Basic Math for Computer Graphics

Sungkil Lee
Sungkyunkwan University

Objectives

- Much of graphics is just translating math directly into code.
  - The cleaner the math, the cleaner the resulting code.
  - Also, clean codes result in better performance in many cases.

- In this lecture, we review various tools from high school and college mathematics.

- This chapter is not intended to be a rigorous treatment of the material; instead, intuition and geometric interpretation are emphasized.

Sets and Mappings

- Sets
  - a is a member of set S

$$a \in S$$

  - Cartesian product of two sets: given any two sets A and B,

$$A \times B = \{(a, b) | a \in A \text{ and } b \in B\}$$

    * As a shorthand, we use the notation A^2 to denote $A \times A$.

Sets and Mappings

- Common sets of interest include:
  - $\mathbb{R}$: the real numbers
  - $\mathbb{R}^+$: the non-negative real numbers (includes zero)
  - $\mathbb{R}^2$: the ordered pairs in the real 2D plane
  - $\mathbb{R}^n$: the points in n-dimensional Cartesian space
  - $\mathbb{Z}$: the integers

Sets and Mappings

- Mappings (also called functions)

$$f : \mathbb{R} \mapsto \mathbb{Z}$$

  - "There is a function called f that takes a real number as input and maps it to an integer."
  - equivalent to the common programming notation :

$$\text{integer } f(\text{real}) \qquad \leftarrow \text{equivalent} \rightarrow \qquad f : \mathbb{R} \mapsto \mathbb{Z}$$

    "There is a function called f which has one real argument and returns an integer."

Trigonometry

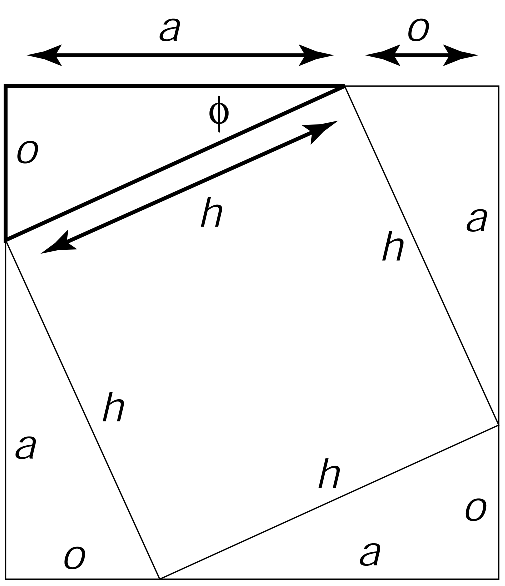
- The conversion between degrees and radians:

$$degrees = \frac{180}{\pi}\, radians$$

$$radians = \frac{\pi}{180}\, degrees$$

Trigonometry

- Trigonometric functions
  - Pythagorean theorem: $a^2 + o^2 = h^2$



$$\sin\phi \equiv o/h \qquad \csc\phi \equiv h/o$$
$$\cos\phi \equiv a/h \qquad \sec\phi \equiv h/a$$

$$\tan\phi \equiv o/a \qquad\qquad \cot\phi \equiv a/o$$

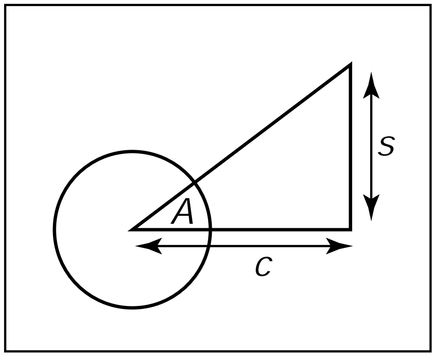
Trigonometry

- The functions are not invertible when considered with the domain $\mathbb{R}$. This problem can be avoided by restricting the range of standard inverse functions, and this is done in a standard way in almost all modern math libraries.

- The domains and ranges are:

$$\begin{aligned}
&\arcsin\ (\text{asin}) : [-1,1] \mapsto [-\pi/2, \pi/2] \\
&\arccos\ (\text{acos}) : [-1,1] \mapsto [0, \pi] \\
&\arctan\ (\text{atan}) : \mathbb{R} \mapsto [-\pi/2, \pi/2] \\
&\arctan 2\ (\text{atan2}) : \mathbb{R}^2 \mapsto [-\pi, \pi]
\end{aligned}$$

Trigonometry



- The $\text{atan2}(s, c)$ is often very useful in graphics: It takes an s value proportional to $\sin A$ and a c value that scales $\cos A$ by the same factor, and returns A.

