Math 210 — Fall 2019
Summary of the Course — Drew Armstrong

Points: Let $\mathbb{R}^n$ denote the set of $n \times 1$ matrices of real numbers:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

We call these "points" in n-dimensional Cartesian space.

Vectors: We will also think of a point $\mathbf{x}$ in $\mathbb{R}^n$ as a directed line segment (a "vector") with its tail at the origin $\mathbf{0}$ and its head at the point $\mathbf{x}$. This idea is subtle because we are allowed to pick up the arrow and move it as long as we don't change its length or direction.

Parallolgram Law: Consider two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. The points $\mathbf{0}, \mathbf{x}, \mathbf{y}$ form three vertices of a 2D parallelogram living in $\mathbb{R}^n$. The fourth vertex of the parallogram is

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

Subtraction of Vectors: Consider two points $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$. The vector with tail at $\mathbf{x}$ and head at $\mathbf{y}$ is represented by the point

$$\mathbf{y} - \mathbf{x} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} y_1 - x_1 \\ y_2 - x_2 \\ \vdots \\ y_n - x_n \end{pmatrix}.$$

Vector Arithmetic: Consider vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in $\mathbb{R}^n$ and numbers a, b in $\mathbb{R}$. Then we have

$$\begin{aligned} &\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}, \\ &\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}, \\ &a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}, \\ &(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}. \end{aligned}$$

Dot Product: Given vectors $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$, we define their dot product as the number

$$\mathbf{x} \bullet \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \bullet \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

More Vector Arithmetic: For all vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in $\mathbb{R}^n$ and numbers a in $\mathbb{R}$ we have

$$\begin{aligned} &\mathbf{x} \bullet \mathbf{y} = \mathbf{y} \bullet \mathbf{x}, \\ &\mathbf{x} \bullet (\mathbf{y} + a\mathbf{z}) = \mathbf{x} \bullet \mathbf{y} + a\mathbf{x} \bullet \mathbf{z}. \end{aligned}$$

Pythagorean Theorem: Given a vector $\mathbf{x}$ in $\mathbb{R}^n$ its "length" $\|\mathbf{x}\|$ is the non-negative number defined by

$$\|\mathbf{x}\|^2 = \mathbf{x} \bullet \mathbf{x} = x_1^2 + x_2^2 + \cdots x_n^2.$$

Law of Cosines: Consider two vectors $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$. These vectors together with their difference $\mathbf{y} - \mathbf{x}$ form the three sides of a 2D triangle in $\mathbb{R}^n$. By applying the formulas above we get

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2(\mathbf{x} \bullet \mathbf{y}).$$

On the other hand, the classical Law of Cosines for triangles tells us that

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\| \cos\theta,$$

where θ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. Then comparing the two equations gives

$$\mathbf{x} \bullet \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta.$$

In particular, this tells us that $\mathbf{x} \perp \mathbf{y}$ if and only if $\mathbf{x} \bullet \mathbf{y} = 0$.

Lines in $\mathbb{R}^2$: A line in the plane can be written in parametric form as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + t \begin{pmatrix} u \\ v \end{pmatrix}.$$

This is the line containing the point (x_0, y_0) and parallel to the vector (u, v). Or it can be expressed by an equation

$$ax + by = c$$

where (a, b) is some vector perpendicular ("normal") to the line. This line contains the origin $(0, 0)$ if and only if $c = 0$. In general, the line has minimum distance $c/\sqrt{a^2 + b^2}$ from the origin.

Planes in $\mathbb{R}^3$: A plane in 3-dimensional space can be written in parametric form as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + s \begin{pmatrix} u_1 \\ v_1 \\ w_1 \end{pmatrix} + t \begin{pmatrix} u_2 \\ v_2 \\ w_2 \end{pmatrix}.$$

This is the plane containing the point (x_0, y_0, z_0) and spanned by the vectors (u_1, v_1, w_1) and (u_2, v_2, w_2). Or it can be expressed by an equation

$$ax + by + cz = d$$

where (a, b, c) is some vector perpendicular ("normal") to the plane. This plane contains the origin $(0, 0, 0)$ if and only if $d = 0$. In general, the plane has minimum distance $d/\sqrt{a^2 + b^2 + c^2}$ from the origin.

