MATH 583A
REVIEW SESSION #1

BOJAN DURICKOVIC

1. Vector Spaces

Very quick review of the basic linear algebra concepts (see any linear algebra textbook): (finite dimensional) *vector space* (or *linear space*), *subspace*, *linear independence*, *basis*, *dimension*.

1.1. **Coordinatization.** Choosing a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for a vector space V enables us to express any vector $\mathbf{x} \in V$ in terms of the elements of the basis:

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{v}_i \,.$$

Thus, we can identify the (abstract) vector $\mathbf{x}$ with a (concrete) column matrix $[x_1, \dots, x_n]^T$, having coefficients in this expansion (*coordinates*) for elements. The set of all column matrices, denoted F^n (F can be either $\mathbb{R}$ or $\mathbb{C}$), forms a vector space which is isomorphic to the abstract vector space V, but we always need to keep in mind the basis with respect to which the coordinatization was carried out. (Coordinates always imply a system of reference!)

There are many possible bases for any vector space, and the matrix representation will be different moving from one basis to another. However, it is easily verified that the transition from one basis to another translates into a linear transformation of the corresponding matrices. Let $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be two bases for the vector space V, and let s_{ij} be the coordinates of $\mathbf{v}_j$ in the basis of $\mathbf{u}$'s,

$$\mathbf{v}_j = \sum_{i=1}^{n} s_{ij} \mathbf{u}_i$$

The matrix S therefore has coordinates of $\mathbf{v}_j$'s for columns. Let $\mathbf{x} \in V$ be any vector; we denote its coordinates in the $\mathbf{u}$- and $\mathbf{v}$-basis by $\mathbf{x}(\mathbf{u}) = [\xi_1, \dots, \xi_n]^T$ and $\mathbf{x}(\mathbf{v}) = [\eta_1, \dots, \eta_n]^T$, respectively. It is easy to verify that

$$\xi_i = \sum_{j=1}^{n} s_{ij} \eta_j \,,$$

i.e. S is the transition matrix from coordinates in the $\mathbf{v}$-basis to coordinates in the $\mathbf{u}$-basis:[1]

$$\mathbf{x}(\mathbf{u}) = S\mathbf{x}(\mathbf{v})$$

Date: September 1, 2006.

[1] This is intuitively clear when you think of the multiplication $S\mathbf{x}$ as a linear combination of columns of S with coefficients x_i. Then x_i are coordinates in the basis of columns of S, and the resulting product will be the coordinates in whatever basis the columns of S are expressed.

The inverse transformation is obtained by applying the inverse matrix:

$$\mathbf{x}(\mathbf{v}) = S^{-1}\mathbf{x}(\mathbf{u})$$

2. Linear Operators

A *linear operator* $A : V \to W$ (V, W vector spaces) is a map that preserves sums and scalar multiples, i.e.

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$$
$$A(\alpha\mathbf{x}) = \alpha A\mathbf{x}$$

$\forall \mathbf{x}, \mathbf{y} \in V, \alpha \in F$.

The *range* (also called *column space* with matrices) of the operator is its range as a function,

$$\operatorname{ran} A = \{A\mathbf{x} \mid \mathbf{x} \in V\} \leq W$$

The *kernel* (or *null space*) of the operator is the subset of its domain that is mapped to the null vector,

$$\ker A = \{\mathbf{x} \in V \mid A\mathbf{x} = 0\} \leq V$$

The linearity of the operator ensures that both the range and the kernel are always subspaces of W and V, respectively. The dimensions of these subspaces are related in the following way.

Proposition.

$$\dim(\operatorname{ran} A) + \dim(\ker A) = \dim V$$

Note that the dimension of the image space W has no bearing in the relation.[2]

The following nice property of linear operators is another direct consequence of their linearity:

Proposition. *A linear operator is uniquely determined by its action on a basis.*

In other words, given a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V, if we specify the images of the elements of the basis $A\mathbf{v}_i = \mathbf{w}_i \in W$, $i = 1, \ldots, n$, the operator is completely determined, since this definition is *extended by linearity* to all vectors in V:

$$A\mathbf{x} = A \sum x_i \mathbf{v}_i = \sum x_i A\mathbf{v}_i = \sum x_i \mathbf{w}_i$$

A linear operator $A : V \to V$ (from a vector space to itself!) is *non-singular* if it is invertible, i.e. if $\exists A^{-1} : V \to V$ such that $AA^{-1} = A^{-1}A = I$. We can tell whether an operator is non-singular by simply looking at its null space:

Proposition. *An operator $A : V \to V$ is non-singular if and only if its kernel is trivial, i.e.* $\ker A = \{0\}$.