Vector Spaces (in Algebratic Math)

- A *vector space* over a *field* F is a set V together with addition and multiplication that satisfy the eight axioms. Elements of V and F are called *vectors* and *scalars*.

Axiom	Meaning
Associativity of addition	$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
Commutativity of addition	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
Identity element of addition	There exists an element $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.
Inverse elements of addition	For every $\mathbf{v} \in V$, there exists an element $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
Distributivity of scalar multiplication with respect to vector addition	$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
Distributivity of scalar multiplication with respect to field addition	$(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$
Compatibility of scalar multiplication with field multiplication	$a(b\mathbf{v}) = (ab)\mathbf{v}$
Identity element of scalar multiplication	$1\mathbf{v} = \mathbf{v}$, where 1 denotes the multiplicative identity in F.

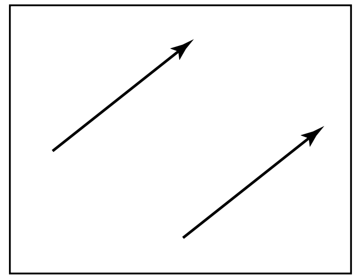
Vector Spaces (in Algebratic Math)

- A *vector space* over a *field* F is a set V together with addition and multiplication that satisfy the eight axioms. Elements of V and F are called *vectors* and *scalars*.

- Mathematical structures related to the concept of a field can be tracked as follows:

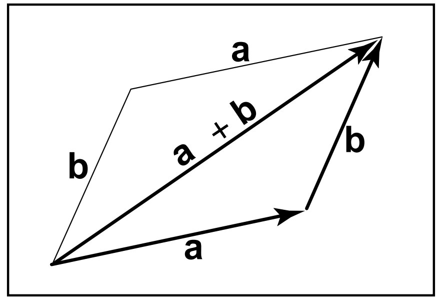
- A *field* is a *ring* whose nonzero elements form an *abelian group* under multiplication.
- A *ring* is an *abelian group* under addition and a *semigroup* under multiplication; addition is commutative, addition and multiplication are associative, multiplication distributes over addition, each element in the set has an additive inverse, and there exists an additive identity.
- An *abelian group* (*commutative group*) is a group in which commutativity ($a \cdot b = b \cdot a$) is satisfied.
- A *semigroup* is a set A in which $a \cdot b$ satisfies associativity for any two elements a and b and opeator $\cdot$.
- A *group* is a set A in which $a \cdot b$ satisfies closure, associativity, identity element, and inverse element for any two elements a and b and opeator $\cdot$.

Vectors (Simply)



- A quantity that encompasses a length and a direction.
- Represented by an arrow and not as coordinates or numbers
- Length: $\|\mathbf{a}\|$
- A unit vector: Any vector whose length is one.
- The zero vector: the vector of zero length. The direction of the zero vector is undefined.

Vectors (Simply)



- Two vectors are added by arranging them head to tail. This can be done in either order.

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$$

- Vectors can be used to store an offset, also called a displacement, and a location (or position) that is a displacement from the origin.

Cartesian Coordinates of a Vector

- Vectors in a n-D vector space are said to be *linearly independent*, iff

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

has only the trivial solution $(a_1 = a_2 = \cdots = a_n = 0)$.

  – The vectors are thus referred to as basis vectors.

  – For example, a 2D vector $\mathbf{c}$ may be expressed as a combination of two basis vectors $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{c} = a_c\mathbf{a} + b_c\mathbf{b},$$

where a_c and b_c are the Cartesian coordinates of the vector $\mathbf{c}$ with respect to the basis vectors $\{\mathbf{a}, \mathbf{b}\}$.

