Class schedule: Algebra Bridge Course, Summer 2024

All slides, papers, and book chapters will be made available on the course webpage.

WEEK 1

Thurs. June 27. Welcome, introductions, course overview, discussion about our graduate program, etc.

Fri. June 28. Four ways to think about the fundamental problem in linear algebra: solving linear equations in n variables.

(1) Matrix form
(2) Row picture
(3) Column picture
(4) Grid picture

The example of $\{2x - y = 0, -x + 2y = 3\}$ was used. Then we did a 3×3 example: $\{2x - y = 0, -x + 2y - z = -1, -3y + 4z = 4\}$. Note how changing up the RHS changes the planes (row picture), but barely changes the column or grid pictures.

Then, we talked about 4 ways to multiply matrices $AB = C$, where A is $m \times n$ and B is $n \times p$.

(1) *Rows times columns*: $C_{ij} = (\text{row } i) \cdot (\text{column } j) = \sum_{k=1}^{n} a_{ik} b_{kj}$.
(2) *By columns*: $A[b_1 \cdots b_p] = [Ab_1 \cdots Ab_p]$. Each column Ab_i is a linear combination of the columns of A.
(3) *By rows*: $\begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} B = \begin{bmatrix} a_1^T B \\ \vdots \\ a_m^T B \end{bmatrix}$. Each row $a_j^T B$ is a linear combination of the rows of B.
(4) *Columns times rows.* This is a sum of n^2 rank-1 matrices, $\sum_{j,k=1}^{n} a_j^T b_k$.

WEEK 2

Mon. July 1. The *four fundamental subspaces.* Given an $m \times n$ matrix A, we introduced the subspaces

(1) *Column space* $C(A)$ in $\mathbb{R}^m$
(2) *Row space* $C(A^T)$ in $\mathbb{R}^n$
(3) *Nullspace* $N(A)$ in $\mathbb{R}^n$
(4) *Left nullspace* $C(A^T)$ in $\mathbb{R}^m$.

We stated, without proving (will do later, in more generality) that $C(A)$ and $C(A^T)$ have the same rank, and that these subspaces come in orthogonal complement pairs:

$$\mathbb{R}^n = C(A) \oplus N(A^T), \qquad \mathbb{R}^m = C(A^T) \oplus N(A).$$

We proved that $N(A) = N(A^T A)$ under "undergraduate notation" (dot products). Then, we discussed inner products, and used $(x, y) = y^T Ax$ as an example, for $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. It was slightly cleaner to prove $N(A) = N(A^T A)$ with this notation.

We discussed how to solve an inhomogeneous system $Ax = b$. The "general solution" has the form $x = x_n + x_p$, where $Ax_n = 0$ and x_p is "any particular solution." As an application, we solved the ODE $y'' + 4y = 8$, which is the inhomomgeneous equation $Ly = 8$, for the linear operator $L = \frac{d^2}{dt^2}$. More generally, the nullspace of an n^{th} order linear differential operation is n-dimensional.

Finally, compared and contrasted systems of equations $Ax = b$, where A is "tall and skinny," (generally, no solution) vs. short and wide (0 or infinitely many solution).

Tues. July 2. *Projections and least squuares.* We started by projecting a vector b onto another vector a. The result $p = xa$ satisfies $b = p + e$, and we derived $x = (a^T b)/(a^T a)$. Alternatively, we could describe $b \overset{P}{\longmapsto} p$ with the rank 1 matrix $P = (aa^T)/(a^T a)$. This matrix satisfies $P^T = P$ and $P^2 = P$, which are defining properties of projection matrices.

More generally, projections arise if we want to solve an underdetermined system $Ax = b$, where $b \notin C(A)$. The "best fit" solution is to solve $A\hat{x} = p$, where p is the projection of b onto $C(A)$. We showed how can be done by solving $A^T A\hat{x} = A^T b$. Reason: If $A = [a_1 \cdots a_n]$ and we write $A\hat{x} = p = \hat{x}_1 a_1 + \cdots + \hat{x}_n a_n$, then $a_i \perp e = (b - A\hat{x})$ means that $a_i^T(b - A\hat{x}) = 0$, which gives the equation $A^T(b - A\hat{x}) = 0$.