Lines in $\mathbb{R}^3$: A line in 3-dimensional space can be written in parametric form as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + t \begin{pmatrix} u \\ v \\ w \end{pmatrix}.$$

This is the line containing the point (x_0, y_0, z_0) and parallel to the vector (u, v, w). However, a line in 3D can **not** be defined by a single equation. It **can** be defined as the solution of a system of two linear equations in three unknowns. Geometrically, this expresses the line as an intersection of two planes.

Systems of Linear Equations: A system of m linear equations in n unknowns has the following form:

$$\begin{cases} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & & & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{cases}$$

Alternatively, we can write it as a matrix equation:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

And we usually shorten it to this:

$$A\mathbf{x} = \mathbf{b}.$$

The Geometric Picture: A single linear equation $\mathbf{a} \bullet \mathbf{x} = b$ represents a flat $(n-1)$-dimensional shape living in $\mathbb{R}^n$, called a "hyperplane." A system of m linear equations $\mathbf{a}_i \bullet \mathbf{x} = b_i$ represents the intersection of m hyperplanes, which forms a flat d-dimensional shape (called a "d-plane") living in $\mathbb{R}^n$. Most likely we will have $d = n - m = \#$ variables $- \#$ equations.

Gaussian Elimination: A system of linear equations $A\mathbf{x} = \mathbf{b}$ can be solved by putting the system in Reduced Row Echelon Form (RREF) by Gaussian elimination. Each non-pivot column in the RREF leads to a free variable, so if there are d non-pivot columns in the RREF the solution will be a d-plane. Each non-pivot column also tells us an explicit non-trivial relation among the columns of A. We call d the "nullity" of A, and write $d = \text{null}(A)$.

Fundamental Theorem I: The set of vectors of the form $A\mathbf{x}$ is called the column space of the matrix A, because it consists of all linear combinations of the columns. If A has shape $m \times n$ then the column space is a subspace of $\mathbb{R}^m$. The dimension of the column space is called the "rank" of A, written $\text{rank}(A)$. It is equal to the number of pivot columns in the RREF. Since the total number of columns in the RREF is n, we obtain

$$\text{rank}(A) + \text{null}(A) = n.$$

Fundamental Theorem II: Let A have shape $m \times n$. Many times I have told you that the equation $A^T\mathbf{e} = \mathbf{0}$ means that the vector $\mathbf{e}$ is perpendicular to all of the columns of A. In other words, the nullspace of A^T is the "orthogonal complement" to the column space of A. It follows from this that their dimensions add to m, i.e., $\text{null}(A^T) + \text{rank}(A) = m$. Combining this with the Fundamental Theorem above gives the following surprising equation:

$$\text{rank}(A^T) = \text{rank}(A).$$

In other words, the row space and the column space of A have the same dimension!

Matrix Multiplication: Let A have shape $\ell \times m$ and let B have shape $m \times n$. Then the matrix AB exists and has shape $\ell \times n$. It is defined by requiring that the following equation holds for all $\mathbf{x}$ in $\mathbb{R}^n$:

$$(AB)\mathbf{x} = A(B\mathbf{x}).$$

However, if we want to actually **compute** the matrix AB we use the following rules:

$((i,j)\text{th entry of } AB) = (i\text{th row of } A)(j\text{th column of } B)$

$(i\text{th row of } AB) = (i\text{th row of } A)\ B$

$(j\text{th column of } AB) = A\ (j\text{th column of } B)$.

