Contents

Lecture 1

Vector spaces and bases

1.1 Abstract vector spaces (review)

A vector space V is an algebraic structure motivated by the idea of parameterizing all possible displacements of a particular object in space. Here is an incomplete list of intuitive notions about displacements that will lend insight to the formal definition (which is notoriously opaque).

1. There is a "neutral displacement" which consists in not moving at all.
2. Displacements can be composed. If "one mile north" and "two miles east" are two examples of displacements, then "one mile north and two miles east" counts too.
3. In the above example, order of composition doesn't matter. You can first go north, then east; or you can first go east, then north. Either path leads to the same overall displacement.
4. Given a displacement, you should be able to state *how much* of it to refer to. For example, the displacement from the top of my head to the bottom of my torso is approximately three times that from the top of my head to my chin.
5. There is no obviously meaningful way to "multiply" two displacements to result in another one. Where would I end up if I were displaced by the product of "New York to Los Angeles" and "Chicago to Seattle"?

Informal Definition 1. *A **field** is a collection of numbers with additive identity 0 and multiplicative identity 1 that you can add, subtract, multiply, and divide (except by 0) in the usual way.*

For our purposes, the field is almost always $\mathbb{C}$, the complex numbers.[1]

Definition 1 (Vector space)**.** *A **vector space** over a field of scalars is a set on which addition and scalar multiplication are defined satisfying the following:*

1. *For all vectors x and y, $x + y = y + x$. (commutativity of addition)*
2. *For all vectors x, y, and z, $(x + y) + z = x + (y + z)$. (associativity of addition)*
3. *There is a neutral vector 0 such that $x + 0 = x$. (additive identity)*
4. *For each vector x there is a vector y such that $x + y = 0$. (additive inverse, "subtraction")*
5. *For each vector x, $1x = x$. (multiplicative identity)*

[1] A fuller definition and theory of fields is not a part of this class, but if you're curious, there are lots of fields out there! Other examples are $\mathbb{R}$, the real numbers; $\mathbb{Q}$, the rational numbers; $\mathbb{Q}[j]$, the rational numbers with $\sqrt{-1}$ thrown in; and $\mathbb{Z}_p$, the integers modulo a prime p.

6. *For each pair of scalars α and β and each vector x, $(\alpha\beta)x = \alpha(\beta x)$. (associativity of scalar multiplication)*

7. *For each scalar α and pair of vectors x and y, $\alpha(x + y) = \alpha x + \alpha y$. (left distributivity of scalar multiplication)*

8. *For each pair of scalars α and β and vector x, $(\alpha + \beta)x = \alpha x + \beta x$. (right distributivity of scalar multiplication)*

There are a *lot* of definitions and facts about vector spaces from 16A or 54 that are necessary to go further, e.g. that a vector space has exactly one additive identity element. We rely on a lot of commonsense technical results about existence and uniqueness, but we will not interrogate them. The following are some of the most important ones.

Definition 2 (Linear independence)**.** *The vectors $v_1, v_2, \ldots, v_n$ are **linearly independent** if for any scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$;*[2] *$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = 0$ implies $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$.*

Definition 3 (Basis)**.** *A **basis** of a vector space V is a maximal list of linearly independent vectors: if you added even one more vector from V, it would not longer be linearly independent.*

Definition 4 (Dimension)**.** *The dimension of a vector space V is the number of vectors in any basis of V.*

Theorem 1 (Criteria for vectors to form a basis)**.** *A list of vectors $v_1, v_2, \ldots, v_m$ is a basis for V if any two of the following three conditions hold:*

1. *$m = \dim V$.*

2. *$v_1, v_2, \ldots, v_m$ generate V.*

3. *$v_1, v_2, \ldots, v_m$ are linearly independent.*

Very important example: k^n as a vector space over k

Given a field k and an integer $k > 0$, we can form the vector space k^n of n-tuples of scalars from k. Vectors are expressions of the form $v = (v_1, v_2, \ldots, v_n)$[3] and are sometimes written as "column vectors":

$$(v_1, v_2, \ldots, v_n) = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

Addition is defined by the rule $(v_1, v_2, \ldots, v_n) + (v_1', v_2', \ldots, v_n') = (v_1 + v_1', v_2 + v_2', \ldots, v_n + v_n')$. This can be communicated concisely using a variable subscript: $(v + v')_i = v_i + v_i'$.[4]

Scalar multiplication is defined by the rule $(\alpha v)_i = \alpha\,(v_i)$.

It can be verified that k^n with addition and scalar multiplication defined this way is a vector space over k.

Theorem 2. *The dimension of k^n is n.*

[2] A note on notation: it's customary to write the shorthand "1, 2, . . . , n" to signify "all of the numbers from 1 to n," even if $n \leq 2$. If n is 0, it's the empty list; if n is 1 or 2 then the enumeration is understood to end at n.

[3] Setting off the components with commas distinguishes this expression from a "row vector."

[4] This equation would be read aloud as "component I of V plus V-prime equals component I of V plus component I of V-prime."

Unimportant, but interesting example: $\mathbb{C}^n$ and its evil twin $\overline{\mathbb{C}^n}$

We will spend most of our time in the vector space $\mathbb{C}^n$, defined in the usual way above. In this section we will define a *different* vector space structure (addition and scalar multiplication) on the set $\mathbb{C}^n$. The vector space we are about to make is evil and strange! (It is utterly useless except for scaring your friends.)

Let's call it "$\overline{\mathbb{C}^n}$." The underlying set is n-tuples of complex numbers, and addition is defined componentwise as usual. Define scalar multiplication by $(\alpha v)_i = \overline{\alpha}\,(v_i)$. That is, "multiplying" by a scalar actually scales the vector's components by the conjugate of the scalar.

(Is there an isomorphism of vector spaces between normal $\mathbb{C}^n$ and $\overline{\mathbb{C}^n}$?[5])

1.2 Coordinates on vector spaces, and abuse of notation

There is a difference between *concept* and *representation* of vector spaces.[6] Vectors and linear transformations [7] have a life of their own and do not mind what we think of them. However, in order to work with them hands-on, we need to represent them in a system of coordinates.

A choice of coordinates is like a digital imaging system. If I photograph a cat, the photo is not the cat; it is a representation of the cat. I can choose to make the photo brighter or darker. I can mirror it or rotate it, and even though the image might be very different, the cat is the same; all I have changed is the means of representation.

We represent vectors and linear maps by choosing bases.

Definition 5 (Coordinates of a vector)**.** *The coordinates of a vector v in the vector space V with respect to the basis $\{v_1, v_2, \ldots, v_n\}$ are the unique scalars $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ such that $v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n$.*

Example: two bases for $\mathbb{C}^3$

We will represent the vector $(1, -\sqrt{3}, \sqrt{3})$ in two different bases.

$$\begin{pmatrix} 1 \\ -\sqrt{3} \\ \sqrt{3} \end{pmatrix} = 1\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \sqrt{3}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \sqrt{3}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{1.1}$$

$$= 1\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + j\begin{pmatrix} 1 \\ -1/2 + \sqrt{3}/2j \\ -1/2 - \sqrt{3}/2j \end{pmatrix} - j\begin{pmatrix} 1 \\ -1/2 - \sqrt{3}/2j \\ -1/2 + \sqrt{3}/2j \end{pmatrix} \tag{1.2}$$

The coordinates of $(1, -\sqrt{3}, \sqrt{3})$ relative to the first basis are $(1, -\sqrt{3}, \sqrt{3})$. This basis is called the standard basis, and its vectors are usually named e_1, e_2, e_3.