As a takeaway message, if S is a subspace with basis $a_1, \ldots, a_r$, then the projection matrix onto SS is $P = A(A^T A)^{-1} A^T$, where $A = [a_1 \cdots a_r]$.

We finished by showing how a classic least squares problems can be viewed as an overdetermined linear system, using an example with three points $(1,1)$, $(2,2)$, and $(3,2)$. Any line $b = C + Dt$ through these points would lead to a linear system $\{C + D = 1,\ C + 2D = 2,\ C + 3D = 2\}$ that has no solution. However, by solving $A^T A\hat{x} = A^T b$ instead, we find that the best fit line is when $C = 2/3$ and $D = 1/2$. We concluded with a remark that we *can* find a degree-2 polynomial $b = C + Dt + Et^2$ that fits this data, because this would lead to a 3×3 system, which has a unique solution.

Wed. July 3. *Orthogonality, least squares, and QR factorization.* We reviewed what it means for a set of vectors to be orthogonal, and orthonormal. A square matrix Q is orthogonal if $Q^T Q = I$, which is equivalent to its columns being orthonormal.

Next, we discussed how to decompose a vector into an orthogonal basis. As an example, note that

$$v = (4,3) = 4e_1 + 3e_2 = (v \cdot e_1)e_1 + (v \cdots e_2)e_2.$$

If we use a different basis, like $v_1 = (\sqrt{2}/2, \sqrt{2}/2)$ and $v_2 = (\sqrt{2}/2, \sqrt{2}/2)$.

$$v = (v \cdot v_1)v_1 + (v \cdots v_2)v_2 = 4.95v_1 + 0.701v_2.$$

We can do this with an orthogonal basis $w_1, \ldots, w_n$ that isn't orthonormal by replacing $v \cdot w_i$ with $(v \cdot w_i)/(w_i \cdot w_i)$.

An example application of this is Fourier series. If $f(x)$ is a piecewise continuous 2π-periodic function, then it can be decomposed uniquely into a sum $f(x) = \frac{a_0}{2} + \sum a_n \cos(nx) + b_n \sin(nx)$. There are some technical details that require analysis to formalized, such as what happens at the points of discontinunity, an the fact that infinite sums are allowed. This all works because the set $\{\frac{1}{\sqrt{2}}, \cos(nx), \sin(nx) \mid n \in \mathbb{Z}\}$ is an orthonormal basis with respect to the inner product $\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)g(x)\,dx$.

We introduced the Gram-Schmidt process, which takes an independent set of vectors, and outputs an orthonormal set. Finally, we mentioned how the Gram-Schmidt process can be described in matrix language. Specificaly, if $M = [a_1, \ldots, a_n]$ is the original basis, and $Q = [q_1, \ldots, q_n]$ the orthonormal basis from computing Gram-Schmidt, then these are related by $M = QR$, where $r_{ij} = q_i \cdot a_j$.

Thurs. July 4. Holiday; no class.

Fri. July 5. Holiday; no class

WEEK 3

Mon. July 8. *Vector spaces and subspaces.* We gave the formal definition of a vector space. For this, we needed the formal definition of a group and a field. We discussed how to formally prove a few "obvious" facts, such as uniqueness of an identity element, uniqueness of inverses, and that $0x = \mathbf{0}$ for all vectors $x \in X$. We formally defined linear maps in this setting, though we discusse linearity earlier.

Next, we discussed subspaces and their sums, such as the difference between $Y + Z$ and $Y \oplus Z$. The former can *always* be defined, whereas the latter occurs only when $Y \cap Z = \{0\}$. An equivalent condition (HW) is that every element in $Y \oplus Z$ can be written uniquely as $y + z$, for $y \in Y$ and $z \in Z$. This lead to the formula

$$\dim(X + Y) = \dim Y + \dim Z + \dim(Y \cap Z),$$

with the caveat that (i) we haven't yet defined dimension, and (ii) we don't yet know how to prove this. We illustrated this with two examples: $Y = xy$-plane and $Z = yz$-plane, and also with $Z = z$-axis.

