Orthogonality Review 3/8/2002 | **Math 21b, O. Knill**

ORTHOGONALITY. $\vec{v}$ and $\vec{w}$ are called **orthogonal** if $\vec{v} \cdot \vec{w} = 0$.

Examples. 1) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 6 \\ -3 \end{bmatrix}$ are orthogonal in $\mathbf{R}^2$. 2) $\vec{v}$ and w are both orthogonal to $\vec{v} \times \vec{w}$ in $\mathbf{R}^3$.

$\vec{v}$ is called a **unit vector** if $||\vec{v}|| = \sqrt{\vec{v} \cdot \vec{v}} = 1$. $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$ are called **orthogonal** if they are pairwise orthogonal. They are called **orthonormal** if they are also unit vectors. A basis is called an **orthonormal basis** if it is orthonormal. For an orthonormal basis, the matrix $A_{ij} = \vec{v}_i \cdot \vec{v}_j$ is the unit matrix.

FACT. Orthogonal vectors are linearly independent and n orthogonal vectors in $\mathbf{R}^n$ form a basis.
Proof. The dot product of a **linear relation** $a_1\vec{v}_1 + \dots + a_n\vec{v}_n = 0$ with $\vec{v}_k$ gives $a_k\vec{v}_k \cdot \vec{v}_k = a_k||\vec{v}_k||^2 = 0$ so that $a_k = 0$. If we have n linear independent vectors in $\mathbf{R}^n$ then they automatically span the space.

ORTHOGONAL COMPLEMENT. A vector $\vec{w} \in \mathbf{R}^n$ is called **orthogonal** to a linear space V if $\vec{w}$ is orthogonal to every vector in $\vec{v} \in V$. The **orthogonal complement** of a linear space V is the set W of all vectors which are orthogonal to V. It forms a linear space because $\vec{v} \cdot \vec{w}_1 = 0, \vec{v} \cdot \vec{w}_2 = 0$ implies $\vec{v} \cdot (\vec{w}_1 + \vec{w}_2) = 0$.

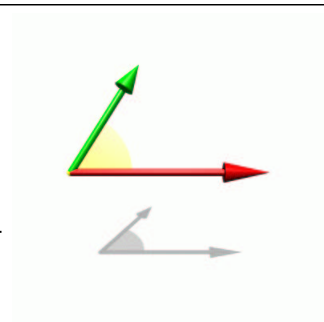
ORTHOGONAL PROJECTION. The **orthogonal projection** onto a linear space V with orthonormal basis $\vec{v}_1, \dots, \vec{v}_n$ is the linear map $T(\vec{x}) = \text{proj}_V(x) = (\vec{v}_1 \cdot \vec{x})\vec{v}_1 + \dots + (\vec{v}_n \cdot \vec{x})\vec{v}_n$. The vector $\vec{x} - \text{proj}_V(\vec{x})$ is in the orthogonal complement of V. (Note that $\vec{v}_i$ in the projection formula are unit vectors!)

PYTHAGORAS: If $\vec{x}$ and $\vec{y}$ are orthogonal, then $||\vec{x} + \vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2$. Proof. Expand $(\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y})$.
PROJECTIONS DO NOT INCREASE LENGTH: $||\text{proj}_V(\vec{x})|| \leq ||\vec{x}||$. Proof. Use Pythagoras: on $\vec{x} = \text{proj}_V(\vec{x}) + (\vec{x} - \text{proj}_V(\vec{x})))$. If $||\text{proj}_V(\vec{x})|| = ||\vec{x}||$, then $\vec{x}$ is in V.
CAUCHY-SCHWARTZ INEQUALITY: $|\vec{x} \cdot \vec{y}| \leq ||\vec{x}||\,||\vec{y}||$. Proof: $\vec{x} \cdot \vec{y} = ||\vec{x}||||\vec{y}|| \cos(\alpha)$.
If $|\vec{x} \cdot \vec{y}| = ||\vec{x}||||\vec{y}||$, then $\vec{x}$ and $\vec{y}$ are parallel.
TRIANGLE INEQUALITY: $||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||$. Proof: $(\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) = ||\vec{x}||^2 + ||\vec{y}||^2 + 2\vec{x} \cdot \vec{y} \leq ||\vec{x}||^2 + ||\vec{y}||^2 + 2||\vec{x}||||\vec{y}|| = (||\vec{x}|| + ||\vec{y}||)^2$.