The coordinates of $(1, -\sqrt{3}, \sqrt{3})$ relative to the second basis are $(1, j, -j)$.

Definition 6 (Coordinates of a matrix)**.** *The representation of the linear function $f : V \to U$ relative to the basis $\{v_1, v_2, \ldots, v_n\}$ for V and the basis $\{u_1, u_2, \ldots, u_m\}$ for U are the $m \times n$ matrix A. Column j of A is the coordinate vector of $f(v_j)$ relative to $\{u_1, u_2, \ldots, u_m\}$.*

Theorem 3. *Coordinate representations of vector spaces are faithful: matrix-vector multiplication represents function application, and matrix-matrix multiplication represents function composition.*

[5] Yes, because all equidimensional vector spaces over the same field are isomorphic. If we use the standard bases for $\mathbb{C}^n$ and $\overline{\mathbb{C}^n}$, the matrix of an isomorphism is I. In my opinion the proof is somewhat confusing.

[6] For this wording I am indebted to Hegel's theory of knowledge as a synthesis of *Begriff* (tr. concept/idea) and *Vorstellung* (tr. representation).

[7] A linear transformation from a vector space to itself is called an operator.

In applied mathematical subjects (such this class), we will often use $\mathbb{C}^n$ with the standard basis. As a result, we can, and sometimes will delude ourselves with notation that appears to identify the representation of a vector with the vector itself. Likewise, sometimes, but not always, we will identify linear functions with their matrix representations.

1.3 Diagonalization

Throughout this section, let $T : V \to V$ be a linear map from vector space V to itself.[8]

Definition 7. *(Eigenvalues and eigenvectors) If a scalar λ and a nonzero vector v satisfy the relationship $T(v) = \lambda v$, then v is called an **eigenvector** of T for **eigenvalue** λ.*

Note that the eigenvector-eigenvalue condition is equivalent to saying that the linear map $(T - \lambda I) : V \to V$ has a nontrivial nullspace, which can be called the eigenspace of T for eigenvalue λ.

Theorem 4. *(Diagonalization) If*

$$\begin{aligned} T(v_1) &= \lambda_1 v_1, \\ T(v_2) &= \lambda_2 v_2, \\ \vdots &= \quad \vdots \\ T(v_n) &= \lambda_n v_n \end{aligned}$$

are eigenvalue-eigenvector pairs such that $\{v_1, v_2, \ldots, v_n\}$ is a basis of V, then the representation of T in this basis is the diagonal matrix

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

where the empty spaces are zero.

Proof. To make column i of the matrix representation of T relative to this basis, we must represent $T(v_i)$ in the basis $\{v_1, v_2, \ldots, v_n\}$. As v_i was chosen to be an eigenvector, $T(v_i) = \lambda_i v_i$. The coordinates of $\lambda_i v_i$ in this basis are λ_i in place i and 0 everywhere else. □

On determinants

For computing and discussing eigenvalues we humans[9] need to use the scalar-valued matrix function $\det : \mathbb{C}^{n \times n} \to \mathbb{C}$, characterized by the following:

1. $\det$ is linear in every column.
2. $\det A = 0$ if the columns of A are linearly dependent.
3. $\det I = 1$.

$\det$ is a polynomial function with a number of important properties and interpretations, many of which are covered in a prerequisite course. Our primary use will be in a function called the characteristic polynomial $\chi_A(s)$ that reveals in its roots the eigenvalues of A. Given a matrix A, $\chi_A(s) = \det(sI - A)$, or sometimes $\det(A - sI)$ (which works equally well).

[8] It is safe to think of V as $\mathbb{C}^n$ and T as a matrix, but you should nevertheless prefer the toolkit to the laws of vector spaces and linear maps to that of coordinate manipulation.

[9] Generally computers have more efficient approximations.

Trivial diagonalization example

Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be dilation by factor of ρ; that is, $T(v) = \rho v$. Choose any basis $\{v_1, v_2, \ldots, v_n\}$. The image of v_i under T can be constituted in this basis as ρ of v_i and 0 of every other basis vector. Therefore the matrix of T in any basis is ρI.

Easy diagonalization example

The vector space is $\mathbb{C}^2$, and the operator is left multiplication by the matrix

$$A = \begin{pmatrix} 1 & j \\ -j & 1 \end{pmatrix} \tag{1.3}$$

This matrix has characteristic polynomial

$$\chi_A(s) = s^2 - 2s = s(s-2) \tag{1.4}$$

and therefore eigenvalues 0 and 2, which we will keep in that order. The eigenvector corresponding to 0 comes from the null space of $A - 0I$. $(1, j)$ will do. The eigenvector corresponding to 2 comes from the null space of $A - 2I$. I'll take $(j, 1)$. To convert eigenbasis coordinates to standard basis coordinates, line up these eigenvectors side by side.

$$V = \begin{pmatrix} 1 & j \\ j & 1 \end{pmatrix} \tag{1.5}$$

The matrix that converts standard basis coordinates to eigenbasis coordinates is V^{-1}:

$$V^{-1} = \begin{pmatrix} 1/2 & -j/2 \\ -j/2 & 1/2 \end{pmatrix} \tag{1.6}$$

Therefore A can be factored as follows.

$$A = \begin{pmatrix} 1 & j \\ j & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1/2 & -j/2 \\ -j/2 & 1/2 \end{pmatrix} \tag{1.7}$$

Hard diagonalization example

Let T be the operator on $\mathbb{C}^3$ that circularly shifts the coordinates one place to the right. It maps e_1 to e_2 and e_2 to e_3, and wraps e_3 around to e_1; so its coordinates in the standard basis are the following matrix.

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{1.8}$$

This matrix can be diagonalized with the help of the characteristic polynomial as above, but we'll take a different, more scenic path. Parameterize an eigenvalue-eigenvector pair (λ, v) and solve the relation directly.

$$A \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} v_3 \\ v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \lambda v_1 \\ \lambda v_2 \\ \lambda v_3 \end{pmatrix} \tag{1.9}$$

This relationship says that $v_1 = \lambda v_2$, $v_2 = \lambda v_3$, and $v_3 = \lambda v_1$. Substituting,

$$\begin{aligned} v_3 &= \lambda v_1 && (1.10) \\ &= \lambda^2 v_2 && (1.11) \\ &= \lambda^3 v_3 && (1.12) \\ v_3(1 - \lambda^3) &= 0 && (1.13) \end{aligned}$$

According to the zero product property, $v_3 = 0$ or $\lambda^3 = 1$. But $v_3 = 0$ would imply $v = 0$, which is not an eigenvector. Therefore $\lambda^3 = 0$: λ is a number that you can cube and get 1 back as result. Parameterize λ in polar form as $re^{j\theta}$.

$$\left(re^{j\theta}\right)^3 = 1 \tag{1.14}$$
$$r^3 e^{3j\theta} = 1e^{j(0+2\pi n)}, \tag{1.15}$$

where n is any integer. Matching magnitude and phase, we have

$$r = 1 \text{ and} \tag{1.16}$$
$$3\theta = 0, 2\pi, 4\pi, 6\pi, \ldots \tag{1.17}$$
$$\theta = 0, \frac{2\pi}{3}, \frac{4\pi}{3} \tag{1.18}$$

Non-diagonalizable example

The following matrix A has eigenvalue 0, but $(A - 0I)$ has a null space of dimension only 1.