---

[2] Think of it as if reducing the dimension of W will tend to reduce the dimension of $\operatorname{ran} A$, while increasing the dimension of $\ker A$.

2.1. **Matrix Representation.** Analogously to the identification of vectors and column matrices for a chosen basis, we can coordinatize a linear operator. Since a linear operator is uniquely determined by its action on the basis, its coordinate representation can be taken as the collection of n column vectors which are the coordinate representations of the vectors $\mathbf{w}_i = A\mathbf{v}_i \in W$, $i = 1, \dots, n$. Combining these n m-dimensional column vectors into a rectangular matrix proves to be a useful representation, since the action of the operator on a vector translates into matrix multiplication.

Again, it should be stressed that although the space of linear operators $V \to W$ is entirely isomorphic to the space of rectangular matrices of corresponding dimensions, the actual correspondence between an operator and a matrix always requires the knowledge of the bases both in V and W. Dealing with abstract operators has the advantage that we don't need to specify bases, while matrices are needed for any practical computation.

The matrix representation of an operator will be different from one pair of bases to another, generally requiring two transformation matrices S and T, for the bases in V and W respectively. (Transformation matrices are always square, but here S and T may have different dimensions if $\dim V \neq \dim W$.) Most frequently, we are dealing with operators from the vector space V to itself, and in that case there is only one basis involved: the transformation of coordinates is then performed by a *similarity transformation*

$$A' = S^{-1}AS$$

The way to think of this transformation is the following: A and A' are matrix representations of an operator in the basis $\mathbf{u} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ and $\mathbf{v} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, respectively, and S is the transformation matrix from the basis $\mathbf{v}$ to the basis $\mathbf{u}$, i.e. its columns are the coordinates of $\mathbf{v}_j$ in the basis $\mathbf{u}$. Reading $S^{-1}AS$ from right to left, we have:

(1) multiplication by S: transformation from basis $\mathbf{v}$ to basis $\mathbf{u}$;
(2) multiplication by A: performing the action of the operator in the basis $\mathbf{u}$;
(3) multiplication by S^{-1}: transforming back from basis $\mathbf{u}$ to basis $\mathbf{v}$.

Although the concrete matrix representation depends on the basis (or basis pair), the properties of the matrix that remain unchanged in a change of basis (i.e. that are invariant with respect to similarity transformations in case of square matrices) are worth identifying: these are the properties of the abstract operator itself.

2.2. **Rank of an operator.** The (column) *rank* of a matrix is the number of linearly independent columns. The rank does not change in similarity transformations (prove this!), thus it is a property of the abstract operator.

Proposition. *The row rank of a matrix is equal to its column rank.*

Proof. Exercise. □

A matrix is said to be (or to have) *full rank* if it has a maximal rank. An $m \times n$ matrix, where $m > n$, is full rank if all its (n) columns are linearly independent. For square matrices, full rank is actually a characterization of non-singularity:

Proposition. *A square matrix is full rank if and only if it is non-singular.*

Full rank and non-singularity are also operator properties independent of the choice of basis. Other properties that don't change in similarity transformations include the determinant and trace. Therefore, although the determinant and trace are computed from a matrix representation, it makes perfect sense to speak of a determinant and trace of an abstract operator.

2.3. **Characteristic polynomial.** The *characteristic polynomial* of a square matrix A is

$$\det(A - \lambda I)$$

The characteristic polynomial of an $n \times n$ matrix is a polynomial of order n:

$$c_A(\lambda) = c_n \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_0$$

Proposition. *The characteristic polynomial of a matrix A has the following coefficients:*

$$c_n = (-1)^n\,, \quad c_{n-1} = (-1)^{n-1}\operatorname{tr} A\,, \quad c_0 = \det A$$

The above proposition can be useful, in particular for quickly computing the characteristic polynomial of a 2×2 matrix:

$$c_A(\lambda) = \lambda^2 - \lambda \operatorname{tr} A + \det A$$

Using the Binet-Cauchy theorem ($\det(AB) = \det A \det B$), it is easy to show the following:

Proposition. *Similar matrices share the same characteristic polynomial.*

Thus, we can speak of a *characteristic polynomial of an operator*, which is independent of the operator's particular matrix representation.

