Bézier Curve Modelling[2]

A. Karamancıoğlu

Eskişehir Osmangazi University

June 3, 2023

[2] These slides are intended for educational use only; not for sale under any circumstances.

1. Complex Numbers
2. Fields
3. Vector Spaces
4. Properties of Vector Spaces
5. Subspaces
6. Sums and Direct Sums
7. Span and Linear Independence
8. Bases
9. Dimension
10. Linear Combinations in Vector Spaces
11. Linear Combinations of Points

12 Sum of a Point and a Vector

13 Affine Space

14 Examples of Affine Spaces

15 Chasles's Identity

16 Affine Combinations, Barycenters

17 Affine Subspaces

18 Affine maps

19 Lagrange Interpolation and Neville's Algorithm

20 De Casteljau's Algorithm

21 Bezier Clipping Algorithm

22 Bézier curve and the control points

[4] These slides are intended for educational use only; not for sale under any circumstances.

23 Bézier Curve Properties

24 Shape preserving linear maps

25 Modifying the parameter interval

26 Derivative of a Cubic Bézier Curve

27 Splitting a Bézier Curve

28 Joining Two Bézier Curves with C^1-Continuity

29 An application: Borders of a half apple

30 Fréchet distance

31 Hausdorff distance

32 Bézier Surfaces

33 Geogebra: A Tool for Bezier Curve Construction

[4] These slides are intended for educational use only; not for sale under any circumstances.

Sources

1) S Axler, Linear Algebra Done Right, 2nd ed., Springer, 1997.
2) J. Gallier, Geometric Methods and Applications, Springer, 2011
3) R, Goldman, Pyramid Algorithms: A Dynamic Programming Approach to Curves and Surfaces for Geometric Modeling, Amazon, 2003

Complex Numbers

A complex number is an ordered pair (a, b), where $a, b \in \mathbb{R}$, but we will write this as $a + bi$. The set of all complex numbers is denoted by $\mathbb{C}$:

$$\mathbb{C} = \{a + bi : \ a, b \in \mathbb{R}\}.$$

If $a \in \mathbb{R}$, we identify $a + 0i$ with the real number a. Thus we can think of $\mathbb{R}$ as a subset of $\mathbb{C}$. Addition and multiplication on $\mathbb{C}$ are defined by

$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

where $a, b, c, d \in \mathbb{R}$.

$\mathbb{C}$ satisfies the following properties:

commutativity

$w + z = z + w$ and $wz = zw$ for all $w, z \in \mathbb{C}$;

associativity

$(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$ and $(z_1 z_2) z_3 = z_1 (z_2 z_3)$ for all $z_1, z_2, z_3 \in \mathbb{C}$;

identities

$z + 0 = z$ and $z1 = z$ for all $z \in \mathbb{C}$;

additive inverse

for every $z \in \mathbb{C}$, there exists a unique $w \in \mathbb{C}$ such that $z + w = 0$;

multiplicative inverse

for every $z \in \mathbb{C}$ with $z \neq 0$, there exists a unique $w \in \mathbb{C}$ such that $zw = 1$;

distributive property

$\lambda(w + z) = \lambda w + \lambda z$ for all $\lambda, w, z \in \mathbb{C}$.

For $z \in \mathbb{C}$, we let $-z$ denote the additive inverse of z. Thus $-z$ is the unique complex number such that

$$z + (-z) = 0.$$

Subtraction on $\mathbb{C}$ is defined by

$$w - z = w + (-z)$$

for $w, z \in \mathbb{C}$. For $z \in C$ with $z \neq 0$, we let $1/z$ denote the multiplicative inverse of z. Thus $1/z$ is the unique complex number such that

$$z\left(\frac{1}{z}\right) = 1.$$

Division on $\mathbb{C}$ is defined by

$$\frac{w}{z} = w\left(\frac{1}{z}\right)$$

for $w, z \in \mathbb{C}$ with $z \neq 0$.

The most common way to formalize this is by defining a field as a set together with two operations, usually called addition and multiplication, and denoted by $+$ and $\cdot$, respectively, such that the following axioms hold; subtraction and division are defined implicitly in terms of the inverse operations of addition and multiplication:

Closure of F under addition and multiplication

For all $a, b \in F$, both $a + b$ and $a \cdot b$ are in F.

Associativity of addition and multiplication

For all a, b, and c in F, the following equalities hold:
$a + (b + c) = (a + b) + c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

Commutativity of addition and multiplication

For all a and b in F, the following equalities hold: $a + b = b + a$ and $a \cdot b = b \cdot a$.

Additive and multiplicative identity
There exists an element of F, called the additive identity element and denoted by 0, such that for all a in F, $a + 0 = a$. Likewise, there is an element, called the multiplicative identity element and denoted by 1, such that for all a in F, $a \cdot 1 = a$. For technical reasons, the additive identity and the multiplicative identity are required to be distinct.
Additive and multiplicative inverses
For every a in F, there exists an element $-a$ in F, such that $a + (-a) = 0$. Similarly, for any a in F other than 0, there exists an element a^{-1} in F, such that $a \cdot a^{-1} = 1$. (The elements $a + (-b)$ and $a \cdot b^{-1}$ are also denoted $a - b$ and $\frac{a}{b}$, respectively.) In other words, subtraction and division operations exist.
Distributivity of multiplication over addition
For all a, b and c in F, the following equality holds:$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$

Example 1 (A field with four elements)

+	O	I	A	B
O	O	I	A	B
I	I	O	B	A
A	A	B	O	I
B	B	A	I	O

·	O	I	A	B
O	O	O	O	O
I	O	I	A	B
A	O	A	B	I
B	O	B	I	A

Why must be the additive and multiplicative identities in a field be different?

If F is any field for which

$$1_F = 0_F$$

we have, for any $x \in F$

$$x = x1_F = x0_F = 0_F$$

so F has only one element, 0_F; if we want to avoid the trivial case

$$F = \{0_F\},$$

we must assume that

$$1_F \neq 0_F.$$

How do we know that $x0_F = 0_F$ for any $x \in F$

We have:

$$0_F = 0_F + 0_F,$$

since 0_F is the additive identity. Then

$$x0_F = x(0_F + 0_F) = x0_F + x0_F,$$

by the distributive law.
Add additive inverse of $x0_F$ to both sides:

$$x0_F + (-x0_F) = x0_F + x0_F + (-x0_F),$$

$$0_F = x0_F + (x0_F + (-x0_F)),$$

$$0_F = x0_F + 0_F,$$

Thus, we obtain

$$x0_F = 0_F.$$

Vector Spaces

The vector space $\mathbb{R}^2$, which one can think of as a plane, consists of all ordered pairs of real numbers:

$$\mathbb{R}^2 = \{(x, y) : \ x, y \in R\}.$$

The vector space $\mathbb{R}^3$, which one can think of as ordinary space, consists of all ordered triples of real numbers:

$$\mathbb{R}^3 = \{(x, y, z) : \ x, y, z \in R\}.$$

To generalize $\mathbb{R}^2$ and $\mathbb{R}^3$ to higher dimensions, suppose n is a nonnegative integer. A list of length n is an ordered collection of n objects (which might be numbers, other lists, or more abstract entities) separated by commas and surrounded by parentheses. Many mathematicians call a list of length n an n-tuple. $(x_1, \ldots, x_n)$.

Thus a list of length 2 is an ordered pair and a list of length 3 is an ordered triple. For $j \in \{1, \ldots, n\}$, we say that x_j is the j th coordinate of the list above. Thus x_1 is called the first coordinate, x_2 is called the second coordinate, and so on.

Sometimes we will use the word list without specifying its length. Remember, however, that by definition each list has a finite length that is a nonnegative integer, so that an object that looks like

$$(x_1, x_2, \ldots),$$

which might be said to have infinite length, is not a list. A list of length 0 looks like this: (). We consider such an object to be a list so that some of our theorems will not have trivial exceptions.