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$$

Therefore A does not have a basis of eigenvectors and is not diagonal in any basis.

1.4 Solving vector differential equations

Diagonalization is able to factor most square matrices $A = V\Lambda V^{-1}$, where Λ is a diagonal matrix of eigenvalues and V has eigenvectors of A in the same order. This fact makes it possible to solve vector differential equations $\dot{x} = Ax + bu$:

$$\frac{\mathrm{d}}{\mathrm{d}t} x = Ax + bu \tag{1.19}$$
$$= V\Lambda V^{-1} x + bu \tag{1.20}$$
$$V^{-1} \frac{\mathrm{d}}{\mathrm{d}t} x = \frac{\mathrm{d}}{\mathrm{d}t} \left(V^{-1} x\right) = \Lambda V^{-1} x + V^{-1} bu \tag{1.21}$$

Choose a new dependent variable $z = V^{-1}x$, and write its differential equation using the Chain Rule.

$$\frac{\mathrm{d}}{\mathrm{d}t} z = \Lambda z + \left(V^{-1} b\right) u \tag{1.22}$$

Each row is a scalar differential equation which we can solve, given initial conditions, using a formula. A solution for x can be reconstructed by $x = Vz$.

Lecture 2

Rigid geometry using inner products

2.1 Inner product spaces

Inner products generalize rigid Euclidean geometry to $\mathbb{C}^n$.

Definition 8. *An inner product on a vector space V over $\mathbb{C}$ is a function that assigns to every pair of vectors x, y a scalar $\langle x, y\rangle$ satisfying the following for all $x, y, z \in V$ and $\alpha \in \mathbb{C}$:*

1. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$. *(additive in the first argument)*
2. $\langle \alpha x, y\rangle = \alpha \langle x, y\rangle$. *(scaling in the second argument)*
3. $\overline{\langle x, y\rangle} = \langle y, x\rangle$. *(conjugate symmetry)*
4. $\langle x, x\rangle$ *is real and nonnegative if x is not the zero vector. (positive-definite)*

Definition 9. *An **inner product space** is a vector space over $\mathbb{C}$ equipped with an inner product.*

Theorem 5. *Let V be an inner product space. Then for $x, y, z \in V$ and $\alpha \in \mathbb{C}$, the following are true:*

1. $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.
2. $\langle x, cy\rangle = \overline{c} \langle x, y\rangle$.
3. $\langle x, 0\rangle = \langle 0, x\rangle = 0$.
4. $\langle x, x\rangle = 0$ *if and only if* $x = 0$.
5. *If* $\langle x, y\rangle = \langle x, z\rangle$ *for all x, then* $y = z$.

Very important example: standard inner product on $\mathbb{C}^n$

Define an inner product on $\mathbb{C}^n$ by $\langle x, y\rangle = y^* x = \sum_{i=1}^n x_i \overline{y_i}$. Let's verify that this proposed inner product satisfies the inner product axioms. It is linear (preserves addition and scaling) in the first argument:

$$\begin{aligned}\langle \alpha x + y, z\rangle &= \sum_{i=1}^n (\alpha x + y)_i \overline{z_i} = \sum_{i=1}^n (\alpha x_i + y_i) \overline{z_i} \\ &= \sum_{i=1}^n (\alpha x_i \overline{z_i} + y_i \overline{z_i}) = \alpha \sum_{i=1}^n x_i \overline{z_i} + \sum_{i=1}^n y_i \overline{z_i} = \alpha \langle x, z\rangle + \langle y, z\rangle .\end{aligned}$$

It is conjugate symmetric:

$$\begin{aligned}\overline{\langle x,y\rangle} &= \overline{\sum_{i=1}^{n} x_i\overline{y_i}} = \sum_{i=1}^{n} \overline{x_i\overline{y_i}} \\ &= \sum_{i=1}^{n} \overline{x_i}\,\overline{\overline{y_i}} = \sum_{i=1}^{n} \overline{x_i}y_i = \langle y,x\rangle\,.\end{aligned}$$

And it's positive-definite, as

$$\langle x,x\rangle = \sum_{i=1}^{n} x_i\overline{x_i}, \tag{2.1}$$

which is a sum in which each term is nonnegative real. Therefore if the whole sum equals zero, every term is zero, and x is the zero vector.

This inner product we just defined is called the *standard inner product*, and we will henceforth marry it to $\mathbb{C}^n$ unless otherwise specified. Occasionally, in notationally reckless contexts, we will write it as y^*x.

Definition 10 (Length of a vector)**.** *The **length** of a vector v, denoted by $\|v\|$, is $\sqrt{\langle v,v\rangle}$.*

The length of a vector scales with the vector. That is, $\|\alpha v\| = |\alpha|\,\|v\|$.

2.2 A "law of cosines"

In Euclidean geometry, the Law of Cosines states for a triangle ABC that $a^2+b^2-2ab\cos\theta = c^2$, where θ is the angle at corner C. We will rediscover this fact in $\mathbb{C}^n$ using inner products.

Let x be the displacement vector from one corner of a triangle to the second, and y the displacement vector from the first to the third. Then the three sides of this triangle are x, y, and $x-y$. An identity for $\|x-y\|^2$ follows.

$$\begin{aligned}\|x-y\|^2 &= \langle x-y, x-y\rangle \\ &= \langle x,x\rangle - \langle x,y\rangle - \langle y,x\rangle + \langle y,y\rangle \\ &= \|x\|^2 + \|y\|^2 - \langle x,y\rangle - \overline{\langle x,y\rangle} \\ &= \|x\|^2 + \|y\|^2 - 2\,\mathrm{Re}\Big\{\langle x,y\rangle\Big\} \\ &= \|x\|^2 + \|y\|^2 - 2\,\|x\|\,\|y\| \left[\frac{\mathrm{Re}\Big\{\langle x,y\rangle\Big\}}{\|x\|\,\|y\|}\right]. \text{ (if neither } x \text{ nor } y \text{ is zero)}\end{aligned}$$

The quotient in square brackets should be understood as $\cos\theta$, where $0 \le \theta < \pi$ is the angle between the vectors x and y.[1] If it is zero, then our identity reduces to the Pythagorean Theorem and we have a right triangle.