Theorem (Hamilton-Cayley)**.** *Every square matrix cancels its own characteristic polynomial, i.e.*

$$c_A(A) = 0$$

This is an important theorem that we will be using for computing the function of a matrix.

3. Spectral Theory

3.1. **Eigenvalue problem.** Given an operator $A : V \to V$, [3] the *eigenvalue problem* consists in finding scalars λ (*eigenvalues*) and non-zero vectors $\mathbf{v}$ (*eigenvectors*) such that

$$A\mathbf{v} = \lambda\mathbf{v}$$

Such a pair $(\lambda, \mathbf{v})$ is collectively termed an *eigenpair*. The set of all eigenvalues of A is called the *spectrum* of A (notation: $\sigma(A)$).

Proposition. *Eigenvectors corresponding to different eigenvalues are linearly independent.*

The existence of eigenvalues is guaranteed by the following theorem.[4]

[3] Note that the eigenvalue problem makes sense only for operators from the space V to itself, i.e. for square matrices!

[4] Note, however, that, except for low dimensions, this is not how eigenvalues are actually computed in practice, due to the prohibitively high cost of computing a determinant. Usually, they are found using some matrix iteration scheme, until they start appearing on the diagonal.

Theorem. *A scalar λ is an eigenvalue of the operator A if and only if λ is a root of the characteristic polynomial,*

$$c_A(\lambda) = 0$$

Proof.

$$\begin{aligned} \lambda \text{ eigenvalue} &\iff (A - \lambda I)v = 0\,, \quad v \neq 0 \\ &\iff \det(A - \lambda I) = 0 \\ &\iff c_A(\lambda) = 0 \end{aligned}$$

□

Thus, a linear operator acting on an n-dimensional vector space has exactly n (not necessarily distinct) eigenvalues. The *algebraic multiplicity* of the eigenvalue is its multiplicity as the root of the characteristic polynomial.

By definition, the existence of eigenvalues implies the existence of corresponding eigenvectors. Every eigenvector spans a one-dimensional *invariant subspace*. However, there may be several eigenvectors corresponding to one eigenvalue, spanning an invariant subspace of dimension greater than one. The dimension of this space is called the *geometric multiplicity* of the eigenvalue.

Proposition. *The geometric multiplicity of an eigenvalue is less than or equal to its algebraic multiplicity.*

Therefore, although there is at least one eigenvector for each eigenvalue, some eigenvalues with multiplicity larger than one may run short of eigenvectors to match their multiplicity. The eigenvalues for which the geometric multiplicity is less than the algebraic one are called *defective*. A defective operator/matrix is one having one or more defective eigenvalues.

3.2. **Diagonalization.** An operator is *diagonalizable* if there is a basis in which the operator is represented by a diagonal matrix. (A matrix is diagonalizable if it can be reduced to diagonal form by similarity transformations.)

Theorem. *An operator $A : V \to V$ is diagonalizable if and only if there is a basis of V composed of eigenvectors of A.*

Proof. (sketch, one direction only) Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of eigenvectors of A. Combining the corresponding column vectors into a matrix $S = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ we have a full rank matrix, which is therefore invertible, and we got ourselves a similarity transformation $S^{-1}AS$. We claim that the result is a diagonal matrix. The matrix AS has for columns $A\mathbf{v}_j = \lambda_j \mathbf{v}_j$, thus can be rewritten as $AS = S\Lambda$ (operations on columns = multiplication from the right; columns are just multiplied by scalars $\implies$ Λ is diagonal). Therefore $S^{-1}AS = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. □

A useful way to think of this is by looking at a transformation to the eigenvector basis:

$$A\mathbf{x} = S\Lambda S^{-1}\mathbf{x}$$

The action of A is therefore obtained by transforming to the eigenvector basis with S^{-1}, in which the action consists simply in multiplying each coordinate with the corresponding eigenvalue (hence multiplication with a diagonal matrix), and then transforming back to the original basis with S.[5]

An immediate consequence is that if all eigenvalues are distinct, the operator is diagonalizable, since there are n (linearly independent) eigenvectors.

Note that defective operators are not diagonalizable — they are short of eigenvectors.

3.3. **Normal operators.** A *normal operator* $N : V \to V$ is an operator that commutes with its adjoint[6],

$$NN^\dagger = N^\dagger N$$

Proposition. *Let N be a normal operator. If (λ, v) is an eigenpair of N, then $(\overline{\lambda}, v)$ is an eigenpair of $N^\dagger$.*

Proof. Exercise. □

The most useful property of normal operators on complex vector spaces is given by the Spectral Theorem:[7]

Theorem (Spectral Theorem for Normal Operators)**.** *An operator $N : V \to V$ is normal if and only if there is an orthonormal basis of V composed of eigenvectors of N.*

Therefore, normal operators on complex spaces are diagonalizable. The situation is somewhat more complicated in the real case, since in that case the characteristic polynomial may have less than n *real* roots. This is not a concern with self-adjoint operators. . .

