REVIEW SHEET FOR FINAL: ADVANCED

MATH 196, SECTION 57 (VIPUL NAIK)

To maximize efficiency, please bring a copy (print or readable electronic) of this review sheet to the Saturday review sesssion.

Please come to the session *only if* you know the meanings of these:

- Subspace
- Linear transformation
- Linear combination
- Linear relation
- Span
- Spanning set
- Linear dependence
- Linear independence
- Basis
- Dimension
- Kernel
- Image
- Rank
- Nullity
- Injectivity
- Surjectivity
- Bijectivity
- Transpose of a matrix

Please go through the basic review sheet, book, or lecture notes, if any of these terms trip you up.

1. Linear dependence, bases and subspaces

Error-spotting exercises ...

(1) *Half-truth*: Consider $\mathbb{R}$ as a vector space. Then, $\mathbb{Z}$, the set of integers, is a subspace of $\mathbb{R}$ because it is closed under addition and contains the zero vector.

(2) *Something doesn't add up*: Consider $\mathbb{R}^2$ as a vector space. Then, the set comprising the vectors $\vec{e}_1, \vec{e}_2, \vec{e}_1 + \vec{e}_2$ and their scalar multiples is a subspace because it contains the zero vector, is closed under addition (after all, $\vec{e}_1 + \vec{e}_2 = \vec{e}_1 + \vec{e}_2$) and is closed under scalar multiplication (by assumption).

(3) *Too non-slanted, too uncrooked*: Suppose U is a vector subspace of $\mathbb{R}^n$. Then, we can obtain a basis for U as follows: start with the standard basis for $\mathbb{R}^n$. Now, pick the vectors from this that are also inside U. These form a basis for U. For instance, if U is the span of $\vec{e}_2$ and $\vec{e}_3$ in $\mathbb{R}^4$, this procedure works.

(4) *Telepathic basis*: Suppose U_1 and U_2 are vector subspaces of $\mathbb{R}^n$. Suppose we are given a basis S_1 for U_1 and a basis S_2 for U_2. Then, $S_1 \cap S_2$ is a basis for the vector space $U_1 \cap U_2$ and $S_1 \cup S_2$ is a basis for the vector space $U_1 \cup U_2$.

(5) *Too trivial to be true*: If a collection of vectors in $\mathbb{R}^n$ satisfies the trivial linear relation, it is termed linearly independent. If, however, it does not satisfy the trivial linear relation, it is termed linearly dependent.

(6) *Throwing out the baby with the bathwater*: Suppose S is a set of vectors in $\mathbb{R}^n$ that spans a subspace V of $\mathbb{R}^n$, and there is a nontrivial linear relation between the vectors of S that uses four of the vectors in S nontrivially (i.e., it has nonzero coefficients for four of the vectors in S). This means

that we can throw out any of those four vectors and still span V. Thus, we can reduce the size of S by 4 and still get a spanning set for V.

2. Coordinates (in focus: similarity of linear transformations)

Error-spotting exercises ...

(1) *Same similar*: Suppose A_1 and B_1 are similar $n \times n$ matrices, and A_2 and B_2 are also similar $n \times n$ matrices. Then, $A_1 A_2$ and $B_1 B_2$ are similar $n \times n$ matrices. Here is the proof: since A_1 and B_1 are similar, there exists a matrix S such that $A_1 = SB_1S^{-1}$. Since A_2 and B_2 are similar, there exists a matrix S such that $A_2 = SB_2S^{-1}$. Then, $A_1A_2 = (SB_1S^{-1})(SB_2S^{-1}) = S(B_1B_2)S^{-1}$. Thus, A_1A_2 and B_1B_2 are similar.

(2) *Shallow roots*: Suppose A and B are $n \times n$ matrices and r is a positive integer. Is it the case that A^r being similar to B^r implies that A is similar to B? Well, this depends on whether r is odd or even. If r is even, then A and B need not be similar. We can get counterexamples even using 1×1 matrices: $[1]$ and $[-1]$ have the same square, but are different.

    On the other hand, if r is odd, then A^r and B^r being similar implies that A and B are similar. Here is the proof. If A^r and B^r are similar, this implies that there exists an invertible $n \times n$ matrix S such that $A^r = SB^rS^{-1} = (SBS^{-1})^r$. So, $A^r = (SBS^{-1})^r$. Since r is odd, we can cancel it from the exponent (note that we do not have the $\pm$ issue that we have with even r) and we get that $A = SBS^{-1}$, so that A and B are similar.

(3) *One-sided scaling*: Any two scalar matrices are similar because they represent the same linear transformation viewed at different scalings.

(4) *One-sided relabeling*: Interchanging the roles of the standard basis vectors $\vec{e}_1$ and $\vec{e}_2$ shows that the matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$$

    are similar to one another.

3. Abstract vector spaces (in focus: function spaces)

Error-spotting exercises ...