Definition 11 (Orthogonal and orthonormal subsets)**.** *A set of vectors $\{v_i\}_{i\in I}$ is **orthogonal** if $\langle v_i, v_j\rangle = 0$ for $i \neq j$. It is **orthonormal** if in addition $\langle v_i, v_i\rangle = 1$.*

Theorem 6 (Pythagoras)**.** *If $\{v_1, v_2, \ldots, v_n\}$ is orthogonal, then $\sum_{i=1}^{n} \|v_i\|^2 = \left\|\sum_{i=1}^{n} v_i\right\|^2$.*

Proof. Split the sum into head and tail, then use induction. ◻

[1] In statistics, this is called Pearson's correlation coefficient.

Πῡθαγόρᾱς didn't know what he was getting himself into! This simple result is richly applicable well beyond geometry class. An electrical interpretation is that when two sinusoidal power sources of different frequencies are superimposed, the resulting power is the sum. Likewise, when two musical instruments play at the same time, the power of the pressure wave is the sum of the instruments' respective powers. (This does not, however, mean that the result is twice as loud, as loudness perception is roughly logarithmic.) You may recall from a statistics class that variances of independent random variables add when the random variables are added. This is because independent random variables have a correlation (cosine) of zero, and are therefore susceptible to the Pythagorean theorem.

2.3 An inequality that bounds sums

Definition 12 (Orthogonal projection)**.** *Let $u, v \in V$, with $v \neq 0$. The* ***orthogonal projection*** *of u onto v is* $\operatorname{proj}_v u = \frac{\langle u,v\rangle}{\langle v,v\rangle} v$.

The motivating property of $\operatorname{proj}_v u$ is that $u - \operatorname{proj}_v u$ is orthogonal to v. proj_v is linear and does not change when v is multiplied by a nonzero scalar.

The following fact asserts that cosines lie between -1 and 1.

Theorem 7 (Cauchy-(Bunjakowsky)-Schwarz)**.** *For all pairs $u, v \in V$, $\left|\langle u, v\rangle\right| \leq \|u\| \, \|v\|$.*

Proof. The proof of the Cauchy-Schwarz inequality[2] is notorious for employing unpredictable magic tricks whence the result comes out of nowhere. I've tried to structure it in a more geometrically luminous, even if longer, way.

Separate u into parallel and perpendicular parts to v:

$$u = \operatorname{proj}_v u + \left(u - \operatorname{proj}_v u\right) \tag{2.2}$$

$$= \frac{\langle u, v\rangle}{\langle v, v\rangle} v + \left(u - \operatorname{proj}_v u\right) \tag{2.3}$$

Apply the Pythagorean Theorem; reduce fractions as needed.

$$\|u\|^2 = \left\| \frac{\langle u, v\rangle}{\langle v, v\rangle} \right\|^2 + \left\| \left(u - \operatorname{proj}_v u\right) \right\|^2 \tag{2.4}$$

Throw away the right summand.

$$\geq \left\| \frac{\langle u, v\rangle}{\langle v, v\rangle} \right\|^2 \tag{2.5}$$

The desired result follows from multiplying through by $\langle v, v\rangle$. ◻

The following result generalizes the fact that two sides of a triangle sum to longer than the third.

Theorem 8 (Triangle Inequality)**.** *Let $u, v \in V$. Then $\|u + v\| \leq \|u\| + \|v\|$.*

Proof. Work from the square of the left side.

$$\|u + v\|^2 = \langle u + v, u + v\rangle \tag{2.6}$$

$$= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \tag{2.7}$$

$$= \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle \tag{2.8}$$

[2] That is, the full version. There are cute proofs that just work for $\mathbb{R}^n$.

Take absolute values of both sides, using the fact that every number is less than or equal to its absolute value, or the scalar "Triangle Inequality."

$$\|u+v\|^2 \leq \|u\|^2 + \|v\|^2 + \left|\langle u, v\rangle\right| + \left|\langle v, u\rangle\right| \tag{2.9}$$

Use Cauchy-Schwarz on the inner products.

$$\|u+v\|^2 \leq \|u\|^2 + \|v\|^2 + 2\,\|u\|\,\|v\| \tag{2.10}$$

The Triangle Inequality follows after taking square roots. □

Example of some orthonormal vectors

The following vectors in are orthonormal in $\mathbb{C}^3$ with the standard inner product:

$$\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\1\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}.$$

2.4 Orthonormal bases and unitary transformations

When dealing with inner product spaces, the best basis is an orthonormal basis. Computing coordinates no longer requires solving a system of equations; it is as direct as computing inner products.

Theorem 9 (Computing coordinates in an orthonormal basis)**.** *If $\{v_1, v_2, \ldots, v_n\}$ is an orthonormal basis for V, then the coordinates of $v \in V$ relative to this basis are $(\langle v, v_1\rangle, \langle v, v_2\rangle, \ldots, \langle v, v_n\rangle)$.*

Proof. We will presume a unique decomposition and solve for a single coordinate.

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n \tag{2.11}$$

Take inner products of both sides with v_i on the right.

$$\langle v, v_i\rangle = \langle \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n, v_i\rangle \tag{2.12}$$

Use linearity in the first argument.

$$\langle v, v_i\rangle = \alpha_1 \langle v_1, v_i\rangle + \alpha_2 \langle v_2, v_i\rangle + \ldots + \alpha_n \langle v_n, v_i\rangle = \alpha_i \tag{2.13}$$

□

The following definition generalizes rigid transformations in Euclidean geometry.

Definition 13 (Unitary transformation)**.** *Let $f : U \to V$ be a linear map of inner product spaces. f is called an **unitary transformation** if it preserves length: $\|f(u)\| = \|u\|$. If $U = V$, f is called an **unitary operator**.*

The following result generalizes the "side-side-side" triangle congruence of classical synthetic geometry—if two triangles have the same side lengths, they have the same angles too.

Theorem 10 (Unitary transformation preserves inner product)**.** *Let $f : U \to V$ be an unitary transformation and let $x, y \in U$. Then $\langle f(x), f(y)\rangle = \langle x, y\rangle$.*

Proof. The plan is to realize the desired product as cross terms of a product of sums.

$$\left\|f(x) + f(y)\right\|^2 = \left\|x + y\right\|^2 \tag{2.14}$$

$$\left\|f(x)\right\|^2 + \left\langle f(x), f(y)\right\rangle + \left\langle f(y), f(x)\right\rangle + \left\|f(y)\right\|^2 = \|x\|^2 + \left\langle x, y\right\rangle + \left\langle y, x\right\rangle + \left\|y\right\|^2 \tag{2.15}$$

$$\left\langle f(x), f(y)\right\rangle + \left\langle f(y), f(x)\right\rangle = \left\langle x, y\right\rangle + \left\langle y, x\right\rangle \tag{2.16}$$

$$\mathrm{Re}\left\langle f(x), f(y)\right\rangle = \mathrm{Re}\left\langle x, y\right\rangle \tag{2.17}$$

Repeating with the identity $\left\|f(x) + jf(y)\right\|^2 = \left\|x + jy\right\|^2$ yields $\mathrm{Im}\left\langle f(x), f(y)\right\rangle = \mathrm{Im}\left\langle x, y\right\rangle$. □

This implies that the image of an orthonormal basis is orthonormal, so unitary transformations are one-to-one.