We defined linear combinations and spanning sets, and discussed the proof of

$$\mathrm{Span}(S) = \bigcap_{S \subseteq Y_\alpha \leq X} Y_\alpha.$$

The RHS of this is the "smallest subspace that contains S." It shows how a subspace can be defined "from the bottom, up" (as linear combinations), or "from the top, down." (as intersections)

Tues. July 9. We started with some examples of mathematical objects that can be defined "from the bottom, up" or "from the top, down." The convex hull is one such example. This equality fails for the defintion of an ideal, if the ring does not have unity (example, $I = (2)$ in $R = 2\mathbb{Z}$. In some cases, like the "smallest normal subgroup containing a set," there isn't a bottom-up definition.

Next, we talked about spanning sets, linear independence, and bases. We proved some basic facts, like how all bases of the same size, called its *dimension*, and gave examples of how this fails in groups (e.g., S_n). We showed how one can always extend an independent set to a basis, and how one can always remove vectors from a dependent set until it is independent.

Wed. July 10. We talked about complements and direct sums, and compared direct sums to direct products. These are the same when things are finite, but they differ in the infinite-dimensional cases. As an example, we compared the space $R \times R \times \cdots$ of infinite sequences ($\cong$ power series) to teh space $R \otimes R \otimes \cdots$ of finite sums ($\cong$ polynomials). Just for fun, we discussed ℓ^p spaces from functional analysis.

We defined what it meant for two vectors to be equivalent modulo Y. Our two running examples were $Y = xy$-plane and $Z = yz$-plane, and also with $Z = z$-axis. We discussed what it meant to be well-defined, and left the proof that addition and scalar multiplication in X/Y being well-defined for the HW.

Thurs. July 11. We spent this day going over some of the main ideas in group theory, with a focus on visualizations. Specifically, we discussed cosets, and how a quotient group G/H is defined iff and only if H is normal. Specifically, what it mean to be well-defined, and how and why can this fail? Then, we moved onto the isomorphism theorems, because they have analogues in linear algebra. Finally, we discussed how every group is a quotient of a free group, and how the construction is done by starting with a free group on the same number of generators, and taking the quotient of the smallest normal subgroup that contain all relators.

Fri. July 12. We started by proving that $\dim Y + \dim X/Y = \dim X$. As a corollary, by applying the first isomorphism theorem, $X/\operatorname{Ker} f \cong \operatorname{Im} f$, gives the rank-nullity theorem: $\dim X = \operatorname{rank} f + \operatorname{nullity} f$. We also proved that

$$\dim U + \dim V - \dim(U \cap V) = \dim X.$$

The first step was to consider the case when $U \cap V = \{0\}$, i.e., $X = U \oplus V$, which we've already done. More generally, if $W = U \cap V$, when we can take the quotient of everything by W, and get that $\overline{X} = \overline{U} + \overline{V}$. This, with the previous results proven today, gives the desired result.

In the second half of class, we discussed the dual space X', which is the space of all linear scalar functions $\ell \colon X \to K$.We showed how every co-vector can be written as $\ell(x) = a_1c_1 + \cdots a_nc_n$, where $x = a_1x_1 + \cdots a_nx_n$. In light of this, one can think of elements in X (vectors) as column vectors, and elements in X' (co-vectors) as row vectors. It is convenient to denote $:= (\ell, x) = \ell(x)$. Given a basis $x_1, \dots, x_n$, the *dual* basis in X' is $\ell_1, \dots, \ell_n$, where $\ell_i(x_j) = \delta_{ij}$. The *Riesz representation theorem* from functional analysis "basically" says that in a Hilbert space, every linear functional is the the inner product on some fixed vector (there are other technical conditions). For a non-example, note that in $\ell_1(\mathbb{R})$, the "dot product with $(1, 1, 1, \dots)$" is a linear functional, despite $(1, 1, 1, \dots) \notin \ell_1(\mathbb{R})$.

We finished with the double dual X'', whose elements can be thought of as "evaluation maps." The space X'' can be canonically identified with X, whereas there is no such basis-free identification with X'.

WEEK 4

Mon. July 15. Mostly for run, we showed two applications of the rank-nullity theorem to polynomials (interpolation and average values). We defined the *annihilator* $Y^\perp$ of a subspace $Y \le X$, and then discussed the *transpose* of a linear map $T \colon X \to U$. This is a map $T \colon U' \to X'$ such that $(\ell, Tx) = (T'\ell, x)$ for all $x \in X$ and $u \in U$.