ANGLE. The **angle** between two vectors $\vec{x}, \vec{y}$ is

$$\alpha = \arccos\left(\frac{\vec{x} \cdot \vec{y}}{||\vec{x}||||\vec{y}||}\right) .$$

CORRELATION. $\cos(\alpha) = \frac{\vec{x} \cdot \vec{y}}{||\vec{x}||||\vec{y}||}$ is called the **correlation** between $\vec{x}$ and $\vec{y}$. It is a number in $[-1, 1]$.



EXAMPLE. The angle between two orthogonal vectors is 90 degrees or 270 degrees. If $\vec{x}$ and $\vec{y}$ represent data showing the deviation from the mean, then $\frac{\vec{x} \cdot \vec{y}}{||\vec{x}||||\vec{y}||}$ is called the **statistical correlation** of the data.

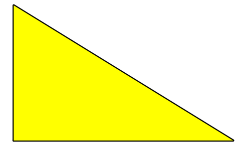
QUESTION. Express the fact that $\vec{x}$ is in the kernel of a matrix A using orthogonality.
ANSWER: $A\vec{x} = 0$ means that $\vec{w}_k \cdot \vec{x} = 0$ for every row vector $\vec{w}_k$ of $\mathbf{R}^n$.
REMARK. We will call later the matrix A^T, obtained by switching rows and columns of A the **transpose** of A. You see already that the image of A^T is orthogonal to the kernel of A.

QUESTION. Find a basis for the orthogonal complement of the linear space V spanned by $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \end{bmatrix}$.

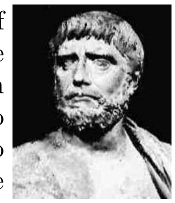
ANSWER: The orthogonality of $\begin{bmatrix} x \\ y \\ z \\ u \end{bmatrix}$ to the two vectors means solving the linear system of equations $x + 2y + 3z + 4w = 0, 4x + 5y + 6z + 7w = 0$. An other way to solve it: the kernel of $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 5 & 6 & 7 \end{bmatrix}$ is the orthogonal complement of V. This reduces the problem to an older problem.

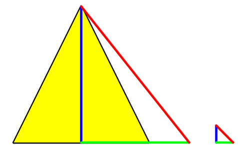
ON THE RELEVANCE OF ORTHOGONALITY.

1) During the pyramid age in Egypt (year -2800 BC until -2300 BC), the Egyptians used ropes divided into length ratios $3 : 4 : 5$ to build triangles. This allowed them to triangulate areas quite precisely: for example to build irrigation (the Nile was reshaping the land constantly) or to build the pyramids: For the **great pyramid at Giza** with a base length of 230 meters, the average error on each side is less then 20cm, an error of less then 1/1000. A key to achieve this was **orthogonality**.

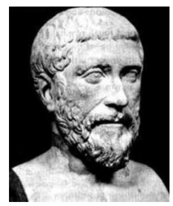


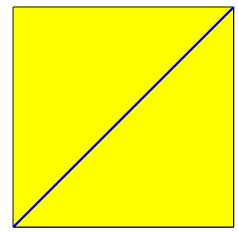
2) During one of Thales (-624 BC until -548 BC) journeys to Egypt, he used a geometrical trick to **measure the height** of the great pyramid. He measured the size of the shadow of the pyramid. Using a stick, he found the relation between the length of the stick and the length of its shadow. The same length ratio applies to the pyramid (**orthogonal** triangles). Thales found also that triangles inscribed into a circle and having as the base as the diameter must have a right angle.