We define F^n to be the set of all lists of length n consisting of elements of F:

$$F^n = \{(x_1, \ldots, x_n) : x_j \in F \text{ for } j = 1, \ldots, n\}.$$

As another example, $\mathbb{C}^4$ is the set of all lists of four complex numbers:

$$\mathbb{C}^4 = \{(z_1, z_2, z_3, z_4) : z_1, z_2, z_3, z_4 \in \mathbb{C}\}.$$

Usual addition is defined on F^n by adding corresponding coordinates:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n).$$

Often the mathematics of F^n becomes cleaner if we use a single entity to denote a list of n numbers, without explicitly writing the coordinates. Thus the commutative property of addition on F^n should be expressed as

$$x + y = y + x$$

for all $x, y \in F^n$, rather than the more cumbersome

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (y_1, \ldots, y_n) + (x_1, \ldots, x_n)$$

Usual scalar multiplication for $a \in F$ and $(x_1, \ldots, x_n) \in F^n$ is defined as

$$a(x_1, \ldots, x_n) = (ax_1, \ldots, ax_n)$$

Vector Space

We define a vector space to be a set V along with an addition and a scalar multiplication on V that satisfy the properties below. By an addition on V we mean a function that assigns an element $u + v \in V$ to each pair of elements $u, v \in V$. By a scalar multiplication on V we mean a function that assigns an element $av \in V$ to each $a \in F$ and each $v \in V$.
Formally, a vector space is a set V along with an addition on V and a scalar multiplication on V such that the following properties hold:
commutativity

$$u + v = v + u \text{ for all } u, v \in V;$$

associativity

$$(u+v)+w = u+(v+w) \text{ and } (ab)v = a(bv) \text{ for all } u,v,w \in V$$
$$\text{and all } a,b \in F;$$

additive identity

$$\text{there exists an element } 0 \in V \text{such that } v+0 = v \text{ for all } v \in V;$$

additive inverse

$$\text{for every } v \in V, \text{ there exists } w \in V \text{ such that } v+w = 0;$$

multiplicative identity

$$1v = v \text{ for all } v \in V;$$

distributive properties

$$a(u+v) = au+av \text{ and } (a+b)u = au+bu \text{ for all } a,b \in F \text{ and all } u,v \in V$$

1.2 Proposition A vector space has a unique additive identity.
Proof: Suppose 0 and $0'$ are both additive identities for some vector space V. Then

$$0' = 0' + 0 = 0,$$

where the first equality holds because 0 is an additive identity and the second equality holds because $0'$ is an additive identity. Thus $0' = 0$, proving that V has only one additive identity. □

1.3 Proposition Every element in a vector space has a unique additive inverse.

Proof: Suppose V is a vector space. Let $v \in V$. Suppose that w and w' are additive inverses of v. Then

$$w = w + 0 = w + (v + w') = (w + v) + w' = 0 + w' = w'.$$

Thus $w = w'$, as desired. $\square$

1.4 Proposition $0v = 0$ for every $v \in V$.
Proof: For $v \in V$, we have

$$0v = (0 + 0)v = 0v + 0v.$$

Adding the additive inverse of 0v to both sides of the equation above gives $0 = 0v$, as desired. $\square$

1.5 Proposition $a0 = 0$ for every $a \in F$.
Proof: For $a \in F$, we have

$$a0 = a(0+0) = a0 + a0.$$

Adding the additive inverse of $a0$ to both sides of the equation above gives $0 = a0$, as desired. $\square$

1.6 Proposition $(-1)v = -v$ for every $v \in V$.

Proof: For $v \in V$, we have

$$v + (-1)v = 1v + (-1)v = (1 + (-1))\, v = 0v = 0.$$

This equation says that $(-1)v$, when added to v, gives 0. Thus $(-1)v$ must be the additive inverse of v, as desired. $\square$

Example 2

The set $\mathbb{R}^3$ with the field $\mathbb{R}$, under the usual addition of vectors and usual multiplication of vectors by scalars form a vector space.

Let $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \in \mathbb{R}^3$. Their sum

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{bmatrix}$$

is in $\mathbb{R}^3$. This shows that the set is closed under usual addition. For any $k \in \mathbb{R}$

$$k \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} ka_1 \\ ka_2 \\ ka_3 \end{bmatrix}$$

is in $\mathbb{R}^3$. This shows that the set is closed under scalar muliplication.

Example 2 (cont.)

Obviously, the commutativity is satisfied:

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

for all $a_1, a_2, a_3, b_1, b_2, b_3 \in \mathbb{R}$. It is straightforward to show that associativity holds.

Additive identity is $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

For $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \in \mathbb{R}^3$, corresponding additive inverse is $\begin{bmatrix} -a_1 \\ -a_2 \\ -a_3 \end{bmatrix}$

Example 2 (cont.)

$1 \in \mathbb{R}$ works as multiplicative identity:

$$1 \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \cdot a_1 \\ 1 \cdot a_2 \\ 1 \cdot a_3 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \text{for all } a_1, a_2, a_3 \in \mathbb{R}.$$

Satisfaction of the distributive properties is shown in a straightforward way.

Since all the axioms are satisfied, the vector set $\mathbb{R}^3$, the field $\mathbb{R}$, usual addition of vectors, and usual multiplication of vectors by scalars form a vector space.

This example can be generalized to the vector sets $\mathbb{R}^n$, for any natural number n.

Example 3

The set of 2×2 matrices with real entries, the field $\mathbb{R}$, under the usual addition of matrices and usual multiplication of matrices by scalars form a vector space.

Example 4

The set of real-valued continuous functions in the interval $[0, 1]$, the field $\mathbb{R}$, under the usual addition of functions and usual multiplication of functions by scalars form a vector space.

Example 5

The set of all polynomials with real coefficients, the field $\mathbb{R}$, under the usual addition of polynomials and usual multiplication of polynomials by scalars form a vector space.

Subspaces

A subset U of V is called a subspace of V if U is also a vector space (using the same addition and scalar multiplication as on V). For example,

$$\{(x_1, x_2, 0) : x_1, x_2 \in F\}$$

is a subspace of F^3.
If U is a subset of V, then to check that U is a subspace of V we need only check that U satisfies the following:

Subspace Test

additive identity
$0 \in U$
closed under addition
$u, v \in U$ implies $u + v \in U$;
closed under scalar multiplication
$a \in F$ and $u \in U$ implies $au \in U$.

Example 6

Vector space $\mathbb{R}^2$ has the following subspaces: Zero vector: $\{0\}$, all the straight lines passing through the origin, $\mathbb{R}^2$ itself.
In particular, consider a line passing through the origin. 0 is on it. Scalar multiple of any point on the line is also on the line. Sum of any two vectors on the line is also on the line. So, any line passing through the origin passes the subspace test.

Example 7

Vector space $\mathbb{R}^3$ has the following subspaces: Zero vector: $\{0\}$, all the straight lines passing through the origin, all the planes passing through the origin, $\mathbb{R}^3$ itself.

Example 8

Set of polynomials with real coefficients having degree less than or equal to 5 is a subspace of the vector space of all polynomials with real coefficients, under usual polynomial addition and usual multiplication of polynomial by a scalar.

Example 9

Set of polynomials with real coefficients having degree equal to 5, call it U, is not a subspace of the vector space of all polynomials with real coefficients, under usual polynomial addition and usual multiplication of polynomial by a scalar.
To show this, consider the polynomials $p_1(s) = s^5 + s^4 + 2$ and $p_2(s) = -s^5 + s^2 + 3$ in U as a counterexample. Their sum $s^4 + s^2 + 5$ is not in U. Since U is not closed under vector addition, it cannot be a subspace.

Sums and Direct Sums

Suppose $U_1, \ldots, U_m$ are subspaces of V. The sum of $U_1, \ldots, U_m$, denoted $U_1 + \cdots + U_m$, is defined to be the set of all possible sums of elements of $U_1, \ldots, U_m$. More precisely,

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : \ u_1 \in U_1, \ldots, u_m \in U_m\}.$$

Note that if $U_1, \ldots, U_m$ are subspaces of V, then the sum $U_1 + \cdots + U_m$ is a subspace of V.

Example 10

Suppose U is the set of all elements of F^3 whose second and third coordinates equal 0, and W is the set of all elements of F^3 whose first and third coordinates equal 0:

$$U = \{(x, 0, 0) \in F^3 : x \in F\} \text{ and } W = \{(0, y, 0) \in F^3 : y \in F\}.$$

Then $U + W = \{(x, y, 0) : x, y \in F\}$.

Example 11

Let

$$U = \{(x, 0, 0) \in F^3 : x \in F\} \text{ and } W = \{(y, y, 0) \in F^3 : y \in F\}.$$

Then $U + W = \{(x, y, 0) : x, y \in F\}$.

Definition Let V be a vector space, $U \subset V$ a subset. Smallest subspace of V that contains U is the intersection of all U-containing subspaces of V.

Suppose $U_1, \ldots, U_m$ are subspaces of V. Clearly $U_1, \ldots, U_m$ are all contained in $U_1 + \cdots + U_m$. Conversely, any subspace of V containing $U_1, \ldots, U_m$ must contain $U_1 + \cdots + U_m$. Thus $U_1 + \cdots + U_m$ is the smallest subspace of V containing $U_1, \ldots, U_m$.