Theorem 11. *Let $f : U \to V$ be an unitary transformation, and let $\{u_1, u_2, \ldots, u_n\}$ and $\{v_1, v_2, \ldots, v_m\}$ be orthonormal bases for U and V, respectively. The matrix of f relative to these bases has orthonormal columns.*

The proof involves a lot of inner product chasing and is in my opinion more technical than illuminating.

Definition 14 (Unitary matrix)**.** *A unitary matrix is a square matrix $U \in \mathbb{C}^n$ with the property that $U^*U = I$.*

The matrix of an unitary operator in an orthonormal basis is unitary.

Theorem 12. *Let λ be an eigenvalue of a unitary operator $f : V \to V$. Then $|\lambda| = 1$.*

Proof. Let $f(v) = \lambda v$. Then

$$\langle v, v\rangle = \left\langle f(v), f(v)\right\rangle = \langle \lambda v, \lambda v\rangle \tag{2.18}$$

$$= \lambda\overline{\lambda}\,\langle v, v\rangle\,, \tag{2.19}$$

so $\lambda\overline{\lambda} = 1$. □

Lecture 3

Adjoint and self-adjoint operators and matrices

An inner product structure on a $\mathbb{C}$-vector spaces induces a "mirrored" twin for every linear transformation, called the adjoint. Linear operators equal their own adjoints have many important properties.

3.1 Adjoint of an operator or matrix

Definition 15 (Adjoint of a linear map)**.** *Let $f : U \to V$ be a linear map between two inner product spaces. The **adjoint** of f, denoted by $f^* : V \to U$, is the unique linear map such that $\langle f(u), v\rangle = \langle u, f^*(v)\rangle$ for all $u \in U$ and $v \in V$.*

Theorem 13 (Technical facts about adjoints)**.** *Let f and g be two linear operators on V.*

1. $(\alpha f + g)^* = \overline{\alpha} f^* + g^*$ *(conjugate linear)*
2. $(fg)^* = g^* f^*$ *(reverses composition)*
3. $(f^*)^*$ *(involutive)*
4. $I^* = I$ *(identity operator is its own adjoint)*

Proof. 1. Let $x, y \in V$. We need to show that $\langle (\alpha f + g)\, x, y\rangle = \langle x, (\overline{\alpha} f^* + g^*)\, y\rangle$.

$$\begin{aligned} \langle (\alpha f + g)\, x, y\rangle &= \alpha \langle fx, y\rangle + \langle gx, y\rangle && (3.1)\\ &= \alpha \langle x, f^* y\rangle + \langle x, g^* y\rangle && (3.2)\\ &= \langle x, (\overline{\alpha} f^* + g^*)\, y\rangle && (3.3)\\ &= \langle x, \overline{\alpha} f^*\rangle + \langle x, g^* y\rangle && (3.4)\\ & && (3.5) \end{aligned}$$

2. Same setup as before. $\langle fgx, y\rangle = \langle gx, f^* y\rangle = \langle x, g^* f^* y\rangle$.
3. Same setup as before. $\langle f^* x, y\rangle = \overline{\langle y, f^* x\rangle} = \overline{\langle fy, x\rangle} = \langle x, fy\rangle$.
4. Same setup as before. $\langle Ix, y\rangle = \langle x, Iy\rangle$.

□

Definition 16 (Adjoint of a matrix)**.** *Let $A \in \mathbb{C}^{m\times n}$. The **adjoint** or **conjugate transpose** of A is the matrix A^* such that $A^*_{ij} = \overline{A_{ji}}$.*

Theorem 14. *Over an inner product space of finite dimension, adjoints exist. In $\mathbb{C}^n$ with the standard inner product, an adjoint matrix is the matrix of the adjoint of the linear operator it represents.*

3.2 Self-adjoint operators

In this section we'll prove a condition for matrices to be orthogonally (-normally) diagonalizable.

Definition 17 (Self-adjoint). *A linear operator or matrix is called **self-adjoint** or **Hermitian** if it is equal to its own adjoint.*

Lemma 1 (Real eigenvalues). *Let $f : V \to V$ be a Hermitian operator and λ an eigenvalue of f. Then λ is real.*

Proof. Let v be a unit eigenvector of f satisfying $f(v) = \lambda v$. Then

$$\begin{aligned} \lambda &= \langle fv, v \rangle &(3.6)\\ &= \langle v, fv \rangle &(3.7)\\ &= \overline{\langle fv, v \rangle} &(3.8)\\ &= \overline{\lambda}. &(3.9) \end{aligned}$$

□

A large class of complex matrices with complex eigenvectors, but real eigenvalues. Wow! Wow!

Definition 18 (Restriction). *Let W be a subspace of V and f a linear operator on V. If $f(W) \subseteq W$, then W is called f**-invariant** or f**-stable**. The linear operator $f\big|_W : W \to W$ defined by $f\big|_W(w) = f(w)$ is called the **restriction** of f to W.*

Theorem 15 (Spectral[1] theorem). *A linear operator $f : V \to V$ is self-adjoint if and only if it is diagonal and real in an orthonormal basis of V.*

Proof. First we show that (diagonal and real in an orthonormal basis) $\implies$ (self-adjoint). Suppose that the matrix of f is diagonal and real in an orthonormal basis. A diagonal real matrix is equal to its conjugate transpose. Therefore, as orthonormal bases faithfully represent inner product spaces and maps between them, f is self-adjoint.

Next we show that (self-adjoint) $\implies$ (diagonal and real in an orthonormal basis). We will use induction on n, the dimension of V. If $n = 1$ then f is already diagonal in any basis.

Next we need to show that if the Spectral Theorem holds on vector spaces of dimension $n - 1$, then it holds on vector spaces of dimension n.

By the Fundamental Theorem of Algebra, f has an eigenvalue λ. Because f is self-adjoint, λ is real. Let v be an eigenvector such that $f(v) = \lambda v$. Both $\operatorname{Span} v$ and its orthogonal complement $(\operatorname{Span} v)^\perp$ are stable under f, the former because it is an eigenspace, the latter in this way: let $\langle v', v \rangle = 0$. Then $\langle f(v'), v \rangle = \langle v', f(v) \rangle = \overline{\lambda} \langle v', v \rangle = 0$.

By the induction hypothesis, $f\big|_{(\operatorname{Span} v)^\perp}$ has $n - 1$ orthogonal eigenvectors in W. They are still eigenvectors, and still orthogonal, when treated as members of V. Furthermore, they are orthogonal to v by construction.

As such f has a basis of orthogonal eigenvectors, so it has a basis of orthonormal eigenvectors as well. □

Lemma 2 (Spectral theorem, factorization version). *Let $A \in \mathbb{C}^{n\times n}$ satisfy $A = A^*$. Then there exist a unitary matrix U and a real diagonal matrix Λ such that $A = U\Lambda U^*$.*

Notice that $U^* = U^{-1}$, so no manual inversion necessary.

[1]The word *spectral* imputes an aura of magic and mystery. That is fitting because this theorem is very, very powerful.