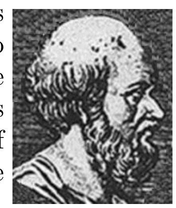


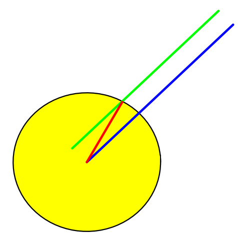
3) The Pythagoreans (-572 until -507) were interested in the discovery that the squares of a right angle would add up as $a^2 + b^2 = c^2$. They were puzzled in assigning a length to the diagonal of a square.



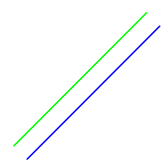


4) Eratosthenes (-274 until 194) realized that while the sun rays were **orthogonal** to the ground in the town of Scene, this did no more do so at the town of Alexandria where they would hit the ground at 7.2 degrees). Because the distance was about 500 miles and 7.2 is 1/50 of 360 degrees, he measured the circumference of the earth as 25'000 miles. This is pretty close to its actual value 24'874 miles.

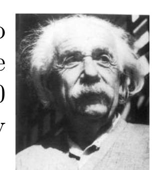




5) Closely related to **orthogonality** is **parallelism**. For a long time mathematicians tried to prove Euclid's parallel axiom using other postulates of Euclid (-325 until -265). These attempts had to fail because there are geometries in which parallel lines always meet (like on the sphere) or geometries, where parallel lines never meet (the Poincaré half plane). Also these geometries can be studied using linear algebra. The geometry on the sphere with **rotations**, the geometry of the Poincare half plane uses 2×2 matrices.

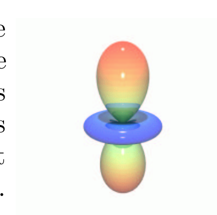


6) The question, whether in reality the angles of a right triangle always add up to 180 degrees became a real issue when geometries where discovered, in which the measurement depends on the position in space. Riemannian geometry, founded 150 years ago, is the foundation of **general relativity**, a theory which describes gravity geometrically: the presence of mass bends space-time.



7) In **probability theory** the notions **independence** or **decorrelated** appear. For example, when throwing dice, the number shown by the first dice is independent and therefore decorrelated from the number shown by the second dice. Decorrelation is identical to **orthogonality**, when vectors are associated to the random variables.

8) In **quantum mechanics**, states of atoms are described by functions which can be viewed as vectors also. The states with energy $-E_B/n^2$ (where $E_B = 13.6eV$ is the Bohr energy) in a hydrogen atom. States in an atom are **orthogonal**. Two states of two different atoms which don't interact are **orthogonal**. One of the challenges in quantum computing, where the computation deals with qubits (=vectors) is that orthogonality is not preserved during the computation. Different states can interact. This coupling is called **decoherence**.



Homework: Section 5.2, 2,14,16,34,40,42

MOTIVATION. The Gram-Schmidt process is an algorithm to build from an arbitrary basis an **orthonormal basis**. Why do we care to have an orthonormal basis?

- The process of producing an orthonormal basis from a matrix A with column vectors $\vec{v}_j$ will be associated to a factorization $A = QR$, which helps to solve linear equations. In physical problems like in astrophysics, the numerical methods to simulate the problems one needs to invert huge matrices in every time step of the evolution. (The reason why this is necessary sometimes is to assure the numerical method is stable (implicit methods)).
- An orthonormal basis looks like the standard basis $\vec{v}_1 = (1, 0, ..., 0), \dots, \vec{v}_n = (0, 0, ..., 1)$. We actually will see that one can turn an orthonormal basis into a standard basis or a mirror of the standard basis.
- If A is written as $A = QR$, then the inverse of A can be found easier, because the inverse of Q and the inverse of R are easy to obtain. See later.
- For many physical problems like in quantum mechanics or dynamics, matrices are **symmetric** $A^* = A$, where $A^*_{ij} = A_{ji}$. For such matrices, there is a natural orthonormal basis which has the nice property that $A\vec{v}_i = \lambda_i \vec{v}_i$. The $\vec{v}_i$ will be called eigenvectors and the λ_i are the eigenvalues. These objects will appear later in the course.
- The **formula for the projection** onto a linear subspace V simplifies with an orthonormal basis $\vec{v}_j$ in V:

$$\text{proj}_V(\vec{x}) = (\vec{v}_1 \cdot \vec{x})\vec{w}_1 + \cdots + (\vec{w}_n \cdot \vec{x})\vec{w}_n \; .$$