(1) *Too big to compare*: Consider differentiation. We can think of this as a linear transformation from $C^1(\mathbb{R})$ (the vector space of all continously differentiable functions on all of $\mathbb{R}$) to $C(\mathbb{R})$ (the vector space of all continuous functions on $\mathbb{R}$). Here is an explanation for why the map is surjective: we know that the kernel of this linear transformation is the vector space of constant functions, which is 1-dimensional. By the rank-nullity theorem, we know that (rank) + (nullity) = dimension of domain. Since the nullity is 1, and the domain is infinite-dimensional, the rank is infinite. This equals the dimension of the co-domain. Since the rank equals the dimension of the co-domain, that means that the image is the whole co-domain, so the differentiation linear transformation is surjective.

(2) *Answer not in the answer key*: We can view differentiation as a linear transformation from $\mathbb{R}(x)$, the vector space of all rational functions in one variable, to $\mathbb{R}(x)$. This linear transformation is not injective, because its kernel is the space of constant functions, which is one-dimensional. The linear transformation is surjective, because we know how to integrate any rational function.

(3) (Can't think of a witty name): Consider the vector space of all rational functions that can be written in the form $r(x)/p(x)$ where p is a fixed polynomial of degree n. This vector space is n-dimensional. A basis for this is given by the rational functions of the form $1/(x - \alpha)$ for all roots α of p, as well as rational functions of the form $1/q(x)$ for all irreducible quadratic factors of p.

(4) *Off by one errors*: Consider differentiation as a linear transformation from P_n to P_n, where P_n is the vector space of all polynomials of degree less than or equal to n. P_n is a n-dimensional space. The linear transformation we obtain is bijective from P_n to P_n. This is because the derivative of any such polynomial is such a polynomial, and every such polynomial is the derivative of a polynomial. Explicitly, if we write a matrix for the linear transformation of differentiation, the matrix is a square matrix and has full rank n.

4. Ordinary least squares regression

Error-spotting exercises ...

(1) *Explaining the past versus predicting the future, aka it's hard to make predictions, especially about the future*: Suppose we are trying to fit a linear function of one variable using some data points (input-output pairs). If we use only two data points, then we can get a unique line through it, with an error vector of zero. In other words, we get a *perfect fit* with zero error.

Suppose that instead we use three data points. Due to measurement error, it is likely that there will be no line that perfectly fits all the three data points. We can still get a fit that minimizes error. However, notice that the magnitude of the error vector is now bigger: the error vector was earlier a zero vector, but now it is (probably) a nonzero vector.

A similar argument can be used to show that *the more data points we have, the larger the magnitude of the error vector for our best fit.* In fact, this is true even when we make an adjustment for the number of coordinates (the error vector with four data points will not just have bigger length in expectation than the error vector with three data points, but the ratio of lengths will be expected to be more than $\sqrt{4}/\sqrt{3}$, i.e., *each coordinate* of the error vector is getting bigger in expectation.

So, this means that, for a given functional form, the greater the number of data points we decide to use, the worse the fit we will obtain. Therefore, to obtain a good fit, we should choose as few data points as possible, though still enough to uniquely determine the function. In the linear case, this ideal number is 2. Less is too little. More is too confusing, because of the possible inconsistencies that arise.

(2) *Off by one errors*: Consider trying to fit a function of one variable, with n data points (i.e., n input-output pairs) where we attempt to fit it using a polynomial of degree (at most) m. Then, the design matrix for the regression (i.e., the coefficient matrix of the linear system) is a $m \times n$ matrix, because there are m parameters and n input-output pairs.

(3) *The best route to success is to avoid listening to negative feedback*: When choosing the design matrix of a linear regression, i.e., choosing the inputs of the input-output pairs, we should attempt to make the design matrix a square matrix of full rank. This is because we want full column rank in order to uniquely determine the parameters, and we need full row rank in order to make sure that a solution *exists*.

(4) *Portrait versus landscape*: Suppose we are trying to find the parameter vector $\vec{\beta}$ given the design matrix X. In other words, we are trying to solve the equation below, with $\vec{\varepsilon}$ chosen to be the vector of minimum length for which the system is consistent:

$$X\vec{\beta} = \vec{y} - \vec{\varepsilon}$$

We know that the vector $\vec{\varepsilon}$ is orthogonal to the image of X. Therefore, it is orthogonal to all the rows of the matrix X. In other words, $X\vec{\varepsilon} = \vec{0}$.

Thus, if we multiply both equations on the left by X, we obtain:

$$X^2\vec{\beta} = X\vec{y}$$

We can solve this system to find the best fit parameter vector $\vec{\beta}$.

5. Extra topic covered in the quizzes: Linear dynamical systems

Error-spotting exercises ...

(1) *Get unreal!*: Suppose A is a $n \times n$ matrix and $\vec{x}$ is a nonzero vector in $\mathbb{R}^n$. Suppose there exists a positive integer r such that $A^r\vec{x}$ is the zero vector in $\mathbb{R}^n$. Since $\vec{x}$ is a *non*zero vector, this forces A^r to be the zero matrix. Hence, A is nilpotent.

