Please send any questions/comments/corrections to hhao@berkeley.edu.

Note: this study guide is a *very* condensed version of the course material (which will still end up being a lot). In particular, most statements won't be proved, and I won't spend space on giving many numerical examples. The goal is rather to give intuition to some of the more difficult concepts, as well as to show how all of the concepts are interconnected (as I think the course text manages to make linear algebra seem like a much more disjointed subject than it actually is). Therefore, most of the below material is in a different order than presented in the text, and I also don't guarantee that 100% of the material on the midterm will be discussed below.

We continue where we left off from Study Guide 1[1].

1 Bases (abstractly)

We clean up a loose but very important end from the previous study guide. Recall that the *dimension* of a vector space V is the size of a basis—all bases of V have the same finite size, or are infinite, so this makes sense. We will always assume that vector spaces have finite dimension (there is a theory for infinite-dimensional vector spaces, but it is more complicated). Of course we will not assume that V is a standard Euclidean space $\mathbf{R}^n$; it is simply some abstract set of elements (vectors) with the appropriate operations. It will be important to take this viewpoint of vectors as purely abstract points.

Recall that the intuition for dimension is the "size" of a vector space V, or in other words, how much "information" is needed to specify a vector in a V. For example, take $V = \mathbf{R}^2$. Then to specify a vector, I need to give two pieces of information: the x and y-coordinates. This corresponds to the fact that $\dim(\mathbf{R}^2) = 2$.

We can do the same procedure with other vector spaces. For instance, let V be the subspace of vectors $(x, y, z) \in \mathbf{R}^3$ such that $x + y + z = 0$. Then after *fixing* a basis, for instance $\{(1, -1, 0), (0, 1, -1)\}$ of V, we see that to specify a vector in V, I need to give two real numbers a and b, so that my choice corresponds to $a(1, -1, 0) + b(0, 1, -1)$.

The key point to emphasize from the previous example is that when you specify a vector in terms of "coordinates" or "coefficients of a linear combination", you are also *implicitly choosing a basis*. The slight issue with this is that *there is no preferred basis for a vector space*; that is, the basis you choose could be different from the basis I choose, so that our coordinates describing the same abstract vector end up being different. Even in $\mathbf{R}^n$ with its "standard basis", it's important to note that this is merely a *convenience*: there is no

[1] https://math.berkeley.edu/~hhao/teaching/54%20Spring%202024/Midterm%201%20Study%20Guide.pdf

necessary reason to use this basis instead of any other basis (in fact, as we will later see,there are many good reasons to *not* use this basis). The upshot is that:

in the absence of any other context, all bases are made equal.

This means that sometimes you should choose a more convenient basis for a computation over the "standard" basis. We will see more examples of this in the future.

Example 1.0.1. Consider the vector $v = e_1 + 2e_2$ inside $\mathbf{R}^2$. Of course, in the standard coordinates, v has the coordinate representation $(1, 2)$. That is, after choosing the standard basis as our "coordinate frame of reference", v is specified by the information $(1, 2)$. But suppose we chose a new basis $\{w_1, w_2\} = \{e_1 + e_2, e_2\}$. Then $v = w_1 + w_2$, so with this new basis as our "frame of reference", v is specified by the information $(1, 1)$. The point is that if you choose different bases, *two different sets of coordinates can define the same abstract vector.*

Example 1.0.2. Here is another example showing that there is no "preferred" choice of basis. Consider the vector space $\mathbf{P}_3$ of polynomials of degree at most 3, and consider the vector x^3+3x^2+3x+2. In the basis $\{x^3, x^2, x, 1\}$ of $\mathbf{P}_3$, our vector has coordinates $(1, 3, 3, 2)$. But in the alternate basis $\{(x+1)^3, (x+1)^2, (x+1), 1\}$, our vector has coordinates $(1, 0, 0, 1)$, which is certainly "simpler".

It is now natural to ask how to "change" coordinates from one basis from another. That is, suppose we have two different bases $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n\}$ of a vector space V. If a vector $x \in V$ has coordinates $(a_1, \ldots, a_n)$ in the v-basis (by which we mean $x = a_1v_1 + \ldots + a_nv_n$), then we want a general procedure to get the coordinates of x in the w-basis. Since we have to substitute some expressions in terms of the w_i's for each of the v_j's, what to do is now clear: we should write each v_j as a linear combination $v_j = \sum_{i=1}^n c_{ij}w_i$, and then write

$$x = \sum_{j=1}^n a_j \sum_{i=1}^n c_{ij}w_i = \sum_{i=1}^n \left(\sum_{j=1}^n c_{ij}a_j \right) w_i.$$

In particular, if C is the matrix with entries (c_{ij}), then $C(a_1, \ldots, a_n)^T$ is the vector whose components give the coordinates of x in the w-basis. This corresponds to the fact that the operation of "changing coordinates" is linear (this is simultaneously tautological and a bit confusing, since for the correct interpretation of this statement you would consider coordinates as a map $V \to \mathbf{R}^n$, which is linear).

Example 1.0.3. Let's look at a particular case of this. What if $V = \mathbf{R}^n$ and $w_1, \ldots, w_n$ is the standard basis $e_1, \ldots, e_n$? Then the matrix C as above is precisely the matrix with the v_i's as columns, so to transform a vector from v-coordinates to standard coordinates,

we simply apply that matrix. For instance, suppose $V = \mathbf{R}^2$, and $\{v_1, v_2\}$ is the basis $\{(1,2)^T, (2,1)^T\}$. Then the vector x with coordinates $(1,1)$ in the v-basis is

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

in standard coordinates. Indeed, $1 \cdot (1,2)^T + 1 \cdot (2,1)^T = (3,3)^T$.

WARNING: The method of the above example *only* works if you are transforming coordinates *from* some arbitrary basis *to* the standard basis. In general, you need to perform row-reduction, or someting similar, to write the v_i's in terms of the w_j's, but when the w_j's are the standard basis, this is tautological.

Example 1.0.4. Set $V = \mathbf{R}^2$ as above, with $\{v_1, v_2\} = \{(1,2)^T, (2,1)^T\}$ again. This time, take $\{w_1, w_2\} = \{(4,2)^T, (3,1)^T\}$. Then $v_1 = (5/2)w_1 - 3w_2$ and $v_2 = (1/2)w_1$. Therefore the change of basis matrix from v-coordinates to w-coordinates is

$$\begin{bmatrix} \frac{5}{2} & \frac{1}{2} \\ -3 & 0 \end{bmatrix}.$$

Now suppose we know the matrix C to transform from v-coordinates to w-coordinates. We may ask for the matrix D to transform from w-coordinates to v-coordinates. Keeping in mind with the general principle that matrices are functions, we see that the matrix D is the inverse of C: that is, if we apply the matrix C, then the matrix D, we have changed from v-coordinates to w-coordinates and back, so we should end up where we started. Therefore D must be the inverse of C! As an application, to change *from* standard coordinates to v-coordinates (the *opposite* of Example 1.0.3), we may alternatively write the vectors of the basis $\{v_1, \ldots, v_n\}$ as the columns of a matrix, and then take its inverse to be the desired change-of-basis matrix.