Suppose $U_1, \ldots, U_m$ are subspaces of V such that $V = U_1 + \cdots + U_m$. Thus every element of V can be written in the form

$$u_1 + \cdots + u_m,$$

where each $u_j \in U_j$. Consider cases where each vector in V can be uniquely represented in the form above. We give it a special name: **direct sum**.

Specifically, we say that V is the **direct sum** of subspaces $U_1, \ldots, U_m$, written

$$V = U_1 \oplus \cdots \oplus U_m,$$

if each element of V can be written uniquely as a sum $u_1 + \cdots + u_m$, where each $u_j \in U_j$.

Example 12

Suppose U is the subspace of F^3 consisting of those vectors whose last coordinate equals 0, and W is the subspace of F^3 consisting of those vectors whose first two coordinates equal 0:

$$U = \{(x, y, 0) \in F^3 : \; x, y \in F\} \text{ and } W = \{(0, 0, z) \in F^3 : z \in F\}.$$

Then $F^3 = U \oplus W$.

Example 13

Consider the following three subspaces of F^3:
$U_1 = \{(x, y, 0) \in F^3 : \; x, y \in F\}$;
$U_2 = \{(0, 0, z) \in F^3 : \; z \in F\}$;
$U_3 = \{(0, y, y) \in F^3 : \; y \in F\}$.
Clearly $F^3 = U_1 + U_2 + U_3$ because an arbitrary vector $(x, y, z) \in F^3$ can be written as

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

where the first vector on the right side is in U_1, the second vector is in U_2, and the third vector is in U_3. However, F^3 does not equal the direct sum of U_1, U_2, U_3 because the vector (x, y, z) can be expressed also as

$$(x, y, z) = (x, y - z, 0) + (0, 0, 0) + (0, z, z),$$

$U_1 = \{(x, y, 0) \in F^3 : \ x, y \in F\};$
$U_2 = \{(0, 0, z) \in F^3 : \ z \in F\};$
$U_3 = \{(0, y, y) \in F^3 : \ y \in F\}.$

Example 13 (cont.)

Two different ways of writing (x, y, z)

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

$$(x, y, z) = (x, y - z, 0) + (0, 0, 0) + (0, z, z),$$

A numerical example:

$$(4, 5, 6) = (4, 5, 0) + (0, 0, 6) + (0, 0, 0),$$

$$(4, 5, 6) = (4, -1, 0) + (0, 0, 0) + (0, 6, 6),$$

$U_1 = \{(x, y, 0) \in F^3 : \; x, y \in F\};$
$U_2 = \{(0, 0, z) \in F^3 : \; z \in F\};$
$U_3 = \{(0, y, y) \in F^3 : \; y \in F\}.$

Example 13 (cont.)

According to a theorem, if $(0, 0, 0)$ could be written in two different ways, then so could (x, y, z). Using this, F^3 does not equal the direct sum of U_1, U_2, U_3 because the vector $(0, 0, 0)$ can be written in two different ways as a sum $u_1 + u_2 + u_3$, with each $u_j \in U_j$. Specifically, we have

$$(0, 0, 0) = (0, 1, 0) + (0, 0, 1) + (0, -1, -1)$$

and, of course,

$$(0, 0, 0) = (0, 0, 0) + (0, 0, 0) + (0, 0, 0),$$

where the first vector on the right side of each equation above is in U_1, the second vector is in U_2, and the third vector is in U_3.

1.8 Proposition Suppose that $U_1, \ldots, U_n$ are subspaces of V. Then $V = U_1 \oplus \cdots \oplus U_n$ if and only if both the following conditions hold:
(a) $V = U_1 + \cdots + U_n$;
(b) the only way to write 0 as a sum $u_1 + \cdots + u_n$, where each $u_j \in U_j$, is by taking all the u_j's equal to 0.

Proof First suppose that $V = U_1 \oplus \cdots \oplus U_n$. Clearly (a) holds (because of how sum and direct sum are defined). To prove (b), suppose that $u_1 \in U_1, \cdots, u_n \in U_n$ and

$$0 = u_1 + \cdots + u_n.$$

Then each u_j must be 0 (this follows from the uniqueness part of the definition of direct sum because $0 = 0 + \cdots + 0$ and $0 \in U_1, \ldots, 0 \in U_n$), proving (b).

Now suppose that (a) and (b) hold. Let $v \in V$. By (a), we can write

$$v = u_1 + \cdots + u_n$$

for some $u_1 \in U_1, \ldots, u_n \in U_n$. To show that this representation is unique, suppose that we also have

$$v = v_1 + \cdots + v_n,$$

where $v_1 \in U_1, \ldots, v_n \in U_n$. Subtracting these two equations, we have

$$0 = (u_1 - v_1) + \cdots + (u_n - v_n).$$

Clearly $u_1 - v_1 \in U_1, \ldots, u_n - v_n \in U_n$, so the equation above and (b) imply that each $u_j - v_j = 0$. Thus $u_1 = v_1, \ldots, u_n = v_n$, as desired. $\square$

1.9 Proposition Suppose that U and W are subspaces of V. Then $V = U \oplus W$ if and only if $V = U + W$ and $U \cap W = \{0\}$. $\square$

Span and Linear Independence

A **linear combination** of a list $(v_1, \dots, v_m)$ of vectors in V is a vector of the form

$$a_1 v_1 + \cdots + a_m v_m,$$

where $a_1, \dots, a_m \in F$. The set of all linear combinations of $(v_1, \dots, v_m)$ is called the **span** of $(v_1, \dots, v_m)$, denoted $\text{span}(v_1, \dots, v_m)$. In other words,

$$\text{span}(v_1, \dots, v_m) = \{a_1 v_1 + \cdots + a_m v_m : \ a_1, \dots, a_m \in F\}.$$

If span $(v_1, \dots, v_m)$ equals V, we say that $(v_1, \dots, v_m)$ spans V. A vector space is called **finite dimensional** if some list of vectors in it spans the space. For example, F^n is finite dimensional because

$$((1, 0, ..., 0), (0, 1, 0, ..., 0), ..., (0, ..., 0, 1))$$

spans F^n.

Example 14

Consider the list $((1,0,0),(0,1,0))$ in $\mathbb{R}^3$.
$(6,2,0)$ is a linear combination of the vectors in the list.
$(2,1,3)$ is not a linear combination of the vectors in the list.
Span of the list, geometrically, is all the points on the xy plane of the xyz coordinate system.

Example 15

Consider the list $((1,1,1),(2,1,4))$ in $\mathbb{R}^3$.
$(3,2,5)$ is in the span of the list.
$(3,3,0)$ is not in the span of the list.

Example 16

Consider the list $((1,1,1),(2,1,4),(0,1,1))$ in $\mathbb{R}^3$. Span of the list equals R^3.

A vector space that is not finite dimensional is called **infinite dimensional**.

Example 17

$P(F)$, the set ol all polynomials with coefficients in F is infinite dimensional. In a set notation, it is

$$P(F) = \{a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_mx^m : m \in N;\ a_0, a_1, a_2, a_3, \ldots a_m \in F\}$$

It is not possible to form a list using vectors in $P(F)$ that spans $P(F)$.

Example 18

The vector space F^∞, consisting of all sequences of elements of F, is also infinite dimensional. In a set notation, it is

$$F^\infty = \{(x_1, x_2, x_3, \ldots) : x_1, x_2, x_3, \ldots \in F\}$$

It is not possible to form a list using vectors in F^∞ that spans F^∞.

Example 19

Real valued continuous functions on $[a, b]$ form a vector space with respect to usual addition and multiplication by scalars. There is no list of vectors in this space which spans the space. Thus, it is an infinite dimensional vector space.

Example 20

$\mathbb{R}^4$ is a finite dimensional vector space. For instance, the list

$$\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right)$$

spans the vector space $\mathbb{R}^4$.

Suppose $v_1, \ldots, v_m \in V$ and $v \in span(v_1, \ldots, v_m)$. By the definition of span, there exist $a_1, \ldots, a_m \in F$ such that

$$v = a_1 v_1 + \cdots + a_m v_m.$$

Consider the question of whether the choice of a's in the equation above is unique. Suppose $\hat{a}_1, \ldots, \hat{a}_m$ is another set of scalars such that

$$v = \hat{a}_1 v_1 + \cdots + \hat{a}_m v_m.$$

Subtracting the last two equations, we have

$$0 = (a_1 - \hat{a}_1) v_1 + \cdots + (a_m - \hat{a}_m) v_m.$$

Thus we have written 0 as a linear combination of $(v_1, \ldots, v_m)$. If the only way to do this is the obvious way (using 0 for all scalars), then each $a_j - \hat{a}_j$ equals 0, which means that each a_j equals $\hat{a}_j$ (and thus the choice of a's was indeed unique)

A list $(v_1, \ldots, v_m)$ of vectors in V is called **linearly independent** if the only choice of $a_1, \ldots, a_m \in F$ that makes $a_1 v_1 + \cdots + a_m v_m$ equal 0 is $a_1 = \cdots = a_m = 0$. For example,

$$((1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0))$$

is linearly independent in F^4.

