REVIEW SHEET FOR MIDTERM 2: BASIC

MATH 196, SECTION 57 (VIPUL NAIK)

We will not be going over this sheet, but rather, we'll be going over the advanced review sheet in the session. Please review this sheet on your own time.

The summaries here are identical with the executive summaries you will find at the beginning of the respective lecture notes PDF files. The summaries are not intended to be exhaustive. Please review the original lecture notes as well, especially if any point in the summary is unclear.

1. Matrix multiplication and inversion

Note: The summary does not include some material from the lecture notes that is not important for present purposes, or that was intended only for the sake of illustration.

(1) *Recall*: A $n \times m$ matrix A encodes a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ given by $T(\vec{x}) = A\vec{x}$.
(2) We can add together two $n \times m$ matrices entry-wise. Matrix addition corresponds to addition of the associated linear transformations.
(3) We can multiply a scalar with a $n \times m$ matrix. Scalar multiplication of matrices corresponds to scalar multiplication of linear transformations.
(4) If $A = (a_{ij})$ is a $m \times n$ matrix and $B = (b_{jk})$ is a $n \times p$ matrix, then AB is defined and is a $m \times p$ matrix. The $(ik)^{th}$ entry of AB is the sum $\sum_{j=1}^{n} a_{ij}b_{jk}$. Equivalently, it is the dot product of the i^{th} row of A and the k^{th} column of B.
(5) Matrix multiplication corresponds to composition of the associated linear transformations. Explicitly, with notation as above, $T_{AB} = T_A \circ T_B$. Note that $T_B : \mathbb{R}^p \to \mathbb{R}^n$, $T_A : \mathbb{R}^n \to \mathbb{R}^m$, and $T_{AB} : \mathbb{R}^p \to \mathbb{R}^m$.
(6) Matrix multiplication makes sense *only if* the number of columns of the matrix on the left equals the number of rows of the matrix on the right. This comports with its interpretation in terms of composing linear transformations.
(7) Matrix multiplication is associative. This follows from the interpretation of matrix multiplication in terms of composing linear transformations and the fact that function composition is associative. It can also be verified directly in terms of the algebraic definition of matrix multiplication.
(8) Some special cases of matrix multiplication: multiplying a row with a column (the inner product or dot product), multiplying a column with a row (the outer product or Hadamard product), and multiplying two $n \times n$ diagonal matrices.
(9) The $n \times n$ identity matrix is an identity (both left and right) for matrix multiplication wherever matrix multiplication makes sense.
(10) Suppose n and r are positive integers. For a $n \times n$ matrix A, we can define A^r as the matrix obtained by multiplying A with itself repeatedly, with A appearing a total of r times.
(11) For a $n \times n$ matrix A, we define A^{-1} as the unique matrix such that $AA^{-1} = I_n$. It also follows that $A^{-1}A = I_n$.
(12) For a $n \times n$ invertible matrix A, we can define A^r for all integers r (positive, zero, or negative). A^0 is the identity matrix. $A^{-r} = (A^{-1})^r = (A^r)^{-1}$.
(13) Suppose A and B are matrices. The question of whether $AB = BA$ (i.e., of whether A and B commute) makes sense only if A and B are both square matrices of the same size, i.e., they are both $n \times n$ matrices for some n. However, $n \times n$ matrices need not always commute. An example of a situation where matrices commute is when both matrices are powers of a given matrix. Also, diagonal matrices commute with each other, and scalar matrices commute with all matrices.
(14) Consider a system of simultaneous linear equations with m variables and n equations. Let A be the coefficient matrix. Then, A is a $n \times m$ matrix. If $\vec{y}$ is the output (i.e., the augmenting column) we

can think of this as solving the vector equation $A\vec{x} = \vec{y}$ for $\vec{x}$. If $m = n$ and A is invertible, we can write this as $\vec{x} = A^{-1}\vec{y}$.

(15) There are a number of algebraic identities relating matrix multiplication, addtion, and inversion. These include distributivity (relating multiplication and addition) and the involutive nature or reversal law (namely, $(AB)^{-1} = B^{-1}A^{-1}$). See the "Algebraic rules governing matrix multiplication and inversion" section in the lecture notes for more information.

Computational techniques-related ...