We will later see more reasons why this point of view is so powerful. For now, we can introduce a notion that we've sometimes been implicitly using, but is important to pin down:

Definition 1.0.5. Two vector spaces V, W are *isomorphic* (as vector spaces) if there is a bijective linear map $T : V \to W$; we call T an *isomorphism* (between V and W). We then write $V \cong W$.

This just means that for linear-algebraic purposes, V and W behave the same: there is a one-to-one and onto function between V and W, so that every $v \in V$ is associated to exactly one vector $w \in W$, such that the vector space structure is preserved. This means that *doing linear algebra in V is the same as doing linear algebra in W*, up to a change of coordinates (given by T).

The reason we should care about this notion is:

Theorem 1.0.6. Let V be a finite-dimensional vector space of dimension n. Then V is isomorphic to $\mathbf{R}^n$.

Therefore, once you know that Euclidean vector spaces $\mathbf{R}^n$, you've really met all the finite-dimensional vector spaces already!

Example 1.0.7. The complex numbers $\mathbf{C}$ are a 2-dimensional vector space over $\mathbf{R}$ (it has basis $\{1, i\}$), so is isomorphic to $\mathbf{R}^2$ as vector spaces.

WARNING: Even after this theorem, it is incorrect to think that it is sufficient to only work with the Euclidean vector spaces. For instance, even though all 2-dimensional subspaces of $\mathbf{R}^3$ are isomorphic to $\mathbf{R}^2$, it often matters very much *which specific subspace* we're working with: the subspace $V = \{(x, y, 0) : x, y, \in \mathbf{R}\}$ might have very different properties (in the context of your problem) than the subspace $W = \{(x, 0, z) : x, z \in \mathbf{R}\}$! So even though V and W have the exact same properties as vector spaces (e.g. it would be valid to say that 3 distinct vectors in V must be linearly dependent, as that is true for $\mathbf{R}^2$), they are distinct objects inside a larger space and must be treated on an unequal footing when working inside $\mathbf{R}^3$.

2 Rank-Nullity

As an aside from our basis discussion, we mention a very powerful theorem that should have been included in Study Guide 1.

Consider the following situation: we have a linear transformation T between finite-dimensional vector spaces V and W. We want to discuss possibilities for T: given some knowledge about V and W (say, their dimensions), can we tell if T can be injective or surjective? Can we say that certain vectors in V are in the kernel of T, or certain vectors in W are in the image? If I have a vector $w \in W$ in the image of T, how can I "reconstruct" what vectors in V map to w under T? And most generally, what "information" about V does T preserve?

Of course, it is very difficult to say anything concretely about this. For instance, there is always the zero map $0 : V \to W$ that sends everything in V to the zero vector in W; this is definitely not injective nor surjective (unless V or W are the zero vector space), and the map "forgets" a lot above V, since T squashes all vectors in V to a single point. But let's consider a more illuminating example:

Example 2.0.1. Consider the map $T : \mathbf{R}^2 \to \mathbf{R}^2$ given by $(x, y) \mapsto (x, 0)$. Geometrically, this projects every point in the plane vertically onto the x-axis. Therefore T "remembers" the data of the x-coordinate, but "forgets" the data of the y-coordinate, since the output will always have y-coordinate 0. In other words, if we draw vertical lines $x = c$ in the plane

(draw this picture for yourself!), for each real number c, T only remembers which vertical line an input point (x, y) is on.

From this geometric picture, we see that the dimension of $\mathrm{Im}(T)$ is 1: we can think of T as assigning the point $(c, 0)$ to each member $\{x = c\}$ of a 1-dimensional collection of vertical lines. The kernel of T is the 1-dimensional line $\{x = 0\}$, and this corresponds to T forgetting one dimension of information: the vertical line. We then have the equality

$$\dim(\mathbf{R}^2) = 2 = 1 + 1 = \dim(\ker(T)) + \dim(\mathrm{Im}(T)),$$

which is to say: of the original 2 parameters needed to specify a vector in $\mathbf{R}^2$, T retained 1 parameter and forgot 1.

This is the intuition for the following theorem:

Theorem 2.0.2 (Rank-Nullity)**.** Let $T : V \to W$ be a linear transformation between vector spaces, where V is finite dimensional. Then

$$\dim(V) = \dim(\ker(T)) + \dim(\mathrm{Im}(T)).$$

The interpretation is the same as sketched out in the rather pedantic Example 2.0.1: if we have a linear map T coming from a dimension $\dim(V)$ vector space V, then T has to retain k dimensions of information (the dimension of $\mathrm{Im}(T)$) and forget l dimensions of information (the dimension of $\ker(T)$), and the only way this makes sense is if $\dim(V) = k + l$.

Remark 2.0.3. In the special case when V and W are Euclidean vector spaces, and $T : \mathbf{R}^n \to \mathbf{R}^m$ is represented by a matrix transformation A (so A is $m \times n$), then the rank-nullity theorem can be written as

$$n = \dim(N(A)) + \dim(C(A)).$$

The number $\dim(C(A))$ has a special name; it is called the *rank* of the matrix A. The number $\dim(N(A))$ is (less commonly) called the *nullity*, hence the name of the theorem.

From row-reduction, we can check that $\mathrm{rank}(A) = \mathrm{rank}(A^T)$, even if A is not square. This is sometimes very useful.

Here is a basic application of the theorem that you already know:

Example 2.0.4. If $n < m$, there is no surjective map from $\mathbf{R}^n$ to $\mathbf{R}^m$, because $n = \dim(N(A)) + \dim(C(A))$ means $\dim(C(A)) \leq n < m$, so $C(A)$ can never be the full m-dimensional space. For similar reasons, if $n > m$, there is no injective map from $\mathbf{R}^n$ to $\mathbf{R}^m$, because $\dim(C(A)) \leq m < n$, so $\dim(N(A))$ must be strictly greater than 0, and there is a nonzero vector in $N(A)$.

Here are two trickier examples that display the power of this theorem:

Example 2.0.5. We claim that there is a nonzero polynomial $f(x)$ of degree at most 2 such that $\int_0^1 f(x)e^x \sin(x)dx = 0$ and $\int_1^2 f(x)e^x \sin(x)dx = 0$. This looks like it has nothing to do with linear algebra, but notice that the map

$$g(x) \mapsto \int_a^b g(x)e^x \sin(x)dx = 0$$

is *linear* in the argument g, no matter what the bounds of integration a and b are. We can then construct the linear map

$$\mathbf{P}_2 \to \mathbf{R}^2, \quad g(x) \mapsto \left(\int_0^1 g(x)e^x \sin(x)dx, \int_1^2 g(x)e^x \sin(x)dx \right).$$

The kernel of this map is precisely the set of polynomials that satisfy the property we want! But because $\mathbf{P}_2$ has dimension 3 and $\mathbf{R}^2$ has dimension 2, rank-nullity (Example 2.0.4) shows that the kernel must be of dimension at least 1; i.e. it must contain some vector other than the zero vector. This is the polynomial f that we want.

Notice that we never actually wrote down what f is, and it is also impossible (or extremely difficult) to solve this problem using calculus techniques alone.

