}^{n} x_j y_j$$

so we get the following properties:
(1) $X \cdot Y = Y \cdot X$ (Symmetry).
(2) $(aX + bY) \cdot Z = a(X \cdot Z) + b(Y \cdot Z)$ (Linearity)
(3) $X \cdot X \geq 0$ and $X \cdot X = 0$ iff $X = \mathbf{0} \in \mathbf{R}^n$ (Positive Definite)

Define the length of $X \in \mathbf{R}^n$ to be $||X|| = \sqrt{X \cdot X} = \sqrt{\sum_{j=1}^{n} x_j^2}$.

For $X, Y \in \mathbf{R}^n$ let $\theta_{X,Y}$ be the angle between X and Y determined by

$$\cos(\theta_{X,Y}) = \frac{X \cdot Y}{(||X||)(||Y||)}.$$

Definition: For $X, Y \in \mathbf{R}^n$, say X and Y are orthogonal (perpendicular) when $X \cdot Y = 0$ and write $X \perp Y$. For non-zero vectors this means $\theta_{X,Y} = \pi/2$.

Theorem (Cauchy-Schwarz Inequality): For any $X, Y \in \mathbf{R}^n$ we have

$$|X \cdot Y| \leq (||X||)(||Y||).$$

Theorem (Triangle Inequality): For any $X, Y \in \mathbf{R}^n$ we have $||X + Y|| \leq ||X|| + ||Y||$.