Tues. July 16. We showed how to define the matrix of a linear map, given an "input basis" $x_1, \dots, x_n$ and an "output basis" $u_1, \dots, u_m$. We saw several examples, such as the projection onto the line $y = x$ with two different bases: $\{e_1, e_2\}$, and the rotation of this by 45°. We pondered an exercise left for the homework about how, given an arbitrary linear map $T \colon X \to U$, to pick an input basis and output basis so the matrix is $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$. We showed that the matrix of the transpose of a linear map was the transpose of the matrix.

Wed. July 17. We started the day talking about change of basis matrices. If A is the matrix with respect to the basis $e_1, \dots, e_n$, then $P^{-1}AP$ is the matrix with respect to the basis $x_1, \dots, x_n$, where $P = [x_1 \cdots x_n]$ is the "change of basis matrix":

$$\begin{array}{lccc}
\textit{new basis} & X & \xrightarrow{\;P^{-1}AP\;} & X \\
 & {\scriptstyle P}\downarrow & & \downarrow{\scriptstyle P} \\
\textit{old basis} & X & \xrightarrow{\quad A \quad} & X
\end{array}$$

We spent the rest of the day talking about multilinear forms (functions). A function $f \colon X^k \to K$ is k-linear if fixing any $k-1$ coordinates leaves a linear function. 1-linear is just "linear" $\ell \colon X \to K$, and 2-linear is bilinear. The space of k-linear functions is n^k-dimensional.

A basis of the 1-linear functionals are the dual vectors $\ell_1, \dots, \ell_n$, where $\ell_j(x_i) = \delta_{ij}$. A basis for the the 2-linear functionals are the functions f_{ij}, $1 \le i, j \le n$, where

$$f_{ij}(x_k, x_\ell) = \begin{cases} 1 & i = k,\ j = \ell \\ 0 & \text{else} \end{cases}$$

This means that every bilinear function can be written uniquely as

$$f(x, y) = \sum_{1 \le i,j \le n} c_{ij} f_{ij}(x, y).$$

A bilinear function is *symmetric* if $f(x, y) = f(y, x)$, and *skew-symmetric* if $f(x, y) = -f(x, y)$. Note that the subspace of symmetric bilinear functions has dimension $n(n+1)/2$, and the skew-symmetric functions have dimension $n(n-1)/2$. Since these sum to n^2, every bilinear function can be written uniquely as a sum of a symmetric and skew-symmetric function.

This generalizes from bilinear to trilinear in the predictable way, but it's notationally messy. For example, the space of trilinear functions has dimension n^3, with

$$f(x, y, z) = \sum_{1 \le i,j,k \le n} c_{ijk} f_{ijk}(x, y, z).$$

A k-linear form $f\colon X^k \to K$ is:

- *symmetric* if $f(x_1, \dots, x_k) = \pi \cdot f(x_1, \dots, x_k) := f(x_{\pi^{-1}(1)}, \dots, x_{\pi^{-1}(k)})$ for all $\pi \in S_k$,
- *skew-symmetric* if $\tau \cdot f(x_1, \dots, x_k) = -f(x_1, \dots, x_k)$ for all transpositions $\tau = (ij) \in S_k$,
- *alternating* if $f(x_1, \dots, x_k) = 0$ whenever $x_i = x_j$.

Alternating implies skew-symmetric; to see this, just take

$$0 = f(x + y, x + y) = f(x, x) + f(x, y) + f(y, x) + f(y, y)$$

The converse is false, but only if $\mathbb{Z} = \mathbb{F}_2$.

Thurs. July 18. We further explored determinants. Along the way, we took a detour to the symmetric and alternating groups, and used some visuals involving Cayley graphs and the permutahedron.

By expanding the determinant as

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \cdots,$$

we can derive formulas such as

$$\det A = \sum_{\pi \in S_n} (\operatorname{sign} \pi) a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)} = \sum_{\pi \in S_n} (\operatorname{sign} \pi) a_{\pi(1),1} a_{\pi(2),2} \cdots a_{\pi(n),n} = \det A^T.$$

We saw why the space of alternating n-linear functions was one-dimensional. We also defined the universal property of determinant, which allowed us to quickly establish properties such as $\det AB = (\det A)(\det B)$.