Example 2.0.6. Let's consider the vector space $\mathbf{R}^{m\times n}$ of $m \times n$ matrices under the usual operations of matrix addition and scalar multiplication, and $v \in \mathbf{R}^n$ be a fixed nonzero vector. Let

$$W = \{M \in \mathbf{R}^{m\times n} : Mv = 0\}$$

be the subspace of matrices that send v to 0. This is a subspace, and our goal is to compute $\dim(W)$.

Notice that the map $T : \mathbf{R}^{m\times n} \to \mathbf{R}^n$ given by $M \mapsto Mv$ is linear (make sure you understand what this means! The *matrix* is now the argument to the map, not the vector!), with nullspace precisely W. Since v is nonzero, the image of T is all of $\mathbf{R}^n$ (convince yourself why!). Therefore the image of T has dimension n, and $\dim(\mathbf{R}^{m\times n})$ is visibly mn (a basis is given by matrices that have a 1 in some position and zeros elsewhere), so that $\dim(W) = mn - n = m(n-1)$ by rank-nullity.

3 Eigen-stuff

Given a random $n \times n$ matrix, it is usually very difficult to give a (simple) geometric description of what it does. On the other hand, giving a geometric description of a diagonal matrix is very easy: we simply scale each of the axes by the corresponding element on the diagonal. Likewise, the image of any standard basis vector is a multiple thereof. The goal of all of the "eigen-stuff" is to generalize this concept. In light of *scaling* being the key concept, we define:

Definition 3.0.1. Let $T : V \to V$ be a linear map from a vector space to itself. An *eigenvector* of T is any *nonzero* vector $v \in V$ such that $Tv = \lambda v$ for some scalar λ. We say that λ is an *eigenvalue* of T and v is a λ-eigenvector (or similar terminology). If λ is an eigenvalue of T, the subspace of all vectors $w \in V$ such that $Tw = \lambda w$ is called the λ-*eigenspace* of T.

There are three important things to note about the definition. First, T must be a linear transformation from a vector space to itself—in the language of matrices, this means that it only makes sense to talk about eigenvalues when we have a square matrix. Second, to avoid degenerate cases, we don't allow 0 to be an eigenvector, since $T\mathbf{0} = \lambda\mathbf{0}$ for all λ (and we want each eigenvector to correspond to exactly 1 eigenvector). Third, any scalar multiple[2] of an eigenvector is again an eigenvector with the same eigenvalue. Therefore when we say that "v is an eigenvalue of T with eigenvalue λ," we are also implicitly saying that any nonzero scalar multiple of v satisfies the same.

Remark 3.0.2. In fact, the correct point of view is to consider *complex-linear* maps of *complex vector spaces*—that is, we should be doing linear algebra over the complex numbers instead of the real numbers, allowing matrices, vectors, and scalars with complex entries. This ensures that every linear map $V \to V$ has the "correct" number of eigenvalues in the finite-dimensional case (which turns out to be $\dim(V)$), which makes the theory much more cohesive. In fact, even matrices with real entries can have complex eigenvalues, as we shall see, so it doesn't make sense to exclude complex numbers from our discussion, even from the beginning.

In these notes, to avoid any confusion, I will stick to real eigenvalues and eigenvectors besides mentioning some examples in passing. However, you should know that the entire theory for real eigenvalues works in the same way for complex eigenvalues—I will deliberately be vague about whether I mean real scalars or complex scalars, to emphasize this point. The book has some strange and very ad hoc discussion on complex eigenvalues in the 2×2 case that is not worth repeating here. You should read Section 5.5 of the book for the exam, and immediately forget it afterwards.

Example 3.0.3. A linear map $T : V \to V$ is invertible if and only if it doesn't have 0 as an eigenvalue. Indeed, we know that T being invertible is equivalent to it being injective, and being injective is equivalent to $Tv = 0 = 0v$ only having the zero solution $v = 0$. Since the zero vector is never an eigenvector, we conclude that invertibility implies 0 is not an eigenvalue, and vice versa.

Example 3.0.4. The only eigenvalue of the identity matrix (of any dimension) is 1.

Example 3.0.5. Let V be the vector space of smooth functions, and T the differentiation operator $T : V \to V$. Then every $\lambda \in \mathbf{R}$ is an eigenvalue of T, because the corresponding eigenvector is $e^{\lambda x}$.

[2]From now on, when we say "scalar multiple," we will usualy mean a *nonzero* scalar multiple.

Example 3.0.6. A real-valued matrix need not have any eigenvalues (this is why only allowing complex eigenvalues is very necessary). For instance, consider any rotation matrix

$$A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

for an angle $0 < \theta < 2\pi$. This is invertible, so by Example 3.0.3 it cannot have 0 as an eigenvalue. By the very definition of an eigenvector, we know that if a vector v in the plane $\mathbf{R}^2$ is an eigenvector of A with nonzero eigenvalue, then Av points in the same direction as v (thinking about the arrow depiction of a vector). But A is a *rotation* by a nonzero angle, so Av cannot point in the same direction as v. The conclusion is that A *has no eigenvectors or eigenvalues*!

Example 3.0.7. If A is an $n \times n$ matrix and λ is an eigenvalue of A, then λ^n is an eigenvalue of A^n for any positive integer n (also for any negative integer n if A is invertible). More generally, if $p(x) = a_n x^n + \ldots + a_0$ is a polynomial, and we "plug in A into this polynomial" to obtain a matrix

$$p(A) = a_n A^n + a_{n-1} A^{n-1} + \ldots + a_1 A + a_0 I,$$

then $p(\lambda)$ is an eigenvalue of $p(A)$.

In the best-case scenario, V has a *basis of eigenvectors*: we mean that there is a basis $\{v_1, \ldots, v_n\}$ of V where each v_i satisfies $Tv_i = \lambda_i v_i$ for some scalar λ_i. This means that if we use the above eigenbasis for V, then the linear transformation T "looks diagonal."

To be more concrete, let's think about the case when $V = \mathbf{R}^n$ and T is represented by a square matrix A. Let S be the change-of-basis matrix from $\{v_1, \ldots, v_n\}$ to the standard basis, so the columns of S are the v_i. If D is the diagonal matrix with the λ_i on the diagonal, then $A = SDS^{-1}$. Indeed, for each v_i,

$$SDS^{-1}v_i = SDe_i = S\lambda_i e_i = \lambda_i(Se_i) = \lambda_i v_i = Av_i$$

by the definition of S. Then SDS^{-1} and A agree on a basis, so they must be the same linear map.

This relation is so important that it deserves a name:

Definition 3.0.8. Let A and B be two $n \times n$ matrices. If there is an invertible matrix S such that $A = SBS^{-1}$, then we say A and B are *similar* matrices.

The intuition for similar matrices is that

Similar matrices represent the same abstract linear transformation, but in different coordinate systems.

Note that if A and B are similar matrices, with $A = SBS^{-1}$, then A^n and B^n are similar, with $A^n = SB^nS^{-1}$. In fact $p(A) = Sp(B)S^{-1}$ for any polynomial $p(x)$ (see Example 3.0.7).