A list of vectors in V is called **linearly dependent** if it is not linearly independent. In other words, a list $(v_1, \ldots, v_m)$ of vectors in V is linearly dependent if there exist $a_1, \ldots, a_m \in F$, not all 0, such that $a_1 v_1 + \cdots + a_m v_m = 0$.

Example 21

Consider the list $((1,1,0),(0,2,4))$ in $\mathbb{R}^3$. Linear independence equation

$$c_1(1,1,0) + c_2(0,2,4) = (0,0,0)$$

reveals that the list is linearly independent.

Example 22

Consider the list $((1,1,0),(0,2,4),(1,3,4))$ in $\mathbb{R}^3$. Linear independence equation

$$c_1(1,1,0) + c_2(0,2,4) + c_3(1,3,4) = (0,0,0)$$

has nontrivial solutions (for example $c_1 = 1, c_2 = 1, c_3 = -1$), therefore, the list is linearly dependent.

2.4 Linear Dependence Lemma If $(v_1, \ldots, v_m)$ is linearly dependent in V and $v_1 \neq 0$, then there exists $j \in \{2, \ldots, m\}$ such that the following hold:
(a) $v_j \in span(v_1, \ldots, v_{j-1})$;
(b) if the $j-$th term is removed from $(v_1, \ldots, v_m)$, the span of the remaining list equals span$(v_1, \ldots, v_m)$.
Proof Suppose $(v_1, \ldots, v_m)$ is linearly dependent in V and $v_1 \neq 0$. Then there exist $a_1, \ldots, a_m \in F$, not all 0, such that

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Not all of $a_{2,3}, \ldots, a_m$ can be 0 (because $v_1 \neq 0$). Let j be the largest element of $\{2, \ldots, m\}$ such that $a_j \neq 0$. Then

$$v_j = -\frac{a_1}{a_j} v_1 - \cdots - \frac{a_{j-1}}{a_j} v_{j-1}, \tag{1}$$

proving (a).

To prove (b), suppose that $u \in span(v_1, \ldots, v_m)$. Then there exist $c_1, \ldots, c_m \in F$ such that

$$u = c_1 v_1 + \cdots + c_m v_m.$$

In the equation above, we can replace v_j with the right side of (1), which shows that u is in the span of the list obtained by removing the $j-$th term from $(v_1, \ldots, v_m)$. Thus (b) holds. $\square$

Example 23

Consider the list of vectors

$$L \triangleq \left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 5 \\ 8 \\ 1 \end{bmatrix} \right)$$

The third vector is a linear combination of the preceding two. According to the linear dependence lemma, we conclude that the list above is linearly dependent.

According to the lemma

$$\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 5 \\ 8 \\ 1 \end{bmatrix} \right)$$

also spans L.

Theorem 1

In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

Proof Suppose that $(u_1, \ldots, u_m)$ is linearly independent in V and that $(w_1, \ldots, w_n)$ spans V. We need to prove that $m \leq n$. We do so through the multistep process described below; note that in each step we add one of the u's and remove one of the w's.

Step 1

The list $(w_1, \ldots, w_n)$ spans V, and thus adjoining any vector to it produces a linearly dependent list. In particular, the list $(u_1, w_1, \ldots, w_n)$ is linearly dependent. Thus by the linear dependence Lemma 2.4, we can remove one of the w's so that the list B (of length n) consisting of u_1 and the remaining w's spans V.

Step j

The list B (of length n) from step $j-1$ spans V, and thus adjoining any vector to it produces a linearly dependent list. In particular, the list of length $(n+1)$ obtained by adjoining u_j to B, placing it just after $u_1, \ldots, u_{j-1}$, is linearly dependent. By the linear dependence Lemma 2.4, one of the vectors in this list is in the span of the previous ones, and because $(u_1, \ldots, u_j)$ is linearly independent, this vector must be one of the w's, not one of the u's. We can remove that w from B so that the new list B (of length n) consisting of $u_1, \ldots, u_j$ and the remaining w's spans V. After step m, we have added all the u's and the process stops. If at any step we added a u and had no more w's to remove, then we would have a contradiction. Thus there must be at least as many w's as u's. □

2.7 Proposition Every subspace of a finite-dimensional vector space is finite dimensional. □

Bases

A **basis** of V is a list of vectors in V that is linearly independent and spans V. For example,

$$((1,0,...,0),(0,1,0,...,0),...,(0,...,0,1))$$

is a basis of F^n, called the standard basis of F^n.

2.8 Proposition A list $(v_1,\ldots,v_n)$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form

$$v = a_1 v_1 + \cdots + a_n v_n, \tag{2}$$

where $a_1,\ldots,a_n \in F$.

Proof First suppose that $(v_1,\ldots,v_n)$ is a basis of V. Let $v \in V$. Because $(v_1,\ldots,v_n)$ spans V, there exist $a_1,\ldots,a_n \in F$ such that (2) holds. To show that the representation in (2) is unique, suppose that $b_1,\ldots,b_n$ are scalars so that we also have

$$v = b_1 v_1 + \cdots + b_n v_n.$$

Subtracting the last equation from (2), we get

$$0 = (a_1 - b_1)v_1 + \cdots + (a_n - b_n)v_n.$$

This implies that each $a_j - b_j = 0$ (because $(v_1, \ldots, v_n)$ is linearly independent) and hence $a_1 = b_1, \ldots, a_n = b_n$. We have the desired uniqueness, completing the proof in one direction.
For the other direction, suppose that every $v \in V$ can be written uniquely in the form given by (2). Clearly this implies that $(v_1, \ldots, v_n)$ spans V. To show that $(v_1, \ldots, v_n)$ is linearly independent, suppose that $a_1, \ldots, a_n \in F$ are such that

$$0 = a_1 v_1 + \cdots + a_n v_n.$$

The uniqueness of the representation (2) (with $v = 0$) implies that $a_1 = \cdots = a_n = 0$. Thus $(v_1, \ldots, v_n)$ is linearly independent and hence is a basis of V. $\square$

Example 24

$((1,1,1),(1,1,0),(1,0,0))$ is a basis for $\mathbb{R}^3$. Every element in $\mathbb{R}^3$ is a unique linear combination of this list.

Example 25

$((1,1,1),(1,1,0),(1,0,0),(3,2,1))$ is not a basis for $\mathbb{R}^3$. Even though this list spans $\mathbb{R}^3$, it is not linearly independent. Every element of $\mathbb{R}^3$ is expressed as many different linear combinations of this list.

Theorem 2

Every spanning list in a vector space can be reduced to a basis of the vector space.

2.11 Corollary Every finite-dimensional vector space has a basis.

Theorem 3

Every linearly independent list of vectors in a finite dimensional vector space can be extended to a basis of the vector space.

2.13 Proposition Suppose V is finite dimensional and U is a subspace of V. Then there is a subspace W of V such that $V = U \oplus W$.

Dimension

Theorem 4

Any two bases of a finite-dimensional vector space have the same length.

The **dimension** of a finite-dimensional vector space is defined to be the length of any basis of the vector space.

2.15 Proposition If V is finite dimensional and U is a subspace of V, then $\dim U \leq \dim V$.

2.16 Proposition If V is finite dimensional, then every spanning list of vectors in V with length $\dim V$ is a basis of V.

2.17 Proposition If V is finite dimensional, then every linearly independent list of vectors in V with length $\dim V$ is a basis of V.

Theorem 5

If U_1 and U_2 are subspaces of a finite-dimensional vector space, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Example 26

Let U_1 and U_2 be subspaces of $\mathbb{R}^4$ defined by

$$U_1 = span\left(\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}\right), \quad U_2 = span\left(\begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}\right)$$

$$U_1 \cap U_2 = span\left(\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\right)$$

dim U_1=2, dim U_2=2, dim $U_1 \cap U_2$=1. Applying the theorem:

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2),$$

$$\dim(U_1 + U_2) = 2 + 2 - 1 = 3.$$

A Remark

A vector space is a set V together with operations $+ : V \times V \to V$ and $\cdot : F \times V \to V$ that satisfy certain conditions, where F is a field.

One may notice that these conditions makes $(V, +)$ into an Abelian group (i.e., commutative group). This means that if you take V and remove the scalar multiplication operator, the elements of V forms a group and commute with each others.