3.4. **Self-adjoint operators.** An operator L is *self-adjoint* (or *Hermitian*) if $L^\dagger = L$. Self-adjoint operators are a special case of normal operators, hence they are diagonalizable. They also have the additional nice property that all their eigenvalues are real. This theorem is frequently used:

Theorem (Spectral Theorem)**.** *Let $L : V \to V$ be a self-adjoint operator. Then:*

(1) The eigenvalues of L are real.
(2) There is an orthonormal basis of V composed of eigenvectors of L.

3.5. **Computing the function of a matrix.** Taylor series enable us to give meaning to a function of a square matrix, since we know how to compute powers of a matrix. If the matrix is in diagonal form, this task is very simple:

$$\Lambda^k = \operatorname{diag}(\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k)\,,$$

so that

$$f(\Lambda) = \operatorname{diag}(f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n))\,.$$

---

[5] If you are having problems remembering whether A is $S^{-1}\Lambda S$ or $S\Lambda S^{-1}$, where the columns of S are eigenvectors, remember that you need to have $S\Lambda$ in there, since you need to multiply eigenvectors with corresponding eigenvalues, and *columns* are scaled by multiplying *from the right* with a diagonal matrix.

[6] The *adjoint* $A^\dagger$ (sometimes also denoted A^*) of a matrix A is its complex conjugate transpose, $A^\dagger = \overline{A}^T$.

[7] It is called "Spectral Theorem" since it allows us to decompose the entire space V into mutually orthogonal invariant subspaces of N. The words "spectral", "invariant subspace" and the prefix "eigen" often denote the same context. It's all about the eigenvalue problem.

But how to compute the function of a non-diagonal matrix? The obvious approach is diagonalizing (assuming the matrix is diagonalizable), computing the function of a diagonal form, and then transforming back to the original basis. However, there is a way of computing this without performing the change of basis. To do this, we don't need the actual eigenvectors of the matrix, but only its eigenvalues.

Consider the exponential function to be specific. Its Taylor series converges on the whole set of real numbers:

$$\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad \forall x \in \mathbb{R}.$$

The exponential of a matrix is thus defined as

$$\mathrm{e}^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

We can compute this infinite series in a finite number of operations by using the Hamilton-Cayley theorem: since the matrix A cancels its characteristic polynomial, its nth power can be expressed in terms of lower powers:[8]

$$A^n = \sum_{j=0}^{n-1} c_j A^j$$

The series thus reduces to a finite sum:

$$\mathrm{e}^A = \sum_{k=0}^{n-1} \xi_k A^k = \xi_{n-1} A^{n-1} + \cdots + \xi_1 A + \xi_0 I,$$

where ξ_k are some coefficients to be determined. These are found by applying the equation to all n eigenvectors, yielding a linear system for $\xi_0, \ldots, \xi_{n-1}$ involving the n eigenvalues:

$$\begin{aligned}
\mathrm{e}^{\lambda_1} &= \xi_{n-1}\lambda_1^{n-1} + \cdots + \xi_1\lambda_1 + \xi_0 \\
\mathrm{e}^{\lambda_2} &= \xi_{n-1}\lambda_2^{n-1} + \cdots + \xi_1\lambda_2 + \xi_0 \\
&\vdots \\
\mathrm{e}^{\lambda_n} &= \xi_{n-1}\lambda_n^{n-1} + \cdots + \xi_1\lambda_n + \xi_0
\end{aligned}$$

You should recognize the Vandermonde matrix. Its determinant will be nonzero if all eigenvalues are different; for now our method is limited to this case. Then we have a unique solution for $\xi_0, \ldots, \xi_{n-1}$.

We have thus found e^A using nothing but the eigenvalues of A, and all that needed to be computed are the first $n-1$ powers of A and the solution of one $n \times n$ system.

Exercise. Compute the exponential of the matrix

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

(a) by diagonalizing and undiagonalizing; (b) by the above method.

---

[8] The coefficients c_j here coincinde possibly up to a sign with the coefficients of the characteristic polynomial: remember that the leading coefficient of c_A is $(-1)^n$.