Lemma 3 (Spectral theorem, projection version)**.** *Let $f : V \to V$ satisfy $f = f^*$. Let $\{v_1, v_2, \ldots, v_n\}$ be orthonormal eigenvectors of f with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.*

Then $f = \sum_{i=1}^n \lambda_i \operatorname{proj}_{v_i}$. That is, every self-adjoint operator is a real *linear combination of orthogonal projections.*

Lemma 4 (Spectral theorem, dyad version)**.** *The last and final form of the Spectral Theorem can be seen either as an expansion of the factorization version into outer products or as a translation of the projection version into orthonormal coordinates. Let $A \in \mathbb{C}^n$ satisfy $A = A^*$. Let $\{v_1, v_2, \ldots, v_n\}$ be orthonormal eigenvectors of f with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.*

Then $f = \sum_{i=1}^n \lambda_i v_i v_i^*$.

Application: direction of maximum amplification

Given a nonzero matrix $A \in \mathbb{C}^{n\times n}$, we might wonder what spatial direction gives you the best bang for your buck under left multiplication by A. That is, we are interested in the maximum amplification that A can exert on any vector:

$$\max_v \frac{\|Av\|}{\|v\|} \tag{3.10}$$

We can narrow our search to unit vectors.

$$\max_{\|v\|=1} \sqrt{\langle Av, Av\rangle} = \max_{\|v\|=1} \sqrt{\langle A^*Av, v\rangle} \tag{3.11}$$

A^*A is self-adjoint: $(A^*A)^* = (A^*)(A^*)^* = A^*A$. Diagonalize it as $A^*A = U\Lambda U^*$.

$$= \max_{\|v\|=1} \sqrt{\langle U\Lambda U^* v, v\rangle} \tag{3.12}$$

$$= \max_{\|v\|=1} \sqrt{\langle \Lambda U^* v, U^* v\rangle} \tag{3.13}$$

Change variables to $w = U^*v$.

$$= \max_{\|w\|=1} \sqrt{\langle \Lambda w, w\rangle} \tag{3.14}$$

This maximum is achieved when $w = e_i$, where λ_i is a maximal entry of Λ.

$$= \sqrt{\lambda_{\max}(A^*A)} \tag{3.15}$$

The furthest that A can magnify any vector is the square root of $\lambda_{\max}$, a maximal eigenvalue of A^*A. It quantifies how "big" A is, and is sometimes called are σ_1 or $\|A\|_2$, the operator norm of A.

Lecture 4

SVD I

Last lecture we computed a positive scalar, σ_1, that characterizes the magnitude of maximum amplification that a matrix A can effect on any vector.

4.1 Magnitude and direction(s)

The singular value decomposition does this and more:

singular values are the proportions of maximum amplification of orthogonal directions:

1. σ_1 is a maximum amplification of any direction.
2. σ_2 is a maximum amplification of any direction orthogonal to the direction σ_1 amplifies.
3. σ_3 is a maximum amplification of any direction orthogonal to the direction σ_1 amplifies *and* the direction σ_2 amplifies.
4. ...
5. $\sigma_{\operatorname{rank} A}$ is a minimum amplification of any direction not in the null space of A.

right singular vectors are the directions that are detected for amplification: v_1 is how σ_1 finds what it wants, etc.

left singular vectors are the directions that are output after amplification: σ_1 detects the quantity of v_1 in its input, amplifies it, and outputs its amplified version along u_1, etc.

The SVD can be seen as a generalization of the maxim "a vector is magnitude and direction" to matrices: "a matrix is magnitudes, input directions, and output directions." It is customary to state the magnitudes and directions of a matrix in order of importance.

Definition 19 (SVD, abstract version)**.** *Let $f : V \to U$ be a linear map of inner product spaces. A **singular value decomposition** of f is a choice of*

- *an orthonormal basis $\{v_1, v_2, \ldots, v_m\}$ for V (**right singular vectors**)*
- *an orthonormal basis $\{u_1, u_2, \ldots, u_n\}$ for U (**left singular vectors**), and*
- *positive scalars $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ (**singular values**); such that*

$f(v_i) = \sigma_i u_i$ for $1 \leq i \leq r = \operatorname{rank} f$.

Lemma 5 (A "square"[1] of a map)**.** *Let $f : V \to U$ be a linear map of inner product spaces.*

1. *The composite $f^* f : V \to V$ is self-adjoint.*

[1] In the same sense that $\overline{z}z$ is a nonnegative real number whose "size" is the square of the complex number z.

2. *The eigenvalues of $f^* f$ are nonnegative.*
3. *The null space of $f^* f$ is the same as that of f.*
4. *The rank of $f^* f$ is the same as that of f.*

Proof.
1. Proved yesterday.
2. Let $f^* f v = \lambda v$. Then $\left\langle f^* f v, v \right\rangle = \left\langle f v, f v \right\rangle = \lambda \langle v, v \rangle$, establishing λ as a ratio of nonnegative numbers.
3. Obviously the null space of f is contained in the null space of $f^* f$. We'll show that the null space of $f^* f$ is contained the null space of f. Let $f^* f v = 0$. Then $\left\langle f^* f v, v \right\rangle = \left\langle f v, f v \right\rangle = 0$ and $f v = 0$.
4. Follows from previous part by the rank-nullity theorem.

□

4.2 SVD exists

Theorem 16 (SVD exists). *Let $f : V \to U$ be a linear map of vector spaces, with $\dim V = n$ and $\dim U = m$.*

Proof. (Construction)

1. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0$ be the positive eigenvalues of $f^* f$, and let $\{v_1, v_2, \ldots, v_r\}$ be corresponding orthonormal eigenvectors.
2. Choose $\sigma_i = \sqrt{\lambda_i}$ for $i = 1 \leq i \leq r$.
3. Choose $u_i = f(v_i)/\sigma_i$ for $i = 1 \leq i \leq r$.
4. Complete $\{v_i\}_{1 \leq i \leq r}$ and $\{u_i\}_{1 \leq i \leq r}$ to orthonormal bases for V and U, respectively (the former having n vectors and the latter m).

(Verification) The basis $\{v_i\}$ for V is orthonormal, the purported singular values are in nonincreasing order, and $f(v_i) = \sigma_i u_i$ by construction.

All that remains to show is that $\{u_i\}_{1 \leq i \leq m}$ is orthonormal. By construction, $\{u_i\}_{n < i \leq m}$ is orthonormal, and is orthogonal to $\{u_i\}_{1 \leq i \leq r}$. So we just have to show that $\{u_i\}_{1 \leq i \leq r}$ is orthonormal. For $1 \leq i, j \leq r$,

$$\left\langle u_i, u_j \right\rangle = \left\langle \frac{f(v_i)}{\sigma_i}, \frac{f(v_j)}{\sigma_j} \right\rangle \tag{4.1}$$

$$= \frac{1}{\sigma_i \sigma_j} \left\langle (f^* f)(v_i), v_j \right\rangle \tag{4.2}$$

$$= \frac{1}{\sigma_i \sigma_j} \left\langle \sigma_i^2 v_i, v_j \right\rangle \tag{4.3}$$

$$= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{4.4}$$

□

Example: SVDs of the identity map/matrix

Let's find out what the SVDs of $I \in \mathbb{C}^{n\times n}$ are.