A vector space structure could be extended to more involved structures. For instance, a vector space that has a notion of size, or the norm, of objects in that space is called **normed space**.
A norm $||\cdot||$ is a function that has the following properties.
It returns nonnegative values, and only the zero vector has zero norm:

$$||u|| \geq 0 \;\; \forall u \in V, \;\; ||u|| = 0 \rightarrow u = 0$$

The norm of a vector multiplied by a scalar is itself scaled:

$$\alpha u|| = |\alpha|||u||, \forall \alpha \in F, \;\; u \in V$$

The triangle inequality holds:

$$||u + v|| \leq ||u|| + ||v|| \; \forall u, v \in V$$

A normed vector space is also a metric space. There is a notion of distance between objects and it is just the norm of the difference of them:

$$d(u, v) = ||u - v|||$$

This also induces a topology so we can talk about concepts such as continuity and convergence.

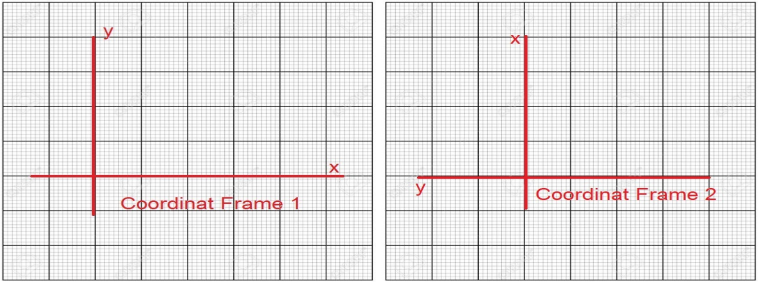
A **Banach space** is a vector space with a metric that allows the computation of vector length and distance between vectors and is complete in the sense that a Cauchy sequence of vectors always converges to a well-defined limit that is within the space.

A Cauchy sequence is a sequence whose terms become very close to each other as the sequence progresses. Formally, the sequence $\{a_n\}_{n=0}^{\infty}$ is a Cauchy sequence if, for every $\epsilon > 0$, there is an $N > 0$ such that

$$m, n > N \to ||a_n - a_m|| < \epsilon$$

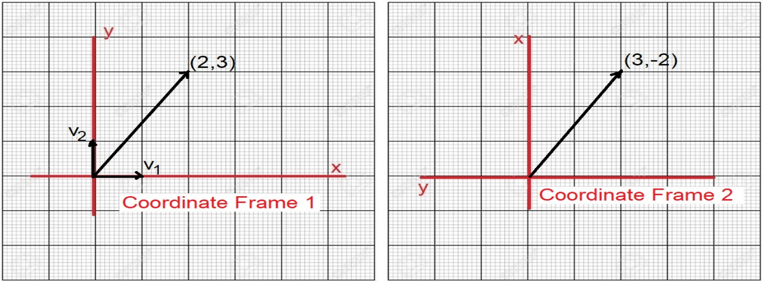
If a space has an inner product which generates the norm then it is an **inner product space**. If an inner product space is complete, which is the same thing as a Banach space with an inner product it is called a **Hilbert space**.

Consider two coordinate frames CF1 and CF2 in $\mathbb{R}^2$. CF2 is obtained by rotating CF1 by 90 degrees in the CCW direction.

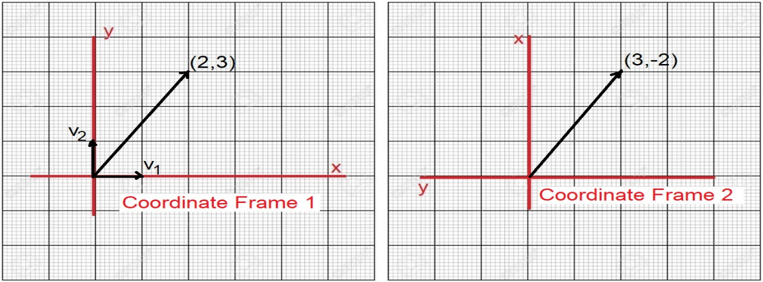


For the vectors $v_1 = (1, 0)$ and $v_2 = (0, 1)$ in CF1, we find their linear combination $2v_1 + 3v_2$, then express the result in CF2.

For the vectors $v_1 = (1, 0)$ and $v_2 = (0, 1)$ in CF1, we find their linear combination $2v_1 + 3v_2$, then express the result in CF2.



In CF2 we obtained the vector $(3, -2)$.



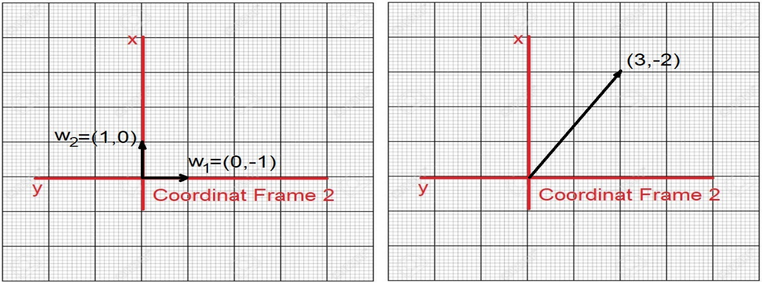
A recap:

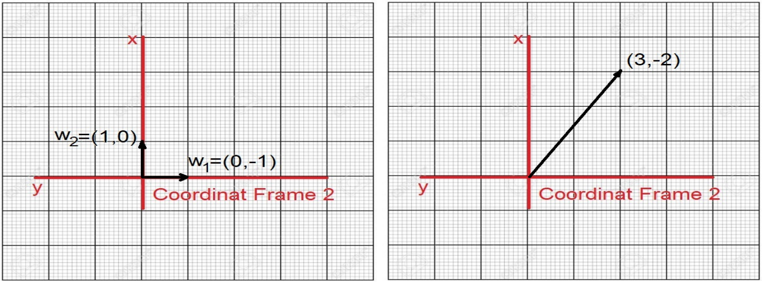
1) v_1 and v_2 are expressed in CF1: ($v_1 = (1, 0)$ and $v_2 = (0, 1)$)

2) Their linear combination $2v_1 + 3v_2$ in CF1 is $(2, 3)$.

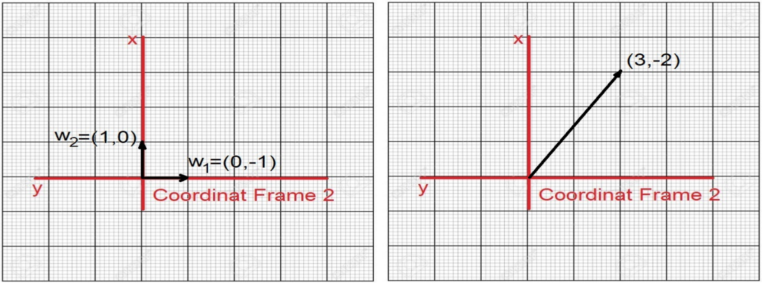
3) In CF2 it is $(3, -2)$

Let us expreess v_1 and v_2 in CF2. And do the linear combination operation there.





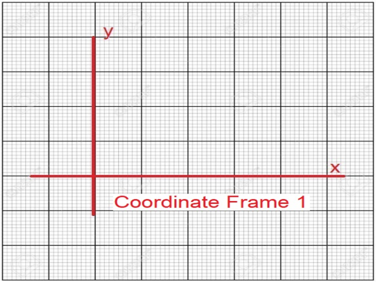
1) v_1 and v_2 in CF2 are $w_1 = (0, -1)$ and $w_2 = (1, 0)$)
2) Their linear combination $2w_1 + 3w_2$ in CF2 is $(3, -2)$.

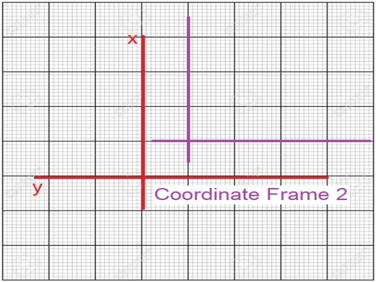


Linear combination followed by basis change EQUALS basis changes followed by linear combination.