From the above discussion, we see that if $\mathbf{R}^n$ has a basis of eigenvectors of A, then A is similar to a diagonal matrix. The converse is also true, since if $A = SDS^{-1}$, then the columns of S given a basis of $\mathbf{R}^n$ consisting of eigenvectors of A. In this case:

Definition 3.0.9. We say a square $n \times n$ matrix A is *diagonalizable* if it is similar to a diagonal matrix. **This is equivalent to $\mathbf{R}^n$ having a basis of eigenvectors of A.**

The notion of diagonalizability is subtle. For instance, there is no correlation between invertibility and diagonalizability. As an example, the 2×2 identity matrix is certainly diagonalizable—it is already diagonal. But the matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ is invertible yet not diagonalizable: its only eigenvector is $(1, 0)^T$ with corresponding eigenvalue 1, so the eigenvectors do not form a basis of $\mathbf{R}^2$.

On the other hand, we can say:

Proposition 3.0.10. If $v_1, \ldots, v_k$ are eigenvectors of $T : V \to V$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then the v_i are linearly independent.

Corollary 3.0.11. If $T : V \to V$ has $n = \dim(V)$ distinct eigenvalues, then T is diagonalizable, because we get n linearly independent eigenvectors of T, which necessarily form a basis in n-dimensional space.

Of course, the converse of this statement is not true. For example, the 2×2 identity matrix only has 1 as an eigenvalue (clear from the definition), but is diagonalizable. To guide your intuition, you should have examples and counterexamples in your head regarding each of the statements about eigenvalues and diagonalizability in this section.

3.1 The Characteristic Polynomial

To actually find eigenvalues of linear transformations represented by matrices, we make the following observation. Suppose $Av = \lambda v$ for some nonzero vector v. Then $(A - \lambda I)v = 0$, so that $A - \lambda I$ is a matrix with nonzero kernel. This implies that it has determinant 0, so that for any eigenvalue λ, we have $\det(A - \lambda I) = 0$. The converse is also true, and so we define:

Definition 3.1.1. The *characteristic polynomial* of a matrix A is the degree-n polynomial $\mathrm{char}_A(\lambda) = \det(A - \lambda I)$.

The eigenvalues of A are therefore the roots of $\mathrm{char}_A(\lambda)$. Of course, the characteristic polynomial need not have any real roots—consider for instance a quadratic polynomial with negative discriminant (this is again why we should be working in the complex numbers). It can also have repeated roots. If λ_1 is a root of char_A of multiplicity k, then we say λ_1 is an eigenvalue of A of *algebraic multiplicity* k.

Example 3.1.2. The eigenvalues of the matrix $A = \begin{bmatrix} 1 & 2 \\ -2 & 5 \end{bmatrix}$ are the roots to

$$\det(A - \lambda I) = (1 - \lambda)(5 - \lambda) + 4 = \lambda^2 - 6\lambda + 9.$$

This has 3 as a double root, so 3 is an eigenvalue of algebraic multiplicity 2.

Example 3.1.3. The eigenvalues of the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ are the roots to

$$\det(A - \lambda I) = \lambda^2 + 1.$$

This has complex roots $\pm i$, both of which appear with algebraic multiplicity 1.

Example 3.1.4. The eigenvalues of any triangular matrix A are the entries a_{ii} on the diagonal, because $\det(A - \lambda I) = (\lambda - a_{11})(\lambda - a_{22}) \cdot \ldots \cdot (\lambda - a_{nn})$.

The following result may also be helpful in checking your eigenvalue computations:

Proposition 3.1.5. Let A be an $n \times n$ matrix. The product of all eigenvalues (**including the complex ones if they exist**) is equal to $\det(A)$. The sum of all eigenvalues is equal to $\sum_{i=1}^{n} a_{ii}$, the sum of the elements on the main diagonal of A (this is called the *trace* of A, and is another important invariant associated to a matrix).

Also, remember that if two matrices are similar, they are related by a change of basis. That is, they represent the same linear transformation *geometrically*, but in different coordinate systems. Eigenvalues are also *geometric* concepts: they are abstractly defined as "how much some vector scales under a linear transformation," which is not a definition that depends on the coordinate system we choose. Therefore it should follow that eigenvalues are preserved under change-of-basis, and this is true:

Proposition 3.1.6. Similar matrices have the same characteristic polynomial, and hence the same eigenvalues with the same algebraic multiplicities.

Beware that the eigenvectors change, since we have changed to a different coordinate system.

3.2 How to Diagonalize

Let's go back to geometry, and consider when it is possible to diagonalize an $n \times n$ matrix A. By definition, this occurs when $\mathbf{R}^n$ has a basis of eigenvectors of A. Therefore if $\lambda_1, \ldots, \lambda_k$ are the eigenvalues of A (there need not be n distinct eigenvalues!), we need that the dimensions of the λ_i-eigenspaces sum to n, so that an eigenbasis of $\mathbf{R}^n$ is given by putting together the individual bases for each eigenspace (note that linear independence is guaranteed by Proposition 3.0.10). This motivates the following definition:

Definition 3.2.1. Let A be a matrix with eigenvalue λ. The *geometric multiplicity* of λ is the dimension of the λ-eigenspace. By the very definition of eigenvalue, the geometric multiplicity is always at least 1.

Proposition 3.2.2. If λ is an eigenvalue of A, then the geometric multiplicity of λ is *at most* the algebraic multiplicity.

So suppose a matrix A has n (real) eigenvalues, counted with algebraic multiplicity. By this, we mean that the algebraic multiplicities of the eigenvalues sum up to n. This is clearly a necessary condition for diagonalizability, since as we saw earlier, the diagonal matrix that A is similar to must have the eigenvalues of A, with the correct multiplicities, as its diagonal entries. Now, we know from Proposition 3.2.2 that for each eigenvalue λ, the size of the basis for the λ-eigenspace is at most the algebraic multiplicity of λ. But for diagonalizability to occur, the sum of the sizes of each of those individual bases must equal n, the same as the sum of all the algebraic multiplicites. The only way for this to occur is if the geometric multiplicities are the same as the algebraic multiplicities:

Diagonalizability means geometric multiplicity equals algebraic multiplicity for every eigenvalue, and the sum of the multiplicities is n.

Example 3.2.3. Suppose we have a 4×4 matrix A with 3 distinct eigenvalues $1, 2, 3$ (not counting with algebraic multiplicity). Then if d_1, d_2, d_3 are the geometric multiplicities of the 1, 2, and 3-eigenspaces respectively, we have the inequalities $d_1, d_2, d_3 \geq 1$ and $d_1+d_2+d_3 \leq 4$.

Suppose further that we know that $d_1 = 2$. Then we must have $d_2 = d_3 = 1$, so that $d_1 + d_2 + d_3 = 4$ and A is diagonalizable. We can also deduce that the algebraic multipliticies of $1, 2, 3$ must be $2, 1, 1$ respectively.

Finally, let's go through a full example of diagonalizing a 3×3 matrix.

Example 3.2.4. Let's try to diagonalize $A = \begin{bmatrix} 0 & -4 & -6 \\ -1 & 0 & -3 \\ 1 & 2 & 5 \end{bmatrix}$. The characteristic polynomial is

$$\det(A - \lambda I) = \det \begin{bmatrix} -\lambda & -4 & -6 \\ -1 & -\lambda & -3 \\ 1 & 2 & 5-\lambda \end{bmatrix} = \lambda^2(5-\lambda) + 12 + 12 - 6\lambda - 6\lambda - 4(5-\lambda)$$

$$= -\lambda^3 + 5\lambda^2 - 8\lambda + 4.$$

It is easy to see that 1 is a root of this polynomial, and then

$$-\lambda^3 + 5\lambda^2 - 8\lambda + 4 = -(\lambda - 1)(\lambda^2 - 4\lambda + 4) = -(\lambda - 1)(\lambda - 2)^2.$$

Therefore A has eigenvalues 1 (algebraic multiplicity 1) and 2 (algebraic multiplicity 2).