Cartesian Coordinates of a Vector

- Assuming vectors, $\mathbf{x}$ and $\mathbf{y}$, are orthonormal,

  $$\mathbf{a} = x_a\mathbf{x} + y_a\mathbf{y}$$

  the length of $\mathbf{a}$ is

  $$\|\mathbf{a}\| = \sqrt{{x_a}^2 + {y_a}^2}$$

  By convention we write the coordinates of $\mathbf{a}$ either as an ordered pair (x_a, y_a) or a column matrix:

  $$\mathbf{a} = \begin{bmatrix} x_a \\ y_a \end{bmatrix}$$

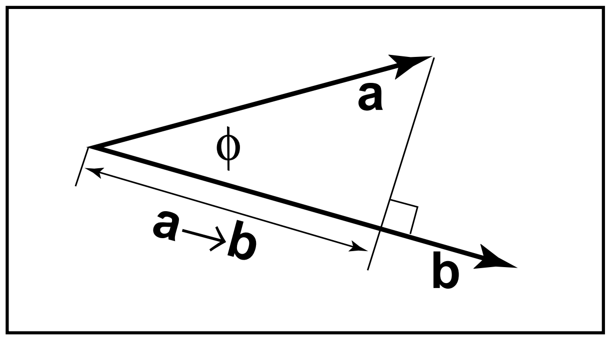
  $$\mathbf{a}^\top = \begin{bmatrix} x_a & y_a \end{bmatrix}$$

Dot Product

- The simplest way to multiply two vectors (also called the scalar product).

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \phi$$

- The projection of one vector onto another: the length $\mathbf{a} \to \mathbf{b}$ of the projection of $\mathbf{a}$ that is projected onto $\mathbf{b}$



$$\mathbf{a} \to \mathbf{b} = \|\mathbf{a}\| \cos \phi = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|}$$

Dot Product

- If 2D vectors $\mathbf{a}$ and $\mathbf{b}$ are expressed in Cartesian coordinates,

$$\begin{aligned}\mathbf{a}\cdot\mathbf{b} &= (x_a\mathbf{x} + y_a\mathbf{y}) \cdot (x_b\mathbf{x} + y_b\mathbf{y}) \\ &= x_a x_b(\mathbf{x}\cdot\mathbf{x}) + x_a y_b(\mathbf{x}\cdot\mathbf{y}) + x_b y_a(\mathbf{y}\cdot\mathbf{x}) + y_a y_b(\mathbf{y}\cdot\mathbf{y}) \\ &= x_a x_b + y_a y_b\end{aligned}$$

- Similarly in 3D,

$$\mathbf{a}\cdot\mathbf{b} = x_a x_b + y_a y_b + z_a z_b$$

Cross Product

$$\|\mathbf{a} \times \mathbf{b}\| = \|\mathbf{a}\| \|\mathbf{b}\| \sin \phi$$

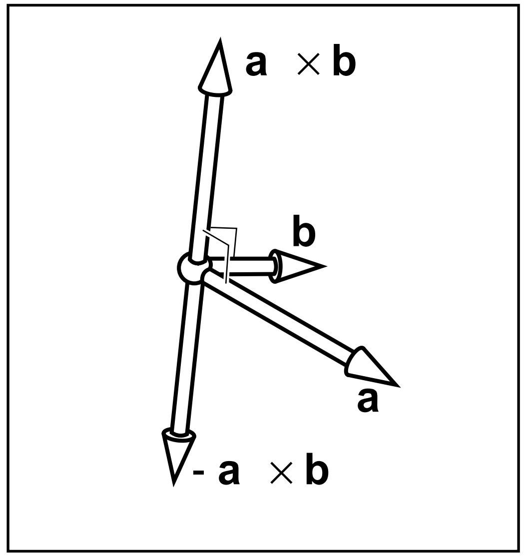
- By definition the unit vectors in the positive direction of the $x-$, $y-$ and $z-$axes are given by

$$\mathbf{x} = (1, 0, 0),$$
$$\mathbf{y} = (0, 1, 0),$$
$$\mathbf{z} = (0, 0, 1),$$

  and we set as a convention that $\mathbf{x} \times \mathbf{y}$ must be in the plus or minus $\mathbf{z}$ direction.

$$\mathbf{z} = \mathbf{x} \times \mathbf{y}$$

Cross Product



- The "right-hand rule":
  Imagine placing the base of your right palm where $\mathbf{a}$ and $\mathbf{b}$ join at their tails, and pushing the arrow of $\mathbf{a}$ toward $\mathbf{b}$. Your extended right thumb should point toward $\mathbf{a} \times \mathbf{b}$.

2D Implicit Curves

- Curve: A set of points that can be drawn on a piece of paper without lifting the pen.

- A common way to describe a curve is using an implicit equation.