Linear Combinations of Points

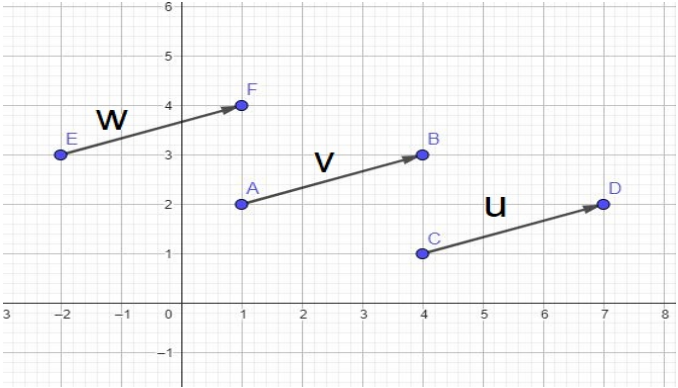
The result we obtained in the previous section is not valid if CF2 is obtained by translating CF1. Consider such coordinate frames presented below:





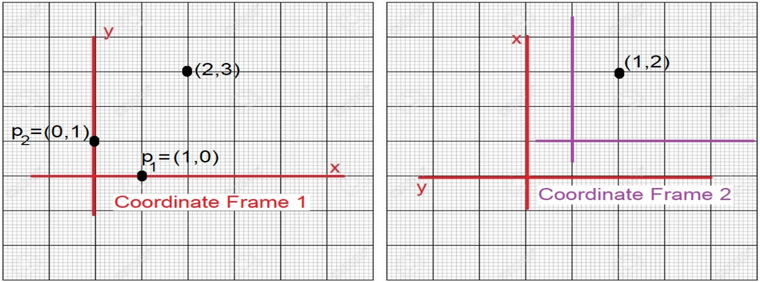
A digression

Vectors are identified with their direction and magnitude. In this sense, u, v, and w below are equivalent. When positions matter, it is more convenient to use points.

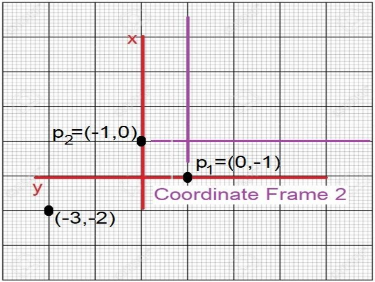


EOD

We have two points given: p_1 and p_2. Linear combination $2p_1 + 3p_2$ is shown on CF1 as $(2,3)$. This point is $(1,2)$ on CF2.

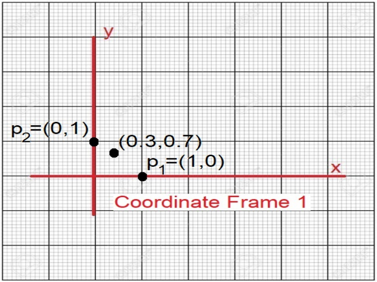


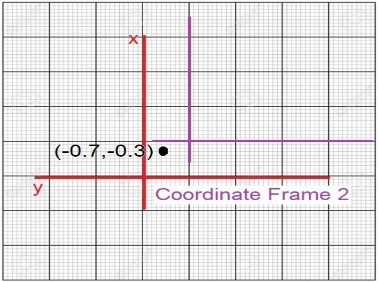
Points p_1 and p_2 are $(0,-1)$ and $(-1,0)$ on CF2. The same linear combination $2p_1 + 3p_2$ equals $(-3,-2)$.



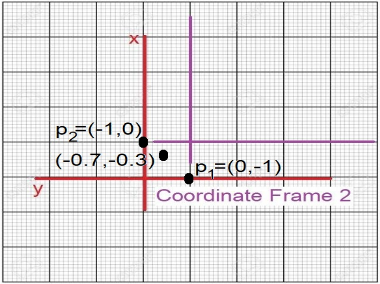
Linear combination followed by translation equals $(1,2)$. However, translations followed by linear combination equals $(-3,-2)$.

For the points p_1 and p_2, this time we have different weights 0.3 and 0.7. Linear combination $0.3p_1 + 0.7p_2$ is shown on CF1 as $(0.3, 0.7)$. This point is $(-0.7, -0.3)$ on CF2.





Points p_1 and p_2 are $(0, -1)$ and $(-1, 0)$ on CF2. The same linear combination $0.3p_1 + 0.7p_2$ equals $(-0.7, -0.3)$.



Linear combination followed by translation EQUALS translations followed by linear combination!!!

The reason for equality is that the sum of the weights equals 1.

Repeat this for different p_1 and p_2; and for different weights as long as their sum equals 1, reach the same result.

Sum of a Point and a Vector

We duplicate $\mathbb{R}^3$ into two copies: the first copy corresponding to points, where we forget the vector space structure, and the second copy corresponding to vectors, where the vector space structure is important. Definition. Given any point $a = (a_1, a_2, a_3)$ and any vector $v = (v_1, v_2, v_3)$, their sum results in a point:

$$a + v = (a_1 + v_1, a_2 + v_2, a_3 + v_3),$$

which can be thought of as the result of translating a to b using the vector v. This action $+ : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ satisfies

$$a + 0 = a,$$

$$(a + u) + v = a + (u + v),$$

and for any two points a, b, there is a unique vector $\overrightarrow{ab}$ such that

$$b = a + \overrightarrow{ab}$$

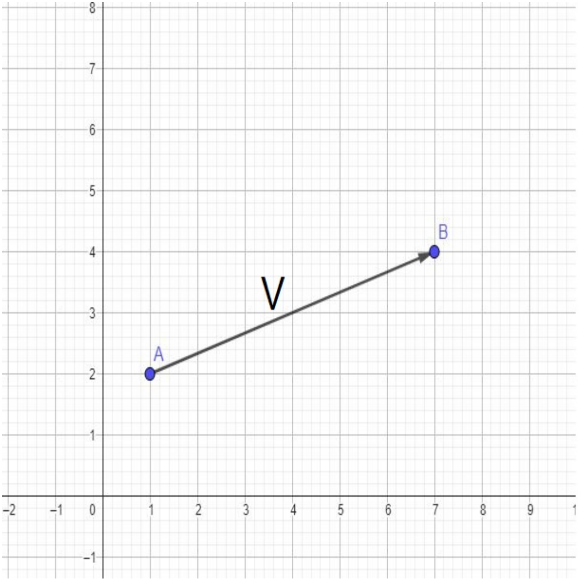


Figure 1: Sum of point A and vector V results point B

Affine Space

We consider two (distinct) sets E and $\overrightarrow{E}$, where E is a set of points (with no structure) and $\overrightarrow{E}$ is a vector space (of vectors) acting on the set E.

Definition An **affine space** is either the degenerate space reduced to the empty set, or a triple $< E, \overrightarrow{E}, + >$ consisting of a nonempty set E (of points), a vector space $\overrightarrow{E}$ (of translations, or vectors), and an action $+ : \; E \times \overrightarrow{E} \to E$, satisfying the following conditions.

(A1) $a + 0 = a$, for every $a \in E$.

(A2) $(a + u) + v = a + (u + v)$, for every $a \in E$, and every $u, v \in \overrightarrow{E}$.

(A3) For any two points $a, b \in E$, there is a unique $u \in \overrightarrow{E}$ such that $a + u = b$.

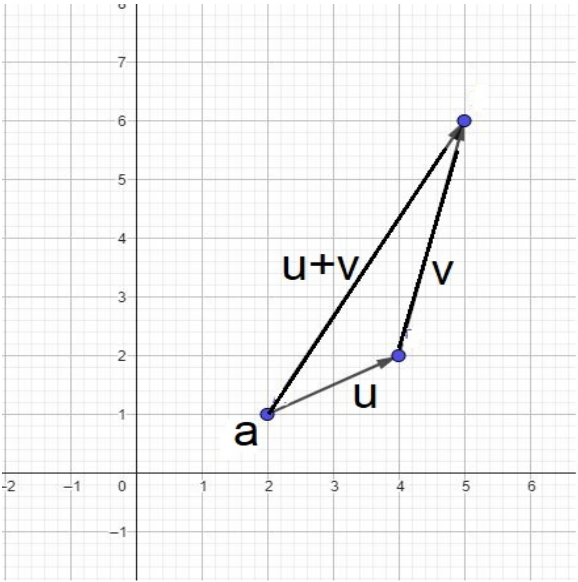


Figure 2: An example for $(a + u) + v = a + (u + v)$

The unique vector $u \in \overrightarrow{E}$ such that $a + u = b$ is denoted by $\overrightarrow{ab}$, or sometimes by $b - a$. We even write $b = a + (b - a)$.
The **dimension** of the affine space $< E, \overrightarrow{E}, + >$ is the dimension $\dim(\overrightarrow{E})$ of the vector space $\overrightarrow{E}$. It is denoted by $\dim(E)$.
Note that

$$\overrightarrow{a(a+v)} = v \tag{3}$$

for all $a \in E$ and all $v \in \overrightarrow{E}$. Adding a to both sides of (3), we have

$$a + \overrightarrow{a(a+v)} = a + v$$

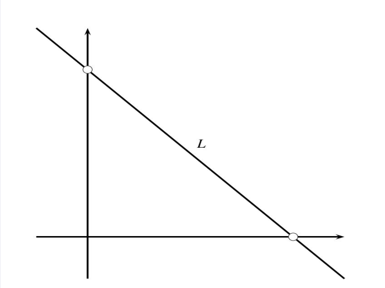
Because $a + v$ is obtained by a unique v, we conclude $\overrightarrow{a(a+v)} = v$.
Notation

$$A^n := < \mathbb{R}^n, \mathbb{R}^n, + >$$

Example 27

Consider the subset L of A^2 consisting of all points (x, y) satisfying

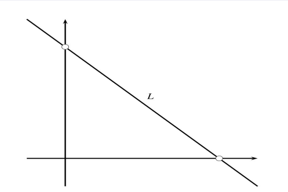
$$x + y - 1 = 0$$



Define the action $+ : L \times \mathbb{R} \to L$ of $\mathbb{R}$ on L defined such that for every point $(x, 1 - x)$ on L and any $u \in \mathbb{R}$,

$$(x, 1 - x) + u = (x + u, 1 - x - u)$$

Example 27 (cont.)