To see whether A is diagonalizable or not, we need to see whether the geometric multiplicty of 2 is 1 or 2, since we know that 1 has geometric multiplicity 1. The 2-eigenspace is equivalently the nullspace of $A - 2I$, which is

$$A - 2I = \begin{bmatrix} -2 & -4 & -6 \\ -1 & -2 & -3 \\ 1 & 2 & 3 \end{bmatrix}.$$

All the columns are a scalar multiple of the first, so $A-2I$ has rank 1, and thus its nullspace has dimension $3 - 1 = 2$ by rank-nullity. Therefore the geometric multiplicity of 2 is 2, verifying that A is diagonalizable.

To actually complete the diagonalization of A, we need to find bases of the 1 and 2-eigenspaces. Via row reduction or some other method, we see that the nullspace for $A - 2I$ has basis $\{(1, 1, -1)^T, (2, -1, 0)^T\}$. Next, we have

$$A - I = \begin{bmatrix} -1 & -4 & -6 \\ -1 & -1 & -3 \\ 1 & 2 & 4 \end{bmatrix},$$

which we know has 1-dimensional nullspace. It is then not too hard to see that the 1-eigenspace has basis $\{(2, 1, -1)^T\}$, so that an eigenbasis for $\mathbf{R}^3$ is given by

$$\{(2, 1, -1)^T, (1, 1, -1)^T, (2, -1, 0)^T\}.$$

Therefore

$$A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & -1 \\ -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & -1 \\ -1 & -1 & 0 \end{bmatrix}^{-1}.$$

4 Inner Products

A vector space is a rather "bare" structure. That is, using only the abstract vector space operations given to us (addition and scaling of vectors), we have no way of measuring geometric notions like length, area, angle, etc. To fix this, we need to define an additional operation, called an *inner product*.

Definition 4.0.1. Let V be a vector space, not necessarily finite-dimensional. An *inner product* on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbf{R}$ (taking in two vectors) satisfying the following:

1. Symmetry: $\langle v, w \rangle = \langle w, v \rangle$ for all vectors v and w.

2. Bilinearity: $\langle v+v', w\rangle = \langle v, w\rangle + \langle v', w\rangle$ and $\langle cv, w\rangle = c\langle v, w\rangle$ for a scalar c. That is, if we hold one of the arguments fixed, the inner product is linear in the other argument (linearity in the second argument follows from symmetry).

3. Positive-definiteness: $\langle v, v\rangle \geq 0$ for all vectors v, and equality $\langle v, v\rangle = 0$ occurs if and only if $v = 0$.

If V has an inner product, we call V an inner product space.

Here are some examples:

Example 4.0.2. If $V = \mathbf{R}^n$, the usual dot product $v \cdot w = \sum_{i=1}^n v_i w_i$ is an inner product on $\mathbf{R}^n$

Example 4.0.3. Suppose $V = \mathbf{P}_n$, the space of polynomials of degree at most n. Let $a_1, \ldots, a_{n+1}$ be $n+1$ distinct real numbers. Then

$$\langle p, q\rangle = \sum_{i=1}^{n+1} p(a_i)q(a_i)$$

is an inner-product on $\mathbf{P}_n$. Positive-definitness is the hardest condition to check (it usually is), and here it follows because if a degree-n polynomial has at least $n+1$ roots, then it must be the 0 polynomial.

Example 4.0.4. Suppose $V = C[0, 1]$, the space of continuous functions on the unit interval. We define an inner product on V to be

$$\langle f, g\rangle = \int_0^1 fg.$$

You should convince yourself why this is an inner product. Inner products similar to this one are the basis for the theory of *Fourier series*, to be discussed at the end of the course.

Note that a vector space can be equipped with many inner products. For instance, if $\langle \cdot, \cdot\rangle$ is an inner product on a vector space V, then any scaling $\langle \cdot, \cdot\rangle' = c\langle \cdot, \cdot\rangle$ is another inner product on V. Again, you should have examples of many different types of inner products on different types of vector spaces, not just the standard dot product on $\mathbf{R}^n$.

As promised, the inner product allows us to define lengths and (certain) angles.

Definition 4.0.5. The *norm*, or *length*, of a vector v in an inner product space is

$$\|v\| = \langle v, v\rangle^{1/2}.$$

A vector of norm 1 is called a *unit vector*, or *normalized*. If v is any nonzero vector, then the unit vector $\frac{1}{\|v\|}v$ is the *normalization* of v. The *distance* between two vectors v and w is $\|v - w\| = \|w - v\|$.

For example, in $\mathbf{R}^n$ with the standard dot product, $\|v\| = \sum_{i=1}^n v_i^2$, and so vectors of norm 1 satisfy $\sum_{i=1}^n v_i^2 = 1$.

The norm of a vector satisfies the usual properties we want "length" to have, even in more abstract situations. For example, we know that for any scalar c, $\|cv\| = |c| \, \|v\|$, and we also have the

Proposition 4.0.6 (Triangle inequality)**.** For all vectors v and w in an inner product space, we have

$$\|v + w\| \leq \|v\| + \|w\| .$$

This follows from the Cauchy-Schwarz inequality:

Proposition 4.0.7 (Cauchy-Schwarz)**.** For all vectors v and w in an inner product space, we have

$$|\langle v, w \rangle| \leq \|v\| \, \|w\| .$$

Moreover, equality holds if and only if one of v, w is a scalar multiple of the other.

Example 4.0.8. As a sample application of Cauchy-Schwarz, we will show that for positive real numbers $x_1, \ldots, x_n$, the inequality

$$(x_1 + \ldots + x_n)(1/x_1 + \ldots + 1/x_n) \geq n^2$$

always holds. This doesn't look like it has anything to do with linear algebra, but consider the vectors $(\sqrt{x_1}, \ldots, \sqrt{x_n})$ and $(1/\sqrt{x_1}, \ldots, 1/\sqrt{x_n})$ in $\mathbf{R}^n$. Then applying Cauchy-Schwarz to these two vectors gives the desired inequality.

We now turn to angles.

Definition 4.0.9. A set of vectors $\{v_1, \ldots, v_k\}$ is *orthogonal* if for all $i \neq j$, $\langle v_i, v_j \rangle = 0$. We say that the set $\{v_1, \ldots, v_k\}$ is an *orthogonal set*. If the v_i are also all unit vectors, we say that the set is *orthonormal*.

In $\mathbf{R}^n$ with the dot product, the notion of "orthogonal" corresponds to the intuitive geometric notion of perpendicularity. Then from our intuition with Euclidean space, we expect that orthogonal vectors are linearly independent—they should "point in different directions." This is true:

Proposition 4.0.10. An orthogonal set of vectors $\{v_1, \ldots, v_k\}$ in an inner product space V is linearly independent. In particular, if an orthogonal/orthonormal set of vectors spans V, then we call it an *orthogonal/orthonormal basis*.