(1) The arithmetic complexity of matrix addition for two $n \times m$ matrices is $\Theta(mn)$. More precisely, we need to do mn additions.
(2) Matrix addition can be completely parallelized, since all the entry computations are independent. With such parallelization, the arithmetic complexity becomes $\Theta(1)$.
(3) The arithmetic complexity for multiplying a generic $m \times n$ matrix and a generic $n \times p$ matrix (to output a $m \times p$ matrix) using naive matrix multiplication is $\Theta(mnp)$. Explicitly, the operation requires mnp multiplications and $m(n-1)p$ additions. More explicitly, computing each entry as a dot product requires n multiplications and $(n-1)$ additions, and there is a total of mp entries.
(4) Matrix multiplication can be massively but not completely parallelized. All the entries of the product matrix can be computed separately, already reducing the arithmetic complexity to $\Theta(n)$. However, we can parallelized further the computation of the dot product by parallelizing addition. This can bring the arithmetic complexity (in the sense of the depth of the computational tree) down to $\Theta(\log_2 n)$.
(5) We can compute powers of a matrix quickly by using repeated squaring. Using repeated squaring, computing A^r for a positive integer r requires $\Theta(\log_2 r)$ matrix multiplications. An explicit description of the minimum number of matrix multiplications needed relies on writing r in base 2 and counting the number of 1s that appear.
(6) To assess the invertibility and compute the inverse of a matrix, augment with the identity matrix, then row reduce the matrix to the identity matrix (note that if its rref is not the identity matrix, it is not invertible). Now, see what the augmented side has turned to. This takes time (in the arithmetic complexity sense) $\Theta(n^3)$ because that's the time taken by Gauss-Jordan elimination (about n^2 row operations and each row operation requires $O(n)$ arithmetic operations).
(7) We can think of pre-processing the row reduction for solving a system of simultaneous linear equations as being equivalent to computing the inverse matrix first.

Material that you can read in the lecture notes, but not covered in the summary.

(1) Real-world example(s) to illustrate matrix multiplication and its associativity (Sections 3.4 and 6.3).
(2) The idea of fast matrix multiplication (Section 4.2).
(3) One-sided invertibility (Section 8).
(4) Noncommuting matrix examples and finite state automata (Section 10).

2. Geometry of linear transformations

(1) There is a concept of *isomorphism* as something that preserves essential structure or feature, where the concept of isomorphism depends on what feature is being preserved.
(2) There is a concept of *automorphism* as an isomorphism from a structure to itself. We can think of automrohpisms of a structure as *symmetries* of that structure.
(3) Linear transformations have already been defined. An *affine linear transformation* is something that preserves lines and ratios of lengths within lines. Any affine linear transformation is of the form $\vec{x} \mapsto A\vec{x} + \vec{b}$. For the transformation to be linear, we need $\vec{b}$ to be the zero vector, i.e., the transformation must send the origin to the origin. If A is the identity matrix, then the affine linear transformation is termed a *translation*.
(4) A linear *isomorphism* is an invertible linear transformation. For a linear isomorphism to exist from $\mathbb{R}^m$ to $\mathbb{R}^n$, we must have $m = n$. An affine linear isomorphism is an invertible affine linear transformation.
(5) A linear automorphism is a linear isomorphism from $\mathbb{R}^n$ to itself. An affine linear automorphism is an affine linear isomorphism from $\mathbb{R}^n$ to itself.

(6) A self-isometry of $\mathbb{R}^n$ is an invertible function from $\mathbb{R}^n$ to itself that preserves Euclidean distance. Any self-isometry of $\mathbb{R}^n$ must be an affine linear automorphism of $\mathbb{R}^n$.
(7) A self-homothety of $\mathbb{R}^n$ is an invertible function from $\mathbb{R}^n$ to itself that scales all Euclidean distances by a factor of λ, where λ is the factor of homothety. We can think of self-isometries precisely as the self-homotheties by a factor of 1. Any self-homothety of $\mathbb{R}^n$ must be an affine linear automorphism of $\mathbb{R}^n$.
(8) Each of these forms a group: the affine linear automorphisms of $\mathbb{R}^n$, the linear automorphisms of $\mathbb{R}^n$, the self-isometries of $\mathbb{R}^n$, the self-homotheties of $\mathbb{R}^n$.
(9) For a linear transformation, we can consider something called the *determinant*. For a 2×2 linear transformation with matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the determinant is $ad - bc$.

We can also consider the *trace*, defined as $a + d$ (the sum of the diagonal entries).
(10) The trace generalizes to $n \times n$ matrices: it is the sum of the diagonal entries. The determinant also generalizes, but the formula becomes more complicated.
(11) The determinant for an affine linear automorphism can be defined as the determinant for its linear part (the matrix).
(12) The sign of the determinant being positive means the transformation is orientation-preserving. The sign of the determinant being negative means the transformation is orientation-reversing.
(13) The magnitude of the determinant gives the factor by which volumes are scaled. In the case $n = 2$, it is the factor by which areas are scaled.
(14) The determinant of a self-homothety with factor of homothety λ is $\pm\lambda^n$, with the sign depending on whether it is orientation-preserving or orientation-reversing.
(15) Any self-isometry is volume-preserving, so it has determinant ± 1, with the sign depending on whether it is orientation-preserving or orientation-reversing.
(16) For $n = 2$, the orientation-preserving self-isometries are precisely the translations and rotations. The ones fixing the origin are precisely rotations centered at the origin. These form groups.
(17) For $n = 2$, the orientation-reversing self-isometries are precisely the reflections and glide reflections. The ones fixing the origin are precisely reflections about lines passing through the origin.
(18) For $n = 3$, the orientation-preserving self-isometries fixing the origin are precisely the rotations about axes through the origin. The overall classification is more complicated.