Define the action $+ : L \times \mathbb{R} \to L$ of $\mathbb{R}$ on L defined such that for every point $(x, 1-x)$ on L and any $u \in \mathbb{R}$,

$$(x, 1-x) + u = (x+u, 1-x-u)$$

This action makes L into an affine space. For example, for any two points $a = (a_1, 1-a_1)$ and $b = (b_1, 1-b_1)$ on L, the unique (vector) $u \in \mathbb{R}$ such that $b = a + u$ is $u = b_1 - a_1$.

Example 28

Consider the subset H of A^3 consisting of all points (x, y, z) satisfying the equation $x + y + z - 1 = 0$. The set H is the plane passing through the points $(1, 0, 0), (0, 1, 0)$, and $(0, 0, 1)$. The plane H can be made into an affine space by defining the action $+ : H \times \mathbb{R}^2 \to H$ of $\mathbb{R}^2$ on H defined such that for every point $(x, y, 1 - x - y)$ on H and any $\begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2$

$$(x, y, 1 - x - y) + \begin{pmatrix} u \\ v \end{pmatrix} = (x + u, y + v, 1 - x - u - y - v).$$

Example 29

Consider the subset P of A^3 consisting of all points (x, y, z) satisfying the equation

$$x^2 + y^2 - z = 0.$$

The set P is a paraboloid of revolution, with axis Oz. The surface P can be made into an affine space by defining the action $+ : P \times \mathbb{R}^2 \to P$ of $\mathbb{R}^2$ on P defined such that for every point $(x, y, x^2 + y^2)$ on P and any $\begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2$

$$(x, y, x^2 + y^2) + \begin{pmatrix} u \\ v \end{pmatrix} = (x + u, y + v, (x + u)^2 + (y + v)^2)$$

Chasles's Identity

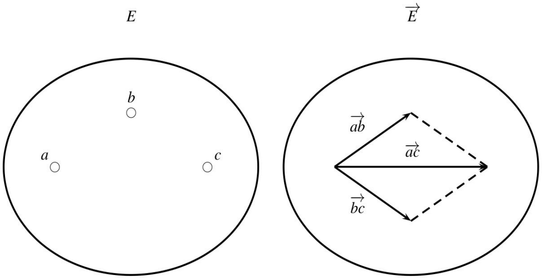
Given any three points $a, b, c \in E$, since $c = a + \overrightarrow{ac}$, $b = a + \overrightarrow{ab}$ and $c = b + \overrightarrow{bc}$, we get

$$c = b + \overrightarrow{bc} = (a + \overrightarrow{ab}) + \overrightarrow{bc} = a + (\overrightarrow{ab} + \overrightarrow{bc})$$

Thus

$$\overrightarrow{ab} + \overrightarrow{bc} = \overrightarrow{ac}$$

which is known as Chasles's identity



Notation

Let $Q = \{3, 5, 9\}$ the

$$\sum_{i \in Q} i^2 = 3^2 + 5^2 + 9^2$$

Alternative representation of the set $\{a_3, a_8, a_{11}, a_d\}$ is

$$(a_i)_{i \in Q} \text{ where } Q = \{3, 8, 11, d\}.$$

Alternative representation of the set $\{\overrightarrow{a_1 a_3}, \overrightarrow{a_1 a_7}, \overrightarrow{a_1 a_8}, \overrightarrow{a_1 a_9}\}$ is

$$(\overrightarrow{a_1 a_j})_{j \in S - \{1\}} \text{ where } S = \{1, 3, 7, 8, 9\}.$$

A fundamental concept in linear algebra is that of a linear combination. The corresponding concept in affine geometry is that of an affine combination, also called a barycenter.

Lemma Given an affine space E, let $(a_i)_{i \in I}$ be a family of points in E, and let $(\lambda_i)_{i \in I}$ be a family of scalars. For any two points $a, b \in E$, the following properties hold:

(1) If $\sum_{i \in I} \lambda_i = 1$, then

$$a + \sum_{i \in I} \lambda_i \overrightarrow{aa_i} = b + \sum_{i \in I} \lambda_i \overrightarrow{ba_i}.$$

(2) If $\sum_{i \in I} \lambda_i = 0$, then

$$\sum_{i \in I} \lambda_i \overrightarrow{aa_i} = \sum_{i \in I} \lambda_i \overrightarrow{ba_i}.$$

Lemma Given an affine space E, let $(a_i)_{i\in I}$ be a family of points in E, and let $(\lambda_i)_{i\in I}$ be a family of scalars. For any two points $a, b \in E$, the following properties hold:
(1) If $\sum_{i\in I} \lambda_i = 1$, then

$$a + \sum_{i\in I} \lambda_i \overrightarrow{aa_i} = b + \sum_{i\in I} \lambda_i \overrightarrow{ba_i}.$$

Proof By Chasle's identity, we have

$$\begin{aligned} a + \sum_{i\in I} \lambda_i \overrightarrow{aa_i} &= a + \sum_{i\in I} \lambda_i (\overrightarrow{ab} + \overrightarrow{ba_i}), \\ &= a + (\sum_{i\in I} \lambda_i)\overrightarrow{ab} + \sum_{i\in I} \lambda_i \overrightarrow{ba_i}, \\ &= a + \overrightarrow{ab} + \sum_{i\in I} \lambda_i \overrightarrow{ba_i}, \\ &= b + \sum_{i\in I} \lambda_i \overrightarrow{ba_i}, \end{aligned}$$

Lemma Given an affine space E, let $(a_i)_{i\in I}$ be a family of points in E, and let $(\lambda_i)_{i\in I}$ be a family of scalars. For any two points $a, b \in E$, the following properties hold:
(2) If $\sum_{i\in I} \lambda_i = 0$, then

$$\sum_{i\in I} \lambda_i \overrightarrow{aa_i} = \sum_{i\in I} \lambda_i \overrightarrow{ba_i}.$$

Proof We also have

$$\begin{aligned}\sum_{i\in I} \lambda_i \overrightarrow{aa_i} &= \sum_{i\in I} \lambda_i (\overrightarrow{ab} + \overrightarrow{ba_i}),\\ &= (\sum_{i\in I} \lambda_i)\overrightarrow{ab} + \sum_{i\in I} \lambda_i \overrightarrow{ba_i},\\ &= \sum_{i\in I} \lambda_i \overrightarrow{ba_i},\end{aligned}$$

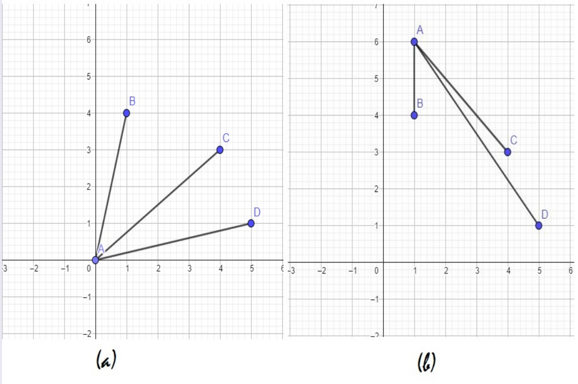
Thus, by the Lemma, for any family of points $(a_i)_{i\in I}$ in E, for any family $(\lambda_i)_{i\in I}$ of scalars such that $\sum_{i\in I}\lambda_i = 1$, the point

$$x = a + \sum_{i\in I}\lambda_i \overrightarrow{aa_i}$$

is independent of the choice of the **origin** $a \in E$. The term "origin" signifies that the vectors $\overrightarrow{aa_i}$ in the above expression have the same initial point a.