Example 4.0.11. In $\mathbf{R}^n$, the set of standard basis vectors is an orthonormal basis. Another orthogonal basis of $\mathbf{R}^2$ is given by $\{(1,1),(-1,1)\}$. This is not orthonormal, but we can normalize the vectors to produce an orthonormal basis $\{(\sqrt{2}/2,\sqrt{2}/2),(-\sqrt{2}/2,\sqrt{2}/2)\}$.

In $\mathbf{R}^n$ with the dot product, we can also define the *angle* between two vectors more generally. The acute angle θ between vectors v, w is given by

$$\cos(\theta) = \frac{v \cdot w}{\|v\| \, \|w\|}.$$

This is often very useful to know.

Example 4.0.12. Consider the vector space $\mathbf{P}_3$ equipped with the inner product $\langle p, q\rangle = \int_{-1}^{1} pq$. The first four *Legendre polynomials*,

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2 - 1), \quad P_3(x) = \frac{1}{2}(5x^3 - 3x)$$

form an orthonormal basis for $\mathbf{P}_3$.

We can also generalize the concept of orthogonality to subspaces.

Definition 4.0.13. Let V be a finite-dimensional inner product space. If W and W' are subspaces of V, we say that W is orthogonal to W' if for all $w \in W, w' \in W'$, we have $\langle w, w'\rangle = 0$ (that is, every vector in W is orthogonal to every vector in W'). Similarly, for a vector $v \in V$, we say v is orthogonal to W if $\langle w, v\rangle = 0$ for all $w \in W$.

We write $W^\perp$ for the subspace $\{v \in V : \langle w, v\rangle = 0\}$ consisting of *all* vectors in V orthogonal to W.

You should convince yourself that if W and W' are orthogonal subspaces, then $W \cap W' = \{0\}$, since the only vector orthogonal to itself is the 0 vector (**this is a common trick to show that a vector is 0, and is worth remembering, along with the trick that the only vector with norm 0 is 0**). You should also convince yourself that $(W^\perp)^\perp = W$ (**WARNING:** This is not true in infinite-dimensional inner product spaces!), and that $\dim(W) + \dim(W') = \dim(V)$.

Example 4.0.14. In $\mathbf{R}^3$, the subspaces $\{(x, y, 0) : x, y \in \mathbf{R}\}$ and $\{(0, 0, z) : z \in \mathbf{R}\}$, representing the xy-plane and the z-axis respectively, are orthogonal complements; they only intersect at $(0, 0, 0)$ and their dimensions add to 3. This is the intuition you should keep in mind when thinking about orthogonal subspaces.

Example 4.0.15. If A is a matrix, then $N(A) = (\mathrm{Row}(A))^\perp$ and $C(A) = N(A^T)$. This is easy to see from the definitions.

The following theorems generalize what you know from Euclidean space, and are worth remembering.

Theorem 4.0.16 (Pythagorean theorem)**.** If v and w are orthogonal vectors in an inner product space, then $\|v+w\|^2 = \|v\|^2 + \|w\|^2$.

Theorem 4.0.17 (Parseval's identity)**.** If $\{v_1, \ldots, v_n\}$ is an orthonormal basis of an inner product space V, then for any $v \in V$, we have

$$v = \sum_{i=1}^{n} \langle v, v_i \rangle v_i,$$

and thus

$$\|v\|^2 = \sum_{i=1}^{n} \langle v, v_i \rangle^2.$$

Parseval's identity follows from the Pythagorean theorem. It's good to keep in mind the special case of Parseval's identity when $V = \mathbf{R}^n$ and the orthonormal basis is the standard basis.

4.1 Orthogonal Projections and Gram-Schmidt

Now that inner products supply us with a notion of distance, we may proceed to ask some natural questions. For instance, suppose we have an inner product space V, a vector $v \in V$, and a subspace W of V. We can ask for the "closest vector in W to v": the vector $w_0 \in W$ that minimizes $\|w - v\|$ as w ranges over all vectors $w \in W$, if it exists. When V is finite-dimensional, such a vector exists and is unique:

Proposition 4.1.1. Let V be a finite dimensional inner product space, let W be a subspace, and let $v \in V$ be a vector. Then v can be written *uniquely* in the form $w_0 + z$, where $w_0 \in W$ and $z \in W^\perp$. The vector w_0, denoted $\text{proj}_W v$, is the *projection* of v onto W, and it minimizes the distance $\|w - v\|$ as w ranges over all vectors $w \in W$ (in particular, if $w' \in W$ is not $\text{proj}_W v$, we have a strict inequality $\|w' - v\| > \|\text{proj}_W v - v\|$. This minimized distance $\|\text{proj}_W v - v\|$ is defined to be the distance of v to W.

If $\{w_1, \ldots, w_k\}$ is an *orthogonal* basis of W, then

$$\text{proj}_W v = \sum_{i=1}^{k} \frac{\langle w_i, v \rangle}{\langle w_i, w_i \rangle} w_i.$$

WARNING: The above formula does not work if the w_i are not orthogonal!

Example 4.1.2. Let's do a numerical example in the simplest case when $V = \mathbf{R}^2$ and W is just a line; i.e. it is spanned by a single vector w. In this case we also write $\text{proj}_w v$ for

$\text{proj}_W v$, and call it the projection of v onto w. Let $v = (2, 0)$ and $w = (-1, -1)$, so W is the line $y = x$. Then

$$\text{proj}_w v = \frac{v \cdot w}{w \cdot w} w = \frac{-2}{2} w = (1, 1).$$

You should draw the picture and convince yourself that the line segment from v to $\text{proj}_w v$ is perpendicular to the line W, so that it represents the shortest distance (which is $\|(1, -1)\| = \sqrt{2}$) from v to W.

Of course, in order to apply Proposition 4.1.1, we need to actually begin with an orthogonal basis of W. We know that we can find *some* basis of W, but there is no guarantee that it be orthogonal. Luckily, there is a procedure to produce an orthogonal (even orthonormal) basis from any basis, called the *Gram-Schmidt* procedure. Let's describe it:

1. Let V be an inner product space with basis $\{v_1, \ldots, v_n\}$ (this doesn't need to be a basis, only a linearly independent set, in which case it is a basis for its span). We want to produce an orthogonal basis from this basis.

2. Set $w_1 = v_1$.

3. Set $w_2 = v_2 - \text{proj}_{w_1} v_2$. By Proposition 4.1.1, w_2 is orthogonal to w_1.

4. Set $w_3 = v_3 - \text{proj}_{\text{span}(w_1, w_2)} = v_3 - \text{proj}_{w_1} v_3 - \text{proj}_{w_2} v_3$. By Proposition 4.1.1, w_3 is orthogonal to w_1 and w_2.

5. In general, for $k \geq 2$, if we have produced $w_1, \ldots, w_{k-1}$, then we set

$$w_k = v_k - \text{proj}_{\text{span}(w_1, \ldots, w_{k-1})} = v_k - \text{proj}_{w_1} v_k - \ldots - \text{proj}_{w_{k-1}} v_k.$$

   Basically, to get w_k, we keep subtracting off the projections of v_k onto $w_1, \ldots, w_{k-1}$. This ensures that w_k is orthogonal to $w_1, \ldots, w_{k-1}$.