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

$$f(x, y) = (x - x_c)^2 + (y - y_c)^2 - r^2$$

- If $f(x, y) = 0$, the points where this equality hold are on the circle with center (x_c, y_c) and radius r.

2D Implicit Curves

- "implicit" equation: The points (x, y) on the curve cannot be immediately calculated from the equation, and instead must be determined by plugging (x, y) into f and finding out whether it is zero or by solving the equation.

- The curve partitions space into regions where $f > 0$, $f < 0$ and $f = 0$.

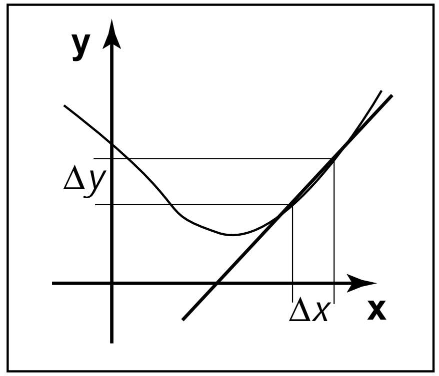
The 2D Gradient

- If the function $f(x, y)$ is a height field with height $= f(x, y)$, the gradient vector points in the direction of maximum upslope, i.e., straight uphill. The gradient vector $\nabla f(x, y)$ is given by

$$\nabla f(x, y) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$$

- The gradient vector evaluated at a point on the implicit curve $f(x, y) = 0$ is perpendicular to the tangent vector of the curve at that point. This perpendicular vector is usually called the *normal* vector to the curve.

The 2D Gradient



- The derivative of a 1D function measures the slope of the line tangent to the curve.
- If we hold y constant, we can define an analog of the derivative, called the partial derivative :

$$\frac{\partial f}{\partial x} \equiv \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

Implicit 2D Lines

- The familiar "slope-intercept" form of the line is

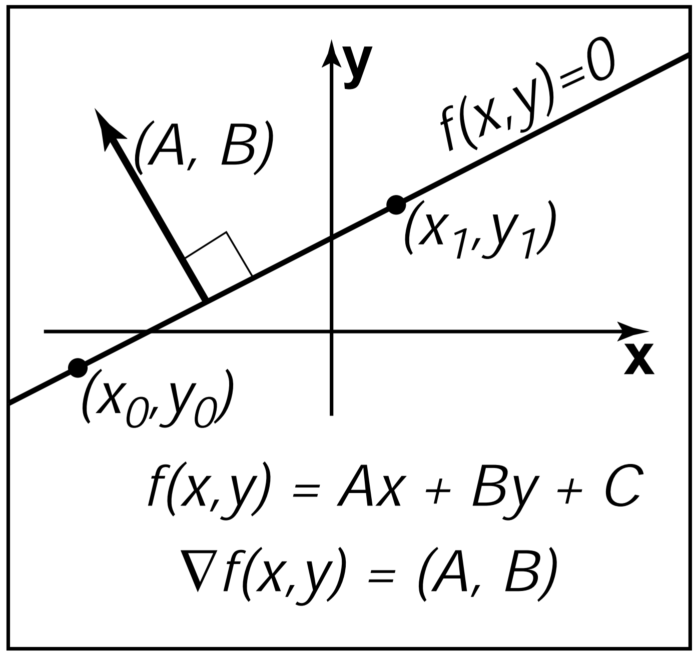
$$y = mx + b$$

This can be converted easily to implicit form

$$\begin{aligned} y - mx - b = 0 \\ Ax + By + C = 0 \end{aligned}$$

Implicit 2D Lines

- The gradient vector (A, B) is perpendicular to the implicit line $Ax + By + C = 0$



2D Parametric Curves

- A parametric curve: controlled by a *single* parameter, t,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} g(t) \\ h(t) \end{bmatrix}$$

  Vector form :

$$\mathbf{p} = f(t)$$

  where f is a vector valued function $f : \mathbb{R} \mapsto \mathbb{R}^2$

2D Parametric Lines

- A parametric line in 2D that passes through points $\mathbf{p}_0 = (x_0, y_0)$ and $\mathbf{p}_1 = (x_1, y_1)$ can be written

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_0 + t(x_1 - x_0) \\ y_0 + t(y_1 - y_0) \end{bmatrix}$$