If $\mathbf{x}$ and $\mathbf{y}$ are column vectors of the same size then the language of matrix multiplication gives us a new notation for the dot product:

$$\mathbf{x}^T\mathbf{y} = \mathbf{x} \bullet \mathbf{y} = \mathbf{y} \bullet \mathbf{x} = \mathbf{y}^T\mathbf{x}.$$

Inverse Matrices: Let A have shape $m \times n$. We say that B is an inverse matrix of A if $AB = I_m$ and $BA = I_n$. But this is impossible unless $m = n$. (Reason: The matrix A has shape $m \times n$. If $m < n$ then RREF(A) has at least one non-pivot column so there exists some nonzero vector $\mathbf{x} \neq \mathbf{0}$ with $B\mathbf{x} = \mathbf{0}$. But then we have $\mathbf{x} = I\mathbf{x} = BA\mathbf{x} = B\mathbf{0} = \mathbf{0}$, which is a contradiction. If $m > n$ then apply the same argument to B.) If A has shape $n \times n$ then it **might** have an inverse. To compute the inverse we do this trick:

$$(A|I) \xrightarrow{\text{RREF}} \left(I|A^{-1}\right)$$

If the trick doesn't work (because A had some row relation or column relation) then we conclude that A has no inverse.

The Determinant: For any square matrix A there is a number $\det(A)$, called the determinant of A, which has the following property:

$$A \text{ is invertible} \quad \Longleftrightarrow \quad \det(A) \neq 0.$$

The determinant of a 2×2 matrix is defined as follows:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

And this leads to an explicit formula for the inverse of a 2×2 matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We did not compute the determinants of larger matrices in this class.

Uniqueness of Inverses: Suppose that we have $AB = I$ and $CA = I$. It follows that

$$C = CI = C(AB) = (CA)B = IB = B.$$

Hence if A has an inverse matrix, this matrix is unique. We give it the special name A^{-1}.

Matrix Arithmetic: Consider matrices A, B, C and numbers x, y. The following formulas hold as long as the respective matrices exist:

$A + B = B + A$

$A + (B + C) = (A + B) + C$

$A(BC) = (AB)C$

$(x + y)A = xA + yA$

$x(AB) = (xA)B = A(xB)$

$A(B + xC) = AB + xAC$

$(A + xB)C = AC + xBC$

$$
\begin{aligned}
&(A+B)^T = A^T + B^T \\
&(AB)^T = B^T A^T \\
&(AB)^{-1} = B^{-1}A^{-1} \\
&(A^T)^{-1} = (A^{-1})^T.
\end{aligned}
$$

WARNING: The following two formulas are NOT generally true:

$$
\begin{aligned}
&AB = BA \\
&(A+B)^{-1} = A^{-1} + B^{-1}.
\end{aligned}
$$

Solutions of a Linear System are Flat: Suppose we have two solutions of a linear system: $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{y} = \mathbf{b}$. Then for any number t we have

$$A(t\mathbf{x} + (1-t)\mathbf{y}) = tA\mathbf{x} + (1-t)A\mathbf{y} = t\mathbf{b} + (1-t)\mathbf{b} = \mathbf{b}.$$

This implies that every point of the line $t\mathbf{x} + (1-t)\mathbf{y}$ is also a solution. This is what I mean when I say that the solutions of a linear system form a d-plane. Geometrically: If m hyperplanes in n-dimensional space meet at two points $\mathbf{x}, \mathbf{y}$ then they also meet at the whole line $t\mathbf{x} + (1-t)\mathbf{y}$.

Least Squares Regression: Suppose that the linear system $A\mathbf{x} = \mathbf{b}$ has no solution. This means that the point $\mathbf{b}$ is not in the column space of A. In this case we want to find some $A\mathbf{x} = \mathbf{p}$ where the distance $\|\mathbf{p}-\mathbf{b}\|$ is as small as possible. This is achieved when $A^T(\mathbf{p}-\mathbf{b}) = \mathbf{0}$, i.e., when the vector $\mathbf{p} - \mathbf{b}$ is perpendicular to the column space of A. By combining these facts we obtain the "normal equation":

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

The most common application of this equation to fit a line to a collection of points.