Finally, we defined the *trace* of a matrix to be the sum of its diagonal entries, and discussed why $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. The trace of A also happens to be the sum of its eigenvalues, but we'll see that next.

Fri. July 19. We revisited the concept of eigenvalues and eigenvectors, and how to solve $Av = \lambda v$. We did a 2×2 matrix as an example, using $A = \begin{bmatrix} 3 & 2 \\ 1 & 4 \end{bmatrix}$. We diagonalized this matrix, by writing $D = P^{-1}AP$, and interpreted that geometrically. By writing $A = PDP^{-1}$, it is easy to quickly compute large powers of A, because $A^k = PD^kP^{-1}$.

We paused for a few examples and applications of this. One is from ODEs, where a 2nd order linear ODE can be written as a 2×2 system. As an example, given $x'' + 2cx' + \omega^2 x = \sin t$, define $v = x'$, and then this becomes

$$\begin{bmatrix} x' \\ v' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2c \end{bmatrix} + \begin{bmatrix} 0 \\ \sin t \end{bmatrix}.$$

We the saw how $x' = Ax$ has solution $x(t) = e^{\lambda t} v$, and showed how this affect the dynamics in the phase space, which is the plot of $x_2(t)$ vs. $x_1(t)$.

After that, we saw applications of population matrices, where the long term behavior is governed by $\lim\limits_{k \to \infty} A^k$. Markov chains are another instance, and we did a simple 2×2 example.

We finished by proving two basic facts about eigenvectors: (i) Every $n \times n$ matrix A has one, and (ii) eigenvectors from distinct eigenvalues are linearly independent.

WEEK 5

Mon. July 22. By computing $p_A(t) = \det(tI - A)$ as a sum of $n!$ scalars of permutation matrices, it is apparent that only one term contains powers of t above $n - 2$. This helps us see that $p_A(t) = t^n - (\operatorname{tr} A)^{n-1} + \cdots + (-1)^n \det(A)$, and that $\operatorname{tr} A$ and $\det A$ are the sum and product of the eigenvalues, respectively.

We remarked that if $Av = \lambda v$, then $A^k v = \lambda^k v$ for all $k \in \mathbb{N}$, and more generally, $q(A)v = q(\lambda)v$ for any polynomial $q(t)$. Next, we moved onto the concept of diagonalization. The following are equivalent for $A \colon X \to X$:

(1) $A = PDP^{-1}$ for some diagonal matrix D.
(2) X has a basis of eigenvectors for A.

The diagonal entries are the eigenvalues, and the columns of P are the eigenvectors.

$$\begin{array}{lccc} \textit{eigenvector basis} & X & \xrightarrow{D = P^{-1}AP} & X \\ & \Big\downarrow P & & \Big\downarrow P \\ \textit{standard basis} & X & \xrightarrow{A = PDP^{-1}} & X \end{array}$$

We discussed what happens if A isn't diagonalizable, which only occurs if it has a repeated eigenvalue but not enough eigenvectors. In this case, the "Spectral Theorem" guarantees that X has as basis of so-called *generalized eigenvectors*. These are vectors in the nullspace of $(A - \lambda I)^k$ for some $k \in \mathbb{N}$. Note that "genuine eigenvectors" occur when $k = 1$. As an example, we considered three possibilities for a 3×3 matrix with characteristic polynomial $p(t) = (t - 2)^2$:

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

In the last example, we have a sequence $v_3 \xmapsto{A-2I} v_2 \xmapsto{A-2I} v_1 \xmapsto{A-2I} 0$ of generalized eigenvectors. More generally, such "generalized eigenvector chains" correspond to Jordan blocks.