Conversely, if A is nilpotent, then $A^r = 0$. Thus, there exists a nonzero vector $\vec{x}$ such that $A^r\vec{x}$ is the zero vector.

The upshot: a $n \times n$ matrix A is nilpotent if and only if there exists a nonzero vector $\vec{x} \in \mathbb{R}^n$ and a positive integer r such that $A^r\vec{x}$ is the zero vector.

(2) *Just because you can return doesn't mean you will*: Suppose A is a $n \times n$ matrix and $\vec{x}$ is a nonzero vector in $\mathbb{R}^n$. Suppose there exists a positive integer r such that $A^r\vec{x} = \vec{x}$. Since A^r sends a nonzero vector to itself, it must be the identity matrix. Thus, $A^r = I_n$. So, $A(A^{r-1}) = (A^{r-1})A = I_n$. Thus, A^{r-1} equals A^{-1}, so in particular, A is invertible.

Conversely, consider the case that A is an invertible $n \times n$ matrix. This means that we can recover the vector $\vec{x}$ from knowledge of the vector $A\vec{x}$. This means that if we apply A enough times to $A\vec{x}$, we get $\vec{x}$. So, there exists s such that $A^s(A\vec{x}) = A^{s+1}(\vec{x}) = \vec{x}$. Set $r = s + 1$, and we have that $A^r\vec{x} = \vec{x}$.

The upshot: a $n \times n$ matrix A is invertible if and only if it has the property that there exists a nonzero vector $\vec{x} \in \mathbb{R}^n$ and a positive integer r such that $A^r\vec{x} = \vec{x}$.

(3) *You can't see all the wonders of the world in a short life*: Let T be the rotation about the origin in $\mathbb{R}^2$ by a fixed angle θ. Starting with any nonzero vector $\vec{x}$, consider the sequence:

$$\vec{x}, T(\vec{x}), T(T(\vec{x})), \dots$$

When we rotate a vector, we preserve its length. Thus, the range of this sequence is the circle centered at the origin of radius equal to the length of $\vec{x}$.

For the following error-spotting exercises, use this (error-free) definition: Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, a (possibly zero, possibly nonzero) real number λ, and a nonzero vector $\vec{x} \in \mathbb{R}^n$, we say that $\vec{x}$ is an eigenvector of T with eigenvalue λ if $T(\vec{x}) = \lambda\vec{x}$.

(4) *What we can't achieve alone, we can do together*: Consider the case $n = 2$ and define T to be the linear transformation with matrix:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Note that T sends $\vec{e}_1$ to $\vec{e}_2$ and sends $\vec{e}_2$ to $\vec{e}_1$. In other words, T interchanges the two standard basis vectors. Since T does not preserve the lines of any of the standard basis vectors, neither of them is an eigenvector for T. Note that any vector would be a linear combination of the standard basis vectors, so T has no eigenvector.

(5) *Compatibility issues*: For any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, the set of eigenvectors of T, along with the zero vector, form a subspace of $\mathbb{R}^n$. Here's the proof. Note that:

- The zero vector is in the set by definition (although the zero vector is not considered an eigenvector, our definition here deliberately adds the zero vector in).
- Suppose vectors $\vec{u}$ and $\vec{v}$ are both eigenvectors for T. This means that there exists a real number λ such that $T(\vec{u}) = \lambda\vec{u}$ and $T(\vec{v}) = \lambda\vec{v}$, then $T(\vec{u} + \vec{v}) = \lambda\vec{u} + \lambda\vec{v}$ which becomes $\lambda(\vec{u} + \vec{v})$.
- Suppose $\vec{v}$ is an eigenvector for T with eigenvalue λ. Then, for any real number a, $T(a\vec{v}) = aT(\vec{v}) = a(\lambda\vec{v}) = \lambda(a\vec{v})$, so $a\vec{v}$ is also an eigenvector for T. Moreover, it has the same eigenvalue.

(6) *Don't be Procrustean!*: Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ with matrix:

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

The eigenvectors for this are $\vec{e}_1$ (with eigenvalue 1) and $\vec{e}_2$ (with eigenvalue 2). Note that there are no more eigenvectors. For instance, $\vec{e}_1 + \vec{e}_2$ is not an eigenvector because its image is $\vec{e}_1 + 2\vec{e}_2$, which is not a multiple of it.

(7) *A missed match*: Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ with matrix:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The matrix is diagonal, so the eigenvectors for this linear transformation are precisely the standard basis vectors $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$.

(8) *Zero's legit*: Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ with matrix:

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

This sends $\vec{e}_3$ to $\vec{e}_2$, sends $\vec{e}_2$ to $\vec{e}_1$, and sends $\vec{e}_1$ to the vector zero. Note that none of the standard basis vectors goes to itself, or for that matter, to a multiple of itself. In other words, T has no eigenvectors.