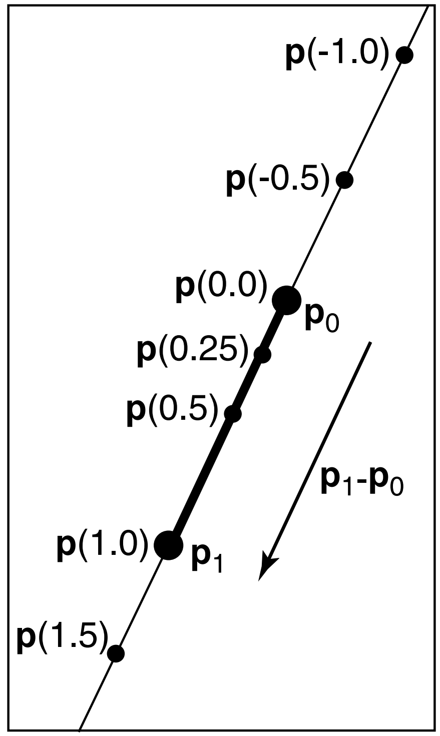
Because the formulas for x and y have such similar structure, we can use the vector form for $\mathbf{p} = (x, y)$:

$$\mathbf{p}(t) = \mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0)$$

Parametric lines can also be described as just a point $\mathbf{o}$ and $\mathbf{a}$ vector $\mathbf{d}$:

$$\mathbf{p}(t) = \mathbf{o} + t(\mathbf{d})$$

2D Parametric Lines



- A 2D parametric line through $\mathbf{P}_0$ and $\mathbf{P}_1$. The line segment defined by $t \in [0, 1]$ is shown in bold.

Linear Interpolation

- Most common mathematical operation in graphics.

- Example of linear interpolation: Position to form line segments in 2D and 3D

$$\mathbf{p} = (1 - t)\mathbf{a} + t\mathbf{b}$$

  - interpolation: $\mathbf{p}$ goes through $\mathbf{a}$ and $\mathbf{b}$ exactly at $t = 0$ and $t = 1$
  - linear interpolation: the weighting terms t and $1 - t$ are linear polynomials of t

Linear Interpolation

- Example of linear interpolation: A set of positions on the x-axis: $x_0, x_1, ..., x_n$ and for each x_i we have an associated height, y_i.
  - A continuous function $y = f(x)$ that interpolates these positions, so that f goes through every data point, i.e., $f(x_i) = y_i$.
  - For linear interpolation, the points (x_i, y_i) are connected by straight line segments.
  - It is natural to use parametric line equations for these segments. The parameter t is just the fractional distance between x_i and x_{i+1}:

$$f(x) = y_i + \frac{x - x_i}{x_{i+1} - x_i}(y_{i+1} - y_i)$$

Linear Interpolation

- In the common form of linear interpolation, create a variable t that varies from 0 to 1 as we move from data A to data B. Intermediate values are just the function $(1-t)A+tB$.

$$t = \frac{x - x_i}{x_{i+1} - x_i}$$

2D Triangles

- With origin and basis vectors, any point $\mathbf{p}$ can be written:

$$\mathbf{p} = (1-\beta-\gamma)\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$$
$$\alpha \equiv 1-\beta-\gamma$$
$$\mathbf{p}(\alpha,\beta,\gamma) = \alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$$

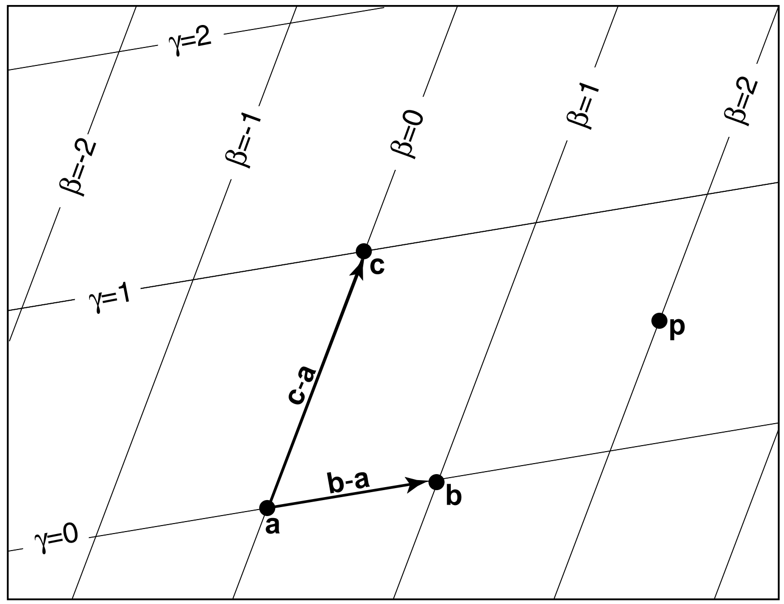
  with the constraint that

$$\alpha + \beta + \gamma = 1$$

- A particularly nice feature of barycentric coordinates is that a point $\mathbf{p}$ is inside the triangle formed by $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ if and only if