1. We have to form the matrix I^*I. It equals I.
2. We need a basis of orthonormal eigenvectors for $I^*I = I$. Any basis $\{v_1, v_2, \dots, v_n\}$ will do. (One could make such a basis using the Gram-Schmidt process by pulling random vectors out of hat.) Every eigenvalue of I^* is 1, so every singular value is 1.
3. The left singular vectors are I times the right singular vectors, divided by 1.

Therefore, the singular value decompositions of I are all orthonormal bases.

SVD as a matrix factorization

Let $A \in \mathbb{C}^{m\times n}$. Define $V \in \mathbb{C}^{n\times n}$ by

$$V = \begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix}, \tag{4.5}$$

where $v_1, v_2, \dots, v_n$ are right singular vectors in an SVD of A. Define

$$U = \begin{pmatrix} u_1 & u_2 & \dots & u_m \end{pmatrix}, \tag{4.6}$$

where $u_1, u_2, \dots, u_n$ are right singular vectors such that $f(v_i) = \sigma u_i$. Define a matrix $\Sigma \in \mathbb{C}^{m\times n}$ by

$$\Sigma_{ii} = \sigma_i. \tag{4.7}$$

Then the following factorization holds:

$$A = U\Sigma V^*. \tag{4.8}$$

Splitting this matrix product into its nonzero outer products,

$$A = \sum_{i=1}^{\operatorname{rank} A} \sigma_i \left(u_i {v_i}^*\right). \tag{4.9}$$

This form is powerful because it allows us to approximate A by simpler matrices. Suppose that A is a noisy measurement of a matrix that we earnestly believe to be of rank $r' < \operatorname{rank} A$. (Full rank matrices are the "hay in the haystack," so noise tends to be full rank.)

$$A \overset{\text{low rank}}{\approx} \sum_{i=1}^{r'} \sigma_i \left(u_i {v_i}^*\right). \tag{4.10}$$

Or suppose that A is a noisy measurement of a matrix we believe in truth to be unitary. (Unitary matrices are common in robotics.) Probably all of the singular values are close to 1. To get the nearest unitary matrix, we can set them all to 1.

$$A \overset{\text{unitary}}{\approx} \sum_{i=1}^{\operatorname{rank} A} u_i {v_i}^*. \tag{4.11}$$

Lecture 5

SVD II

5.1 Computing the SVD (review)

To compute the SVD $U\Sigma V^*$ of a matrix $A \in \mathbb{C}^{m\times n}$ with $\operatorname{rank} A = r$:

1. Form the product A^*A.
2. Identify the r positive eigenvalues of A^*A. Call them λ_i.
3. Identify r orthonormal eigenvectors v_i of A^*A such that $Av_i = \lambda_i$.
4. Define $\sigma_i = \sqrt{\lambda_i}$.
5. Define $u_i = \sigma_i^{-1} A v_i$.
6. Thus far U is an $m \times r$ matrix, Σ is an $r \times r$ matrix, and V is an $n \times r$ matrix. The factorization $A = U\Sigma V^*$ is sometimes reported at this stage, and is called the **truncated SVD**.
7. If the **full SVD** is desired, then complete the columns of U and V to orthonormal bases, and pad Σ with zeros so that it is $m \times n$.

Computation example

We will compute the SVD of the matrix A.

$$A = \begin{pmatrix} 1 & -j \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{5.1}$$

$$A^*A = \begin{pmatrix} 1 & 0 & 1 \\ j & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -j \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{5.2}$$

$$= \begin{pmatrix} 2 & -j \\ j & 2 \end{pmatrix} \tag{5.3}$$

$$\chi(s) = s^2 - 4s + 3 = (s-3)(s-1) \tag{5.4}$$

First eigenvalue and eigenvector:

$$\lambda_1 = 3 \tag{5.5}$$

$$(A^*A - 3I)\, v_1 = \begin{pmatrix} -1 & -j \\ j & -1 \end{pmatrix} v_1 = 0 \tag{5.6}$$

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ j \end{pmatrix} \tag{5.7}$$

Second eigenvalue and eigenvector:

$$\lambda_2 = 1 \tag{5.8}$$

$$(A^*A - I)\, v_2 = \begin{pmatrix} 1 & -j \\ j & 1 \end{pmatrix} v_2 = 0 \tag{5.9}$$

$$v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} j \\ 1 \end{pmatrix} \tag{5.10}$$

Singular values:

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3} \tag{5.11}$$

$$\sigma_2 = \sqrt{\lambda_2} = 1 \tag{5.12}$$

Left singular vectors:

$$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ j \\ 1 \end{pmatrix} \tag{5.13}$$

$$u_2 = \frac{Av_2}{\sigma_2} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ j \end{pmatrix} \tag{5.14}$$

Truncated SVD in the factorization style:

$$A = U_t \Sigma_t V_t^* = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \\ \frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{5.15}$$

and in the dyad style:

$$= \sum_{i=1}^{r} \sigma_i u_i v_i^* = \sqrt{3} \begin{pmatrix} \frac{2}{\sqrt{6}} \\ \frac{j}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \end{pmatrix} + 1 \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{j}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{5.16}$$

For a full SVD, make U square by orthogonalizing and normalizing the columns of $\begin{pmatrix} U_t & I \end{pmatrix}$ from left to right, dropping zero columns.

$$\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & 1 & 0 & 0 \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} & 0 & 1 & 0 \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} & 0 & 0 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{3} & 0 & 0 \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{j}{3} & 1 & 0 \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} & -\frac{1}{3} & 0 & 1 \end{pmatrix} \quad \text{(orthogonalize col. 3)} \tag{5.17}$$

$$\rightsquigarrow \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} & 0 & 0 \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{j}{\sqrt{3}} & 1 & 0 \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & 0 & 1 \end{pmatrix} \quad \text{(normalize col. 3)} \tag{5.18}$$

As there are three orthonormal columns, we are done. The following is the full SVD of A:

$$A = U_f \Sigma_f V_f^* = \begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{j}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{j}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{j}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-j}{\sqrt{2}} \\ \frac{-j}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{5.19}$$

Note that U_f and V_f are both invertible, but Σ_f has rank 2, which is the rank of A.

5.2 SVD of a wide matrix

To compute the SVD as we proved its existence, you need to form the product A^*A. If A has more columns than rows, this matrix is pretty big. The SVD of A can be computed more efficiently by computing the SVD of A^*, then using the following identity.

$$A^* = (U\Sigma V^*)^* = V\Sigma^* U^* \tag{5.20}$$

5.3 Application of SVD: PCA

Recall that the SVD can be used for dimensionality reduction as follows, for a matrix $A \in \mathbb{C}^{m\times n}$ of rank r.

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^* \tag{5.21}$$

$$\approx \sum_{i=1}^{r'} \sigma_i u_i v_i^* \tag{5.22}$$

By halting the sum early, at $r' < r$, we retain the r' biggest summands in a decomposition of A into rank 1 matrices.