Orthogonal Projection: Let A be an $m \times n$ matrix and suppose that the square $n \times n$ matrix $A^T A$ is invertible. Then from the previous topic we find that the orthogonal projection of $\mathbf{b}$ onto the column space of A is given by $\mathbf{p} = A\mathbf{x} = A(A^T A)^{-1}A^T\mathbf{b}$. In other words, then the matrix of the projection function is given by

$$P = A(A^T A)^{-1}A^T.$$

This is an $m \times m$ matrix satisfying $P^2 = P$ and $P^T = P$. If Q is the matrix that projects onto the nullspace of A^T (which consists of all vectors perpendicular to the column space of A) then we have

$$P + Q = I_m \quad \text{and} \quad PQ = QP = 0.$$

Special case: If $A = \mathbf{a}$ is a column vector (i.e., if $n = 1$) then the matrix that projects onto the line $t\mathbf{a}$ is given by

$$P = \mathbf{a}(\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T = \frac{1}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T = \frac{1}{\|\mathbf{a}\|^2}\mathbf{a}\mathbf{a}^T.$$

The matrix that projects onto the hyperplane $\mathbf{a}^T\mathbf{x} = 0$ is given by $I - P$.

Eigenvectors and Eigenvalues: Let A be an $n \times n$ matrix. We say that λ is an eigenvalue of A if there exists a nonzero vector $\mathbf{x} \neq \mathbf{0}$ such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

In this case we say that $\mathbf{x}$ is a λ-eigenvector of A. Equivalently, we say that λ is an eigenvalue of A when there exists a nonzero vector $\mathbf{x} \neq \mathbf{0}$ such that $(A-\lambda I)\mathbf{x} = \mathbf{0}$, i.e., when $\det(A-\lambda I) = 0$. In the case of a 2×2 matrix we can write this characteristic equation explicitly as follows:

$$0 = \det \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} = \lambda^2 - (a+d)\lambda + (ad-bc).$$

Hence the matrix has the following two eigenvalues:

$$\lambda = \frac{(a+d) \pm \sqrt{(a+d)^2 - 4(ad-bc)}}{2}.$$

Arithmetic of Eigenvalues: If a square matrix A satisfies some polynomial equation then its eigenvalues must satisfy the same equation. For example: If λ is an eigenvalue of A then λ^2 is an eigenvalue of A^2, $2\lambda - 1$ is an eigenvalue of $2A - I$ and $3\lambda^2 - \lambda + 5$ is an eigenvalue of $3A^2 - A + 5I$. Furthermore, if A is invertible then λ^{-1} is an eigenvalue of A^{-1}. Application: If P is a projection then the equation $P^2 - P = 0$ implies that every eigenvalue satisfies $\lambda^2 - \lambda = 0$, hence the only possible eigenvalues of P are 0 and 1.

Discrete Dynamical Systems: Suppose you have a linear recurrence relation defined by $\mathbf{v}_{n+1} = A\mathbf{v}_n$. If $\mathbf{v}_0$ is the initial condition then the nth state vector is given by

$$\mathbf{v}_n = A^n \mathbf{v}_0.$$

To solve this equation we first find the eigenvalues of A via the characteristic equation and then we find some corresponding eigenvectors. Suppose we find some eigenvectors:

$$A\mathbf{x} = \lambda \mathbf{x} \qquad \text{and} \qquad A\mathbf{y} = \mu \mathbf{y}.$$

Then we try to express our initial condition in terms of eigenvectors: $\mathbf{v}_0 = a\mathbf{x} + b\mathbf{y}$. If we're successful (i.e., if the matrix A has enough eigenvectors) then we can use this to obtain a "closed form" solution to the recurrence:

$$\begin{aligned} \mathbf{v}_n = A^n \mathbf{v}_0 &= A^n(a\mathbf{x} + b\mathbf{y}) \\ &= aA^n\mathbf{x} + bA^n\mathbf{y} \\ &= a\lambda^n \mathbf{x} + b\mu^n \mathbf{y}. \end{aligned}$$

If the matrix doesn't have enough eigenvectors (i.e., if the matrix is not "diagonalizable") then we might be out of luck.