- For practical reasons, having an orthonormal basis simplifies life - partly because the presence of many zeros $\vec{w}_j \cdot \vec{w}_i = 0$ makes computations easier. This is especially the case for problems with symmetry.
- There is more behind the QR factorization: if we form $A = QR$ in then $A_1 = RQ$, the new matrix $A_1 = Q^{-1}AQ$ shares many properties of A (like the eigenvalues about which we will learn in a few weeks). When iterating this procedure $A \to A_1$ one is lead to interesting topics in differential equations (an example is the Toda lattice).

GRAM-SCHMIDT PROCESS.
Let $\vec{v}_1, ..., \vec{v}_n$ be a basis. Let $\vec{u}_1 = \vec{v}_1$ and $\vec{w}_1 = \vec{u}_1/||\vec{u}_1||$. The Gram-Schmidt process recursively constructs from the already constructed orthonormal set $\vec{w}_1, ..., \vec{w}_{i-1}$ which spans a linear space V_{i-1} the new vector $\vec{u}_i = (\vec{v}_i - \text{proj}_{V_{i-1}}(\vec{v}_i))$ which is orthogonal to V_{i-1}, and then normalizing $\vec{u}_i$ to to get $\vec{w}_i = \vec{u}_i/||\vec{u}_i||$. The vectors $\vec{u}_i$ are orthogonal to the linear space V_{i-1}.

EXAMPLE.

Find an orthonormal basis for $\vec{v}_1 = \begin{vmatrix} 2 \\ 0 \\ 0 \end{vmatrix}$, $\vec{v}_2 = \begin{vmatrix} 1 \\ 3 \\ 0 \end{vmatrix}$ and $\vec{v}_3 = \begin{vmatrix} 1 \\ 2 \\ 5 \end{vmatrix}$.

SOLUTION.

1. $\vec{w}_1 = \vec{v}_1/||\vec{v}_1|| = \begin{vmatrix} 1 \\ 0 \\ 0 \end{vmatrix}$.

2. $\vec{u}_2 = (\vec{v}_2 - \text{proj}_{V_1}(\vec{v}_2)) = \vec{v}_2 - (\vec{w}_1 \cdot \vec{v}_2)\vec{w}_1, = \begin{vmatrix} 0 \\ 3 \\ 0 \end{vmatrix}$. $\vec{w}_2 = \vec{u}_2/||\vec{u}_2|| = \begin{vmatrix} 0 \\ 1 \\ 0 \end{vmatrix}$.

3. $\vec{u}_3 = (\vec{v}_3 - \text{proj}_{V_2}(\vec{v}_3)) = \vec{v}_3 - (\vec{w}_1 \cdot \vec{v}_3)\vec{w}_1 - (\vec{w}_2 \cdot \vec{v}_3)\vec{w}_2 = \begin{vmatrix} 0 \\ 0 \\ 5 \end{vmatrix}$, $\vec{w}_3 = \vec{u}_3/||\vec{u}_3|| = \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix}$.

QR FACTORIZATION.
The formulas can be written as

$$\vec{v}_1 = ||\vec{v}_1||\vec{w}_1 = r_{11}\vec{w}_1$$
$$\cdots$$
$$\vec{v}_i = (\vec{w}_1 \cdot \vec{v}_i)\vec{w}_1 + \cdots + (\vec{w}_{i-1} \cdot \vec{v}_i)\vec{w}_{i-1} + ||\vec{u}_i||\vec{w}_i = r_{i1}\vec{w}_1 + \cdots + r_{ii}\vec{w}_i$$
$$\cdots$$
$$\vec{v}_n = (\vec{w}_1 \cdot \vec{v}_n)\vec{w}_1 + \cdots + (\vec{w}_{n-1} \cdot \vec{v}_n)\vec{w}_{n-1} + ||\vec{u}_n||\vec{w}_n = r_{n1}\vec{w}_1 + \cdots + r_{nn}\vec{w}_n$$