Tues. July 23. We computed a generalized eigenvector basis of $A = \begin{bmatrix} -1 & -1 \\ 1 & -3 \end{bmatrix}$. We defined the *minimal polynomial* $m_A(t)$, which divides $p_A(t)$, as the smallest monic polynomial for which $m_A(A) = 0$. We explored two examples from the HW, and found (most of) the generalized eigenvectors of the following matrices:

$$A = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 2 & 1 & 2 & 1 \\ 0 & 0 & -1 & 0 \\ 4 & 0 & -6 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 2 & 1 & 0 & -4 \\ 1 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The first of these has characteristic polynomial $p_A(t) = (t-1)^2(t+1)^2$, and we wrote down all four possible Jordan canonical forms. The next one has $p_B(t) = (t-1)^4$, two eigenvectors, and $m_B(t) = (t-1)^2$.

For another example, consider an 11×11 matrix A with characteristic polynomial $p_A(t) = (t-\lambda)^{11}$ and minimal polynomial $m_A(t) = (t-\lambda)$. Once possibility for the generalized eigenvectors of such an example is the following:

$$\begin{array}{ccccccccccc} v_5 & \xmapsto{A-\lambda I} & v_4 & \xmapsto{A-\lambda I} & v_3 & \xmapsto{A-\lambda I} & v_2 & \xmapsto{A-\lambda I} & v_1 & \xmapsto{A-\lambda I} & 0 \\ & & & & w_3 & \xmapsto{A-\lambda I} & w_2 & \xmapsto{A-\lambda I} & w_1 & \xmapsto{A-\lambda I} & 0 \\ & & & & & & x_2 & \xmapsto{A-\lambda I} & x_1 & \xmapsto{A-\lambda I} & 0 \\ & & & & & & & & y_1 & \xmapsto{A-\lambda I} & 0 \end{array}$$

We saw how generalized eigenvectors arise in differential equations. For example, $y''+4y'+4y = 0$ leads to a repeated root $(r+2)^2 = 0$, and thus only one solution, $y_1(t) = e^{-2t}$. Another one ends up being a "generalized eigenfunction" $y_2(t) = te^{-2t}$. A similar situation arises when we have a system $x' = Ax$ and A has a repeated eigenvector. The first solution is $x_1(t) = e^{\lambda t}v_1$, and the second solution is $x_2(t) = te^{\lambda t}v_1 + te^{\lambda t}v_2$, where $(A-\lambda I)v_1 = 0$ and $(A-\lambda I)v_2 = v_1$.

Wed. July 24. We defined what it means for a subspace $Y \leq X$ to be A-*invariant*, for some linear map $A\colon X \to X$. Suppose $X = Y \oplus Z$ where Y and Z are A-invariant. If a basis of X is construction from a basis of Y and Z, then the matrix with respect to that basis will be block-diagonal. We related this back to the example of generalized eigenvectors of an 11×11 matrix from the previous day.

We stated the *Spectral Theorem*, which says that if $A\colon X \to X$ is a linear map, then X is a direct sum of generalized eigenspaces of A. That is, $X = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$ where $E_{\lambda_j} = N_{(A-\lambda_j I)^n}$. We proved that if A and B are diagonalizable and commute, then they have a common basis of eigenvectors. We also discussed what this really means.

Thurs. July 25. We spent the first half of the class discussing Jordan canonical form, and going over the relevant HW problems. In particular, we considered matrices that satisfied equations such as $A^n = 0$, $A^2 = A$, and $A^k = A$, what we could say about their minimal polynomials, and other related questions.

Then, we moved onto inner products. We started with a real n-dimensional vector space, and showed how the dot product endows the space with a norm, where $||x||^2 = (x,x)$. We derived $\cos\theta = \frac{(x,y)}{||x||\cdot||y||}$, and the law of cosines:

$$c^2 = a^2 + b^2 - 2ab\cos\theta, \qquad a = ||x||, \quad b = ||y||, \quad c = ||x-y||.$$

Naturally, this works for any inner product, which is a bilinear function $\langle\ ,\ \rangle\colon X \times X \to K$ that is symmetric and positive-definite, and $\cos(\theta)$ can be defined similarly. Then we proved several basic properties about the norm

- Cauchy-Schwarz: $|(x,y)| \leq ||x|| \cdot ||y||$
- Triangle inequality: $||x+y|| \leq ||x|| + ||y||$.