3. Image and kernel of a linear transformation

(1) For a function $f : A \to B$, we call A the domain, B the co-domain, $f(A)$ the range, and $f^{-1}(b)$, for any $b \in B$, the fiber (or inverse image or pre-image) of b. For a subset S of B, $f^{-1}(B) = \bigcup_{b \in S} f^{-1}(b)$.
(2) The sizes of fibers can be used to characterize injectivity (each fiber has size at most one), surjectivity (each fiber is non-empty), and bijectivity (each fiber has size exactly one).
(3) Composition rules: composite of injective is injective, composite of surjective is surjective, composite of bijective is bijective.
(4) If $g \circ f$ is injective, then f must be injective.
(5) If $g \circ f$ is surjective, then g must be surjective.
(6) If $g \circ f$ is bijective, then f must be injective and g must be surjective.
(7) Finding the fibers for a function of one variable can be interpreted geometrically (intersect graph with a horizontal line) or algebraically (solve an equation).
(8) For continuous functions of one variable defined on all of $\mathbb{R}$, being injective is equivalent to being increasing throughout or decreasing throughout. More in the lecture notes, sections 2.2-2.5.
(9) A vector $\vec{v}$ is termed a *linear combination* of the vectors $\vec{v_1}, \vec{v_2}, \dots, \vec{v_r}$ if there exist real numbers $a_1, a_2, \dots, a_r \in \mathbb{R}$ such that $\vec{v} = a_1\vec{v_1} + a_2\vec{v_2} + \cdots + a_r\vec{v_r}$. We use the term *nontrivial* if the coefficients are not all zero.
(10) A subspace of $\mathbb{R}^n$ is a subset that contains the zero vector and is closed under addition and scalar multiplication.

(11) The span of a set of vectors is defined as the set of all vectors that can be written as linear combinations of vectors in that set. The span of any set of vectors is a subspace.
(12) A spanning set for a subspace is defined as a subset of the subspace whose span is the subspace.
(13) Adding more vectors either preserves or increases the span. If the new vectors are in the span of the previous vectors, it preserves the span, otherwise, it increases it.
(14) The kernel and image (i.e., range) of a linear transformation are respectively subspaces of the domain and co-domain. The kernel is defined as the inverse image of the zero vector.
(15) The column vectors of the matrix of a linear transformation form a spanning set for the image of that linear transformation.
(16) To find a spanning set for the kernel, we convert to rref, then find the solutions parametrically (with zero as the augmenting column) then determine the vectors whose linear combinations are being discussed. The parameters serve as the coefficients for the linear combination. There is a shortening of this method. (See the lecture notes, Section 4.4, for a simple example done the long way and the short way).
(17) The fibers for a linear transformation are translates of the kernel. Explicitly, the inverse image of a vector is either empty or is of the form (particular vector) + (arbitrary element of the kernel).
(18) The dimension of a subspace of $\mathbb{R}^n$ is defined as the minimum possible size of a spanning set for that subspace.
(19) For a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ with $n \times m$ matrix having rank r, the dimension of the kernel is $m - r$ and the dimension of the image is r. Full row rank $r = n$ means surjective (image is all of $\mathbb{R}^n$) and full column rank $r = m$ means injective (kernel is zero subspace).
(20) We can define the intersection and sum of subspaces of $\mathbb{R}^n$.
(21) The kernel of $T_1 + T_2$ contains the intersection of the kernels of T_1 and T_2. More is true (see the lecture notes).
(22) The image of $T_1 + T_2$ is contained in the sum of the images of T_1 and T_2. More is true (see the lecture notes).
(23) The dimension of the inverse image $T^{-1}(X)$ of any subspace X of $\mathbb{R}^n$ under a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ satisfies:

$$\dim(\mathrm{Ker}(T)) \leq \dim(T^{-1}(X)) \leq \dim(\mathrm{Ker}(T)) + \dim(X)$$

The upper bound holds if X lies inside the image of T.
(24) Please read through the lecture notes thoroughly, since the summary here is very brief and inadequate.