which means in matrix form

$$A = \begin{vmatrix} | & | & \cdot & | \\ \vec{v}_1 & \cdots & \cdot & \vec{v}_m \\ | & | & \cdot & | \end{vmatrix} = \begin{vmatrix} | & | & \cdot & | \\ \vec{w}_1 & \cdots & \cdot & \vec{w}_m \\ | & | & \cdot & | \end{vmatrix} \begin{vmatrix} r_{11} & r_{12} & \cdot & r_{1m} \\ 0 & r_{22} & \cdot & r_{2m} \\ 0 & 0 & \cdot & r_{mm} \end{vmatrix} = QR$$

where A and Q are $n \times m$ matrices and R is a $m \times m$ matrix.

Any matrix A can be decomposed as $A = QR$, where Q has as columns orthonormal vectors and R is an upper triangular square matrix.

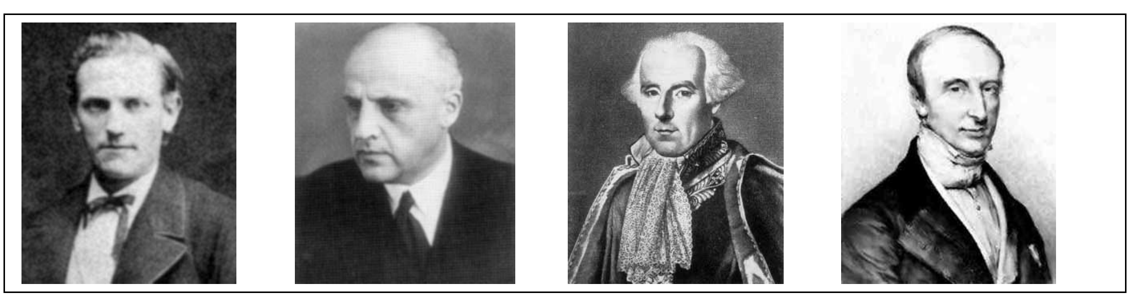
BACK TO THE EXAMPLE. The matrix with the vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ is $A = \begin{vmatrix} 2 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{vmatrix}$.

$\vec{v}_1 = ||\vec{v}_1||\vec{w}_1$
$\vec{v}_2 = (\vec{w}_1 \cdot \vec{v}_2)\vec{w}_1 + ||\vec{u}_2||\vec{w}_2$
$\vec{v}_3 = (\vec{w}_1 \cdot \vec{v}_3)\vec{w}_1 + (\vec{w}_2 \cdot \vec{v}_3)\vec{w}_2 + ||\vec{u}_3||\vec{w}_3$,
so that $Q = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$ and $R = \begin{vmatrix} 2 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{vmatrix}$.

PRO MEMORIA.
While building the matrix R we keep track of the vectors u_i during the Gram-Schmidt procedure. At the end you have vectors $\vec{u}_i, \vec{v}_i, \vec{w}_i$ and the matrix R has the $||\vec{u}_i||$ in the diagonal as well as the dot products $\vec{w}_i \cdot \vec{v}_j$ in the upper right triangle.

PROBLEM. Make the QR decomposition of $A = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix}$. $\vec{w}_1 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$. $\vec{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$.
$\vec{w}_2 = \vec{u}_2$. $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = QR$.

SOME HISTORY.
The recursive formulae of the process were stated by Erhard Schmidt (1876-1959) in 1907 Implicitly the essence of the formulae were in a 1883 paper of J.P.Gram in 1883 which Schmidt mentions in a footnote. The process seems to be a result of Laplace (1749-1827). It was already also used by Cauchy (1789-1857) in 1836.



Gram Schmidt Laplace Cauchy