Let $X \in \mathbb{C}^{n\times p}$ be a matrix of n points in p-space, collected as rows.[1] Each point can represent an observation, and each column a feature.

$$X = \begin{pmatrix} {x_1}^\top \\ {x_2}^\top \\ \vdots \\ {x_m}^\top \end{pmatrix} \tag{5.23}$$

Represent each point as a displacement from the sample average $\overline{x}$, and call this matrix of displacements $\widetilde{X}$.

$$\widetilde{X} = \begin{pmatrix} {x_1}^\top - \overline{x} \\ {x_2}^\top - \overline{x} \\ \vdots \\ {x_m}^\top - \overline{x} \end{pmatrix} \tag{5.24}$$

Use the SVD to write $\widetilde{X}$ as a sum of rank 1 matrices, from most important to least important. (Often $n \gg p$ and $\operatorname{rank} A = p$.)

$$= \sum_{i=1}^{r} \sigma_i u_i v_i^* \tag{5.25}$$

Two data science interpretations of this SVD are the following:

- The vectors v_i are projections onto orthogonal directions of maximum variance, from greatest to least.
- Often times we observe that the singular values fall off a cliff or become negligible. In that even a low-rank approximation to is appropriate.

We will explore both of these in the next lecture.

[1] Usually x_i means column i, but using x_i to represent row i is more common for data science, possibly because the way we write math means that it's easier to picture a lot of rows than a lot of columns.

Lecture 6

Principal component analysis

Suppose we have a matrix $\widetilde{X} \in \mathbb{C}^{n\times p}$ whose rows are centered observations $x_1 - \overline{x}, \ldots, x_n - \overline{x}$ and whose columns correspond to features:[1]

$$\widetilde{X} = \begin{pmatrix} (x_1 - \overline{x})^\top \\ (x_2 - \overline{x})^\top \\ \vdots \\ (x_n - \overline{x})^\top \end{pmatrix} \tag{6.1}$$

If $v \in \mathbb{C}^p$ is a unit vector, the matrix $\widetilde{X}\overline{v}$ is a list of scalar projections of $\widetilde{X}$'s rows onto v.

$$\widetilde{X}\overline{v} = \begin{pmatrix} (x_1 - \overline{x})^\top \overline{v} \\ (x_2 - \overline{x})^\top \overline{v} \\ \vdots \\ (x_n - \overline{x})^\top \overline{v} \end{pmatrix} = \begin{pmatrix} \langle (x_1 - \overline{x}), v \rangle \\ \langle (x_2 - \overline{x}), v \rangle \\ \vdots \\ \langle (x_n - \overline{x}), v \rangle \end{pmatrix} \tag{6.2}$$

Each inner product can be interpreted as having a factor of $\cos\theta$, where θ is the angle between the two vectors. Then $\widetilde{X}\overline{v}$ is larger when θ tends to be small, viz. when v points in the prevailing direction of deviation from $\overline{x}$. We can find v_1 capturing the direction of greatest variation, v_2 an orthogonal direction with second-greatest variation, etc. using the SVD. Define the covariance matrix $Q \in \mathbb{C}^{p\times p}$ as follows:

$$Q = \frac{1}{m-1}\left(\widetilde{X}\right)^* \widetilde{X} \tag{6.3}$$

This can be seen as the sample average value of xx^*.[2]

- Diagonal entry Q_{jj}[3] is the average squared distance that feature p lands from its average value. This called the variance of feature j.
- Off-diagonal entry Q_{jk} is what you expect on the average when you multiply feature j by the complex conjugate of feature k. It captures the correlation between feature j and feature k. If $Q_{jk} \neq 0$, then features j and k tend to move together.
  - If Q_k is positive, that means that features j and k tend to deviate in the same direction. When the former goes up, the latter goes up.

[1] Studying PCA in the context of complex observation would be considered exotic in the practice of statistics, but real-world PCA works just as well if you skip all the conjugation.

[2] Except the denominator is $m-1$ instead of m for statistical reasons. This is called Bessel's correction.

[3] Q_{-1}?

  - If Q_k is negative, then features j and k tend to deviate in opposite directions. When the former goes up, the latter goes up.
  - If Q_k is pure imaginary, then features j and k always move together, but at some angle in the complex plane.[4] When the former goes east, the latter goes north or south.

Let λ_i be the ith greatest eigenvalue of Q, and v_i a unit eigenvector that satisfies $Qv_i = \lambda_i v_i$.

- v_i is the ith **principal component** of $\widetilde{X}$. Projection onto v_1 gets more scalar variance out of $\widetilde{X}$ than any other direction. If a scatter plot of the rows of $\widetilde{X}$ looks like an ellipse in two dimensions, then v_1 is the semi-major axis and v_2 is the semi-minor axis.

- λ_i is the variance of $\widetilde{X}$ after projection onto the direction v_i. If a scatter plot of the rows of $\widetilde{X}$ looks like an ellipse in two dimensions, then $\sqrt{\lambda_1}$ is proportional to the length of the semi-major axis, and $\sqrt{\lambda_2}$ is proportional to the length of the semi-minor axis.

Scatter plots of empirical data come in all shapes and sizes, but after you project $\widetilde{X}$ onto its leading principal components (by taking inner products e.g. ${x_i}^\top \overline{v_j}$), the scatter plots all look quite the same, at least in these ways:

- The points are centered at the origin.

- The point cloud is longest along the axis of the foremost principal component.

6.1 Example: measuring an impedance by hand

Suppose that you are abducted by aliens. You are presented with an unknown linear circuit component, an AC voltage source, an oscilloscope, and a graphing calculator capable of linear algebra. The aliens will set you free, but only if you can tell them what the impedance Z is at a frequency ω.

You beg them to let you into your lab at Berkeley where you have a instrument that measures impedance, but they say no. Here is how you might get a pretty good guess:

1. Wire up your oscilloscope to plot voltage and current while you use your AC voltage source to induce a voltage $v(t) = V\cos(\omega t)$ over the mystery component.

2. Write down the current waveform as $i(t) = I\cos\big(\omega t + \phi\big)$.

3. Now you have a pair of phasors $x = (\tilde{V}, \tilde{I})$, where $\tilde{V} = V$ and $\tilde{I} = Ie^{j\phi}$.

4. Instead of immediately reporting the ratio $Z = \tilde{V}/\tilde{I}$, collect more data at a range of voltages to be extra sure.

5. Now you have a long list of points $\{x_1, x_2, \ldots, x_n\}$ where $x_i = (\tilde{V}_i, \tilde{I}_i)$.

6. Consult the aliens, who are able to visualize data in 4 spatial dimensions, and confirm that your data points do indeed look like a line in $\mathbb{C}^2$.

7. Perform PCA on your data to find the first principal component (a, b). When $\tilde{V}$ moves in a direction of a, $\tilde{I}$ moves in the direction b.

8. Report the slope a/b as your impedance estimate.

[4]The interpretation of this case, I admit, is quite bizarre. I believe it is never part of the statistical form of PCA, which works with real data only.