6. Once we're done, we have an orthogonal basis $\{w_1, \ldots, w_n\}$ of V. We can normalize each vector (divide out by the norm) and produce an orthonormal basis.

Remark 4.1.3. It is advised to do all the normalization at the very end, in order to avoid working with many square roots and fractions in the intermediate steps.

Remark 4.1.4. Notice that the Gram-Schmidt process **depends on the order in which you write your basis**! That is, if we performed the Gram-Schmidt process on two bases that were the same as sets but *different* as *ordered* lists, we would end up with different results.

The book has many worked-out examples of this process in action, so I won't provide another one here. The annoying thing about Gram-Schmidt is that it is a very computationally heavy process, with many places to make a silly arithmetic mistake, so it's good to practice on your own when $\dim(V)$ is small (e.g. 3) to get a feel for the process.

On the other hand, the Gram-Schmidt process is very important *theoretically*. For instance, because of this process and the fact that every vector space has *some* basis, we conclude:

Proposition 4.1.5. Every (finite-dimensional) inner product space has an *orthonormal* basis.

This is convenient, since it is much easier to work with an orthonormal basis instead of some arbitrary basis.

Moreover, if we dig into the details of the Gram-Schmidt process beginning with a basis $\{v_1, \ldots, v_n\}$ and ending with an orthonormal basis $\{w_1, \ldots, w_n\}$, we conclude that

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(w_1, \ldots, w_k)$$

for all $1 \leq k \leq n$. That is, the Gram-Schmidt process preserves the span of the first k vectors in our list, for any k. Now, let's work in $\mathbf{R}^n$ with the standard dot product. Suppose A is the $n \times n$ invertible matrix with the v_i's as columns, and Q is the $n \times n$ invertible matrix with the w_i's as columns. Then $R \coloneqq Q^{-1}A$ is by definition the change-of-basis matrix from the v-coordinate system to the w-coordinate system. But because each v_k is a linear combination of $w_1, \ldots, w_k$ by the above equality of spans, we conclude that the kth column in R has 0 entries after the kth row, so that R must be upper triangular! We conclude:

Proposition 4.1.6 (QR Factorization)**.** If A is an invertible matrix, we can write A as a product QR of square matrices, where the columns of Q are orthonormal and R is upper triangular.

Matrices that share the properties of Q are so important we give them a special name:

Definition 4.1.7. If Q is a square matrix with *orthonormal* columns with respect to the standard dot product on $\mathbf{R}^n$, then Q is called an *orthogonal matrix* (the terminology is a bit unfortunate). Equivalently, an orthogonal matrix Q is a matrix that satisfies $Q^{-1} = Q^T$.

The second definition is the preferred definition since it is more compact.

Orthogonal matrices are generalizations of rotation matrices, in that they *preserve lengths and angles*. By that, I mean:

Proposition 4.1.8. If Q is an orthogonal matrix, then for any vectors $v, w \in \mathbf{R}^n$, we have $Qv \cdot Qw = v \cdot w$. In particular, if $v = w$, we deduce that $\|Qv\| = \|v\|$ for all vectors v.

This also implies that the real eigenvalues of an orthogonal matrix are ± 1, since if v is an eigenvector of Q with eigenvalue λ, then $\|v\| = \|Qv\| = \|\lambda v\| = |\lambda| \, \|v\|$.

Because of their defininining property $Q^{-1} = Q^T$, systems involving orthogonal matrices are very easy to solve: the system $Qx = b$ has the solution $x = Q^{-1}b = Q^T b$, which is convenient because finding the transpose of a matrix is much easier than finding its inverse. In particular, if we already knew the QR factorization of a matrix A, then solving $Ax = b$ is the same as solving $Rx = y$, where $y = Q^T b$. But $Rx = y$ is easy to solve with back substitution, because R is already upper triangular (in fact this is sometimes faster than row reduction even if you don't know the QR factorization, as evidenced by the fact that modern computer algebra software uses a combination of row-reduction and QR-factorization techniques, depending on the situation). As an example:

Example 4.1.9. Suppose the 3×3 matrix A has QR factorization

$$A = QR = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

To solve the linear system $Ax = b$, where $b = (3, 3, 3)^T$, we first set $y = Q^{-1}b = Q^T b$. Then

$$y = \frac{1}{3} \begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \\ 2 & -2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \\ 2 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix}.$$

We then need to solve $Rx = y$. R is already upper-triangular, so we only need to back-substitute to find that the unique solution for x is $(7, -2, 1)^T$.

5 Odds and Ends

In this section, we discuss a few final items that might appear on the midterm.

5.1 Least Squares

Recall that we can solve the linear system $Ax = b$ exactly when b is in the column space of A. Of course, if A is not surjective, this might not occur, so we can ask for the "best possible approximate solution" in the case $b \notin C(A)$. In other words, we are asking for the $x \in \mathbf{R}^n$ such that $\|Ax - b\|$ is minimized (equivalently, such that $\|Ax - b\|^2$ is minimized, hence "least squares"). By Proposition 4.1.1, we are to find the x such that Ax is $\text{proj}_{C(A)} b$. Note that if A is not injective, then there will be more than one possible value of x. The method to do this is via the normal equations:

Proposition 5.1.1. The *normal equations* for the system $Ax = b$ are given by the system

$$A^T Ax = A^T b.$$

This system is always consistent because $\text{rank}(A^T A) = \text{rank}(A^T)$. Any solution $\widehat{x}$ to this system is a least-squares solution to $Ax = b$.

In the case that A is injective, then $A^T A$ is invertible, and the unique least-squares solution is given by

$$\widehat{x} = (A^T A)^{-1} A^T b.$$

Notice that if b is actually in $C(A)$, then the least-squares solutions are just the usual solutions to $Ax = b$, since any such solution has $\|Ax - b\| = 0$, which is obviously minimal.

For the exam, it's worth just memorizing this method and doing a numerical example or two, since there isn't really a new concept introduced here.

5.2 Linear Models

Due to lack of time and interest, I'm not going to say anything about this section. You should just read Section 6.6 of the book to prepare for the midterm, and then you can forget what you've read. Sorry!

5.3 Symmetric Matrices and Quadratic Forms

The theory of eigenvalues/eigenvectors of a symmetric matrix is particularly interesting. For instance, because $A = A^T$, we can show that eigenvectors corresponding to distinct eigenvalues of a symmetric matrix are *orthogonal* (much stronger than "linear independence", as in Proposition 3.0.10). The amazing fact is that

Theorem 5.3.1 (Spectral Theorem)**.** Any symmetric matrix is *orthogonally diagonalizable*. This means that $\mathbf{R}^n$ has an *orthonormal basis* $\{v_1, \ldots, v_n\}$ of eigenvectors of A, so A can be written as

$$A = SDS^{-1} = SDS^T = \sum_{i=1}^{n} d_{ii} v_i v_i^T$$

for D diagonal and S *orthogonal*. Any of the three equivalent forms above for A is called a spectral decomposition of A.

The spectral theorem is difficult to prove, so it is more important that you just know the result. To find the spectral decomposition of a symmetric matrix, you just use the same diagonalization algorithm that you always use.