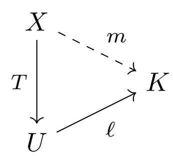
Finally, the norm can be characterized as

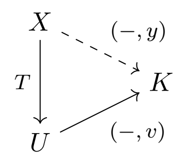
$$||x|| = \max\{(x,y) : ||y|| = 1\}.$$

Fri. July 26. Recall that by the Riesz representation theorem, every linear scalar function can be expressed as the projection onto some fixed vector $y \in X$, i.e., $(-,y)$. This means that if X is endowed with an inner product, then we can canonically identify X with its dual space X' in one of two ways:

$$L\colon X \longrightarrow X', \qquad y \mapsto (-,y), \qquad \text{or} \qquad R\colon X \longrightarrow X', \qquad y \mapsto (y,-).$$

The *adjoint* of a linear map $A\colon X \to U$, is a map $A^*\colon U \to X$ such that $(Ax,u) = (x, A^*u)$ for all $x \in X$ and $u \in U$. In an $\mathbb{R}$-vector space, A^* is just the transpose, and in a $\mathbb{C}$-vector space, it is the conjugate transpose. This is related to the transpose, but it goes between the actual vectors instead of the scalar functions.





That is, if $A'\colon (-,v) \mapsto (-,y)$, then $A'\colon v \mapsto y$.

We reviewed complex numbers, and defined the complex inner product as

$$\langle z, w\rangle = w^H z = \overline{w}^T z = \begin{bmatrix}\overline{w_1} & \overline{w_2} & \cdots & \overline{w_n}\end{bmatrix}\begin{bmatrix}z_1\\ z_2 \\ \vdots \\ z_n\end{bmatrix}.$$

We defined a complex inner product as a positive-definition *sesquilinear* form $\langle\ ,\ \rangle\colon X \times X \to K$. Some proofs for $\mathbb{R}$-vector spaces can be carried over to $\mathbb{C}$-vector spaces by replacing "T" with "H", throughout.

A linear map $A\colon X \to X$ is *self-adjoint* if $A = A^*$, which means that $(Ax,u) = (x,Au)$ for all $x, u \in X$. We proved several basic properties about self-adjoint linear maps:

(1) They have only real eigenvalues.
(2) They have a full set of eigenvectors
(3) Eigenvector from distinct eigenvalues are orthogonal (and so they have an orthonormal set of eigenvectors).

Other linear maps that have this last property include those that are (i) anti-self-adjoint ($A^* = -A$), and (ii) orthogonal ($Q^* = Q^{-1}$). More generally, we showed that the class of linear maps that have this property are called *normal*, which is characterized by commuting with its adjoint, i.e., $NN^* = N^*N$. This was proven by writing such a map into a sum of a self-adjoint and anti-self-adjoint map: $N = S + A$.

WEEK 6

Mon. July 29. A linear map $P\colon X \to X$ is a *projection* if $P^2 = P$, and it is straightforward to show that $X = R_P \oplus N_P$. We showed that these two spaces are orthogonal complements iff P is self-adjoint, in which case it is called an *orthogonal projection*.

A big idea from last week is the if $H\colon X \to X$ is self-adjoint, then it is diagonalizable by an orthogonal matrix, which means that if you "tilt your head the right way," the linear map behave like

$$\begin{bmatrix}2 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 4\end{bmatrix},$$

which is easy to visualize.

If H is self-adjoint, then $X = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_k}$ into orthogonal eigenspaces. If $P_i \colon X \to X$ is the projection onto the i^{th} eigenspace, then

$$I = P_1 + \cdots + P_k, \qquad \text{and} \qquad H = \lambda_1 P_1 + \cdots + \lambda_k P_k.$$

For any function $f(t)$ defined on the eigenvalues, it is well-founded to define $f(H) := f(\lambda_1)P_1 + \cdots + f(\lambda_k)P_k$.

The equation $5x_1^2 - 6x_1x_2 + 5x_2^2 = 9$ is an ellipse, and can be written as $x^T Ax = 9$ for a symmetric matrix A. If we diagonalize it by $A = PDP^T$, and let $z = P^T x$, then $z^T Dz = 9$ which is $8z_1^2 + 2z_2^2 = 9$, which is easy to plot. The change of matrix P is a 45° rotation matrix, and so the original ellipse is a rotation of the one in the z-coordinates.