$$0 < \alpha < 1,$$
$$0 < \beta < 1,$$
$$0 < \gamma < 1.$$

2D Triangles



- A 2D triangle (vertices $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$) can be used to set up a non-orthogonal coordinate system with origin $\mathbf{a}$ and basis vectors $(\mathbf{b}-\mathbf{a})$ and $(\mathbf{c}-\mathbf{a})$. A point is then represented by an ordered pair (β, γ). For example, the point $\mathbf{p} = (2.0, 0.5)$, i.e., $\mathbf{p} = \mathbf{a} + 2.0(\mathbf{b}-\mathbf{a}) + 0.5(\mathbf{c}-\mathbf{a})$.

2D Triangles

- Defined by 2D points $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$. Its area:

$$
\begin{aligned}
area &= \frac{1}{2}\|(\mathbf{b}-\mathbf{a}) \times (\mathbf{c}-\mathbf{a})\| \\
&= \frac{1}{2}\left\| \begin{matrix} x_b - x_a & x_c - x_a \\ y_b - y_a & y_c - y_a \end{matrix} \right\| \\
&= \frac{1}{2}((x_b - x_a)(y_c - y_a) - (x_c - x_a)(y_b - y_a)) \\
&= \frac{1}{2}(x_a y_b + x_b y_c + x_c y_a - x_a y_c - x_b y_a - x_c y_b)
\end{aligned}
$$

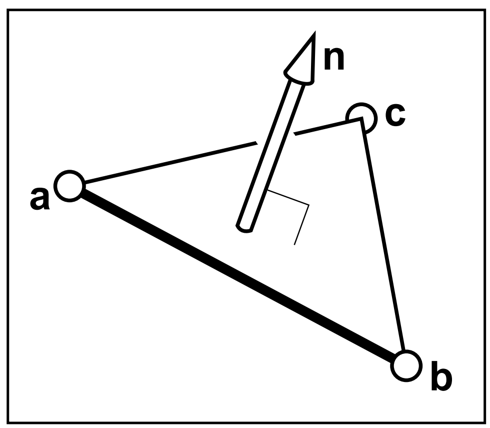
3D Triangles

- Barycentric coordinates extend almost transparently to 3D. If we assume the points $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are 3D,

$$\mathbf{p} = (1 - \beta - \gamma)\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$$

- The normal vector: taking the cross product of any two vectors.

$$\mathbf{n} = (\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})$$



3D Triangles

This normal vector is not necessarily of unit length, and it obeys the right-hand rule of cross products.

- The area of the triangle: Taking the length of the cross product.

$$area = \frac{1}{2}\|(\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})\|$$

Vector Operations in Matrix Form

- We can use matrix formalism to encode vector operations for vectors when using Cartesian coordinates; if we consider the result of the dot product a one by one matrix, it can be written:

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b}$$

- If we take two 3D vectors we get:

$$\begin{bmatrix} x_a & y_a & z_a \end{bmatrix} \begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix} = \begin{bmatrix} x_a x_b + y_a y_b + z_a z_b \end{bmatrix}$$

Matrices and Determinants

- The determinant in 2D is the area of the parallelogram formed by the vectors. We can use matrices to handle the mechanics of computing determinants.

$$\begin{bmatrix} a & A \\ b & B \end{bmatrix} = \begin{bmatrix} a & b \\ A & B \end{bmatrix} = aB - Ab$$

- For example, the determinant of a particular 3×3 matrix

$$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} = 0 \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} + 1 \begin{bmatrix} 5 & 3 \\ 8 & 6 \end{bmatrix} + 2 \begin{bmatrix} 3 & 4 \\ 6 & 7 \end{bmatrix}$$

$$= 0(32 - 35) + 1(30 - 24) + 2(21 - 24)$$

$$= 0$$