Finally, we turn to the concept of quadratic forms.

Definition 5.3.2. A *quadratic form* in variables $x_1, \ldots, x_n$ is a polynomial in the x_i where every term has degree 2.

For example, $x_1x_2 - x_3^2 + 4x_1x_4$ is a quadratic form. On the other hand, $x_1^2 + 2$ and $x_1x_2 + x_3x_4 - x_5$ are not quadratic forms, because they contain terms of degree 0 and 1, respectively.

Suppose we consider the x_i as the components of an n-vector x. Now, if A is any symmetric $n \times n$ matrix, then x^TAx is a 1×1 matrix, which we can consider as a number. It is:

$$x^TAx = \sum_{i,j=1}^{n} a_{ij}x_ix_j.$$

Therefore the map $(x_1, \ldots, x_n) \mapsto x^TAx$ is by definition a quadratic form (note that this map is definitely not linear, even though matrices and matrix products show up). Moreover, by suitable choices of the a_{ij}, we find that every quadratic form arises in this way. For example:

Example 5.3.3. Consider the quadratic form $x_1x_2 - x_3^2 + 4x_1x_4$ mentioned previously. To write it in the form $x \mapsto x^TAx$, we should have a_{ii} be the coefficient of x_i^2 for each i, and $a_{ij} = a_{ji}$ (remember A should be symmetric) be *half* of the coefficient of x_ix_j for each pair of indices $i \neq j$ (make sure you understand why there is a difference)! Therefore with

$$A = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 2 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix},$$

the quadratic form x^TAx is the same as $x_1x_2 - x_3^2 + 4x_1x_4$.

We now explain why we insist on using symmetric matrices. First, if the matrices are symmetric, we have a one-to-one association between quadratic forms and symmetric matrices that doesn't work if we allow more general matrices, since we have ambiguity in the choice of a_{ij} and a_{ji} for each pair $i \neq j$ (recall that in Example 5.3.3, $a_{ij} = a_{ji}$ was forced to be half of the corresponding coefficient in the quadratic form). In short,

Quadratic forms with n variables have a one-to-one correspondence to symmetric $n \times n$ matrices.

Second, the spectral theorem allows us to make *orthogonal* changes of basis to write our quadratic form in a more convenient form.

Definition 5.3.4. If $f(x_1, \ldots, x_n)$ is a quadratic form, then a *cross-term* is any term not of the form cx_i^2; that is, cross-terms are terms of the form dx_ix_j for $i \neq j$.

A quadratic form can be most easily analyzed when there are *no cross terms*, so that none of the x_i "interact" with each other, and the quadratic form is just a sum of scalar multiples of squares of the x_i's. By the method shown in Example 5.3.3, this corresponds to the matrix A being diagonal.

So suppose we start with an arbitrary quadratic form $(x_1, \ldots, x_n) \mapsto x^T Ax$. Because A is symmetric, the spectral theorem tells us that there is an orthogonal matrix Q and a diagonal matrix D such that $A = QDQ^{-1}$, so that $Q^{-1}AQ = D$. Then if we write $x = Qy$, our quadratic form is

$$(x_1, \ldots, x_n) \mapsto x^T Ax = (Qy)^T AQy = y^T Q^T AQy = y^T Q^{-1} AQy = y^T Dy.$$

The point is that if we introduce new variables y_i, governed by the relation $x = Qy \Leftrightarrow Q^{-1}x = y$, our quadratic form has no cross-terms when expressed in terms of the variables y_i. As an example:

Example 5.3.5. Consider the quadratic form

$$x_1^2 - 4x_1x_2 + 4x_2x_3 - x_3^2.$$

This is represented by the symmetric matrix

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

which has spectral decomposition

$$A = QDQ^{-1} = \begin{bmatrix} -1/3 & -2/3 & 2/3 \\ -2/3 & 2/3 & 1/3 \\ 2/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} -3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -1/3 & -2/3 & 2/3 \\ -2/3 & 2/3 & 1/3 \\ 2/3 & 1/3 & 1/3 \end{bmatrix}^{-1}.$$

Therefore if

$$y = Q^{-1}x = \left(\frac{-x_1 - 2x_2 + 2x_3}{3}, \frac{-2x_1 + 2x_2 + x_3}{3}, \frac{2x_1 + x_2 + x_3}{3} \right),$$

then

$$x_1^2 - 4x_1x_2 + 4x_2x_3 + x_3^2 = -3y_1^2 + 3y_2^2,$$

where the right-hand side has no cross terms.

Notice that the coefficients of the y_i's are precisely the eigenvalues of A. It follows that if all eigenvalues are positive (or nonnegative, nonpositive, negative), then the quadratic form f also takes only positive (or nonnegative, nonpositive, negative) values for nonzero arguments, regardless of the values of the x_i's. Indeed, this is clear if the inputs are in terms of $(y_1, \ldots, y_n)$, but each $(x_1, \ldots, x_n)$ is associated to a unique $(y_1, \ldots, y_n)$ by Q ("change of coordinates"). This deserves a special name:

Definition 5.3.6. A quadratic form $f(x_1, \ldots, x_n)$ is

1. Positive definite if whenever the x_i are not all zero, then $f(x_1, \ldots, x_n)$ is strictly positive.
2. Positive semidefinite if whenever the x_i are not all zero, then $f(x_1, \ldots, x_n)$ is nonnegative.
3. Negative semidefinite if whenever the x_i are not all zero, then $f(x_1, \ldots, x_n)$ is nonpositive.
4. Negative definite if whenever the x_i are not all zero, then $f(x_1, \ldots, x_n)$ is strictly negative.

If A is the symmetric matrix associated to the quadratic form f, then these 4 cases occur when all the eigenvalues of A are:

1. Strictly positive.
2. Nonnegative.
3. Nonpositive.
4. Strictly negative.

Finally, if f (or A) does not fall in any of the above categories, then f can take on both positive and negative values, and A must have a mix of both positive and negative eigenvalues. In this case, we call f and A *indeterminate*.

For example, the quadratic form in Example 5.3.5 is indeterminate, because A has eigenvalues 3 and -3. Indeed, f is positive if $y_1 = y_3 = 0$ and $y_2 = 1$, while it is negative if $y_2 = y_3 = 0$ and $y_1 = 1$.

Remark 5.3.7. Those of you that have taken Math 53 may recall that the positive/negative definiteness of the symmetric second-derivative Hessian matrix is used to determine the types of extrema of a multivariate function. For example, positive-definiteness of the Hessian indicates a local minimum, while an indeterminate Hessian indicates a saddle point.

WARNING: You *cannot* determine whether a quadratic form is positive/negative (semi)definite by just looking at the signs of coefficients in the quadratic form or of the entries in the associated symmetric matrix A. You must find the eigenvalues of A. For example, the quadratic forms $f(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2$ and $g(x_1, x_2) = x_1^2 + 3x_1x_2 + x_2^2$ look similar. But f is positive definite because $f(x_1, x_2) = x_1^2/2 + 2(x_1/2 + x_2/2)^2 + x_2^2/2$, while g is indeterminate because $g(1, 0) = 1$ and $g(-1, 1) = -1$.