In general, if A is self-adjoint, then the equation $f(x) = x^T Ax$ is a *quadratic form*. The *Rayleigh quotient* is the function $R_H(x) = (x, Hx)/(x, x)$. The critical points are the eigenvectors, and $R_H(v) = \lambda$, for $Hv = \lambda v$. If the eigenvalues are $\lambda_1 \leq \cdots \leq \lambda_n$, then

$$\lambda_k = \min_{\dim S = k} \left\{ \max_{x \in S \setminus 0} R_H(x) \right\}.$$

Tues. July 30. We spent the day talking about the singular value decomposition (SVD) of a linear map $A \colon X_1 \to X_2$. In particular, this is $A = UDV^*$, and it can be derived by computing

$$A^*A = (UDV^*)(VD^*U^*), \qquad AA^* = (VD^*U^*)(UDV^*).$$

That is, the columns U and V are the orthonormal eigenvectors of A^*A and AA^*, respectively, and D is the "diagonal" matrix with entries $\sqrt{\sigma_1}, \dots, \sqrt{\sigma_r}$, the square roots of the diagonal entries of D^*D. We did an example of this, for the matrix $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$, and used our knowledge of the "four fundamental subspaces" to do it without actually computing A^*A and AA^*, though we did do that to check our work.

Finally, we finished with a discussion of 2-sided, left-, right-, and pseudo-inverses, for a linear map $A \colon K^n \to K^m$. With each one, we drew a "cartoon" of the "four fundamental subspaces" to understand it better.

(1) **2-sided inverse** ($r = n = m$). $Ax = b$ has a unique solution. A^*A and AA^* are both invertible.

(2) **left inverse** ($r = n < m$). Full column rank. $N_A = N_{A^*A} = \{0\}$, so $Ax = b$ has 0 or 1 solutions. A^*A is invertible, and so

$$\underbrace{(A^*A)^{-1}A^*}_{A^{-1}_{\text{left}}} A = I, \qquad AA^{-1}_{\text{left}} = A(A^*A)^{-1}A^* = \text{Proj}_{C(A)}.$$

(3) **right inverse** ($r = m < n$). Full row rank. $N_A = N_{AA^*} = \{0\}$, so $Ax = b$ has infinitely many solutions. AA^* is invertible, and so

$$A \underbrace{A^*(AA^*)^{-1}}_{A^{-1}_{\text{right}}} = I, \qquad A^{-1}_{\text{right}} A = A^*(AA^*)^{-1}A = \text{Proj}_{C(A^*)}.$$

(4) **pseudo-inverse** (any linear map). A linear map $A^\dagger \colon K^m \to K^n$ so that

$$A^\dagger A = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n \times n}, \qquad AA^\dagger = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m \times m}.$$

Wed. July 31. We started the class with an overview of *principal component analysis*, which is related to both SVD, and the Rayleigh quotient. In particular, the algorithm of maximizing the Rayleigh quotient, and then taking the orthogonal complement, and repeating, is analogous to iteratively finding the principal components.

We mentioned the concept of the *norm* of a linear map $A \colon X \to U$, and there are several ways to define this. One is called the *induced norm*, which is

$$|A| := \sup \big\{ ||Ax|| : ||x|| = 1 \big\} = \sup \big\{ \langle Ax, v \rangle : ||x|| = ||v|| = 1 \big\},$$

and does not arise from an inner product. Another example is the *Frobenius norm*, which is $||A|| = \sqrt{\operatorname{tr} A^*A}$, which arises from the inner product $\langle A, B\rangle = \operatorname{tr}(B^*A)$.

We finished the class with applications from PDEs and Fourier series, to illustrate self-adjoint maps. Specifically, we solved the following BVP for the heat equation: $u_t = c^2 u_{xx}$, $u(0,t) = u(\pi,t) = 0$, $u(x,0) = h(x)$. By separating variables, we got two ODEs that were "eigenvalue equations." This BVP is a special case of a *Sturm-Liouville problem*, which has the form $Lv = \lambda v$ for a self-adjoint differential operator L. We gave several other examples, by changing the boundary conditions, initial condition, and the PDE. The general solution is an infinite sum of eigenfunctions. In linear algebra terms, this is just linearity. In physics terms, it is superposition.