The point x is function of the points a_i which, in the sequel, we call them **control points**.

Example 30



Let $\lambda_1 = 0.2, \lambda_2 = 0.4, \lambda_3 = -0.6$. Notice that $\sum_i \lambda_i = 0$. According to the previous lemma

$$\lambda_1 \overrightarrow{AB} + \lambda_2 \overrightarrow{AC} + \lambda_3 \overrightarrow{AD} = (-1.2, 1.4)$$

are equal in both figures, even though A's in the figures are positioned differently.

Definition For any family of points $(a_i)_{i\in I}$ in E, for any family $(\lambda_i)_{i\in I}$ of scalars such that $\sum_{i\in I}\lambda_i = 1$, and for any $a \in E$, the point

$$x = a + \sum_{i\in I}\lambda_i \overrightarrow{aa_i} \tag{4}$$

(which is independent of $a \in E$, by a previous Lemma) is called the **barycenter** (or **barycentric combination**, or **affine combination**) of the points a_i assigned the weights λ_i, and it is denoted by

$$x = \sum_{i\in I}\lambda_i a_i \tag{5}$$

Note that (5) is a point. Since barycentric combination is independent of the pivot, it could be viewed as barycentric combination with 0 pivot:

$$x = \sum_{i\in I}\lambda_i a_i = 0 + \sum_{i\in I}\lambda_i \overrightarrow{0a_i}$$

$$x = a + \sum_{i \in I} \lambda_i \overrightarrow{aa_i} \qquad \text{(cf. 4)}$$

Given a family of weighted points $((a_i, \lambda_i))_{i \in I}$, where $\sum_{i \in I} \lambda_i = 1$, we say that the point $\sum_{i \in I} \lambda_i a_i$ is the barycenter of the family of weighted points $((a_i, \lambda_i))_{i \in I}$.

Note that the barycenter x of the family of weighted points $((a_i, \lambda_i))_{i \in I}$ is the unique point such that

$$\overrightarrow{ax} = \sum_{i \in I} \lambda_i \overrightarrow{aa_i}, \quad \text{for every } a \in E$$

The above expression is obtained from (4) by moving a to the left of the equation. Setting $a = x$ in the equation, the point x is the unique point such that

$$\sum_{i \in I} \lambda_i \overrightarrow{xa_i} = 0$$

$\therefore$ *Vectors with initial points at the barycenter and endpoints at a_i's add up to zero.*

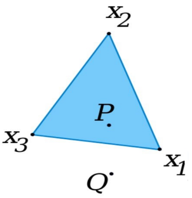
Digression: Convex Combination

Given a finite number of points $x_1, \ldots, x_n$ in a real vector space, a convex combination of these points is a point of the form

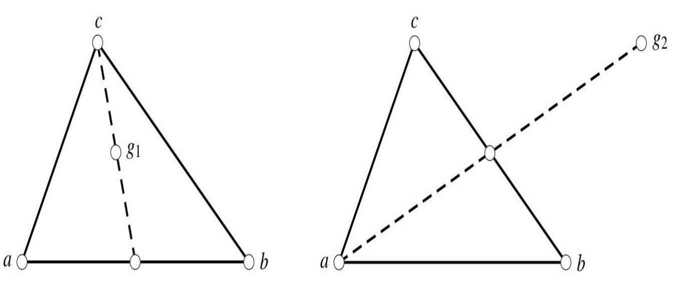
$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$$

where the real numbers α_i satisfy $\alpha_i \geq 0$ and $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 1$.

Given three points x_1, x_2, x_3 in a plane as shown in the figure, the point P is a convex combination of the three points, while Q is not. Q is, however, an affine combination of the three points



Example 31



Barycenters:

$$g_1 = \frac{1}{4}a + \frac{1}{4}b + \frac{1}{2}c = \frac{1}{2}\left(\frac{1}{2}a + \frac{1}{2}b\right) + \frac{1}{2}c$$

$$g_2 = -a + b + c = -a + 2\left(\frac{1}{2}b + \frac{1}{2}c\right)$$

Example 32

Let (a,b,c,d) be a sequence of points in A^2. Observe that $(1-t)^3 + 3t(1-t)^2 + 3t^2(1-t) + t^3 = 1$, since the sum on the left-hand side is obtained by expanding $(t + (1-t))^3 = 1$ using the binomial formula. Thus,

$$(1-t)^3 a + 3t(1-t)^2 b + 3t^2(1-t)c + t^3 d$$

is a well-defined affine combination. Then, we can define the curve $F : A \to A^2$ such that

$$F(t) = (1-t)^3 a + 3t(1-t)^2 b + 3t^2(1-t)c + t^3 d.$$

Such a curve is called a Bézier curve, and (a, b, c, d) are called its control points.

Affine Subspaces

Definition Given an affine space $< E, \overrightarrow{E}, + >$ a subset V of E is an affine subspace (of $< E, \overrightarrow{E}, + >$) if for every family of weighted points $((a_i, \lambda_i))_{i \in I}$ in V such that $\sum_{i \in I} \lambda_i = 1$, the barycenter $\sum_{i \in I} \lambda_i a_i$ belongs to V.

The subspace associated with an affine subspace is often called its **direction**. That is, in the triple $< E, \overrightarrow{E}, + >$, the subspace $\overrightarrow{E}$ is the direction.

Example 33

Consider the subset U of R^2 defined by

$$U = \{(x, y) \in R^2 : \; ax + by = c\}$$

U is an affine subspace of A^2. In fact, it is just a usual line in A^2

Lemma

Let $< E, \overrightarrow{E}, + >$ be an affine space.

(1) A nonempty subset V of E is an affine subspace iff for every point $a \in V$, the set $\overrightarrow{V_a} \triangleq \{\overrightarrow{ax} : \ x \in V\}$ is a subspace of $\overrightarrow{E}$.

$\therefore$ *V_a is the set of all the vectors having their initial point at $a \in V$, and endpoint at all possible points in V.*

Consequently, all the points in V could be expressed as $V = a + \overrightarrow{V_a}$.

Removing the restriction of the vector initial points to a, we obtain

$$\overrightarrow{V} = \{\overrightarrow{xy} : \ x, y \in V\}$$

which is a subspace of $\overrightarrow{E}$, and $\overrightarrow{V_a} = \overrightarrow{V}$ for all $a \in E$. Thus, $V = a + \overrightarrow{V}$.

(2) For any subspace $\overrightarrow{V}$ of $\overrightarrow{E}$ and for any $a \in E$, the set $V = a + \overrightarrow{V}$ is an affine subspace.

By the **dimension of the affine subspace** V, we mean the dimension of $\overrightarrow{V}$. An affine subspace of dimension 1 is called a line, and an affine subspace of dimension 2 is called a plane.

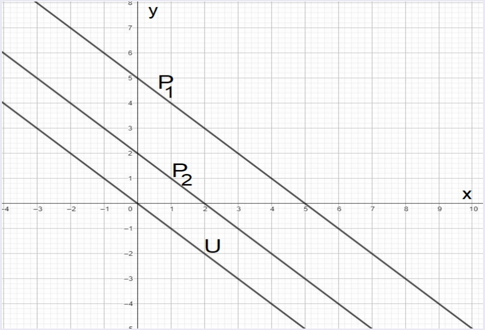
We say that two affine subspaces U and V are **parallel** if their directions $\overrightarrow{U}$ and $\overrightarrow{V}$ are identical.
$\rightarrow U = a + \overrightarrow{U}$ and $V = b + \overrightarrow{U}$ for any $a \in U$ and any $b \in V$
Thus V is obtained from U by the translation $\overrightarrow{ab}$:

$$U + \overrightarrow{ab} = a + \overrightarrow{U} + \overrightarrow{ab} = a + \overrightarrow{ab} + \overrightarrow{U} = b + \overrightarrow{U} = V$$

Example 34

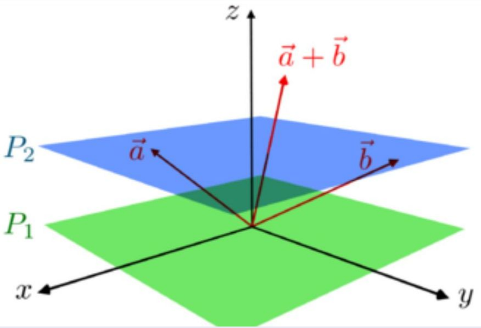
Given the affine space $< E, \overrightarrow{E}, + >$, let $a \in E$ and $\overrightarrow{u} \in \overrightarrow{E}$. We have seen that $a + \overrightarrow{u}$ is a point in E. Consider the sums $a + \overrightarrow{E}$. What we obtain is the set of points E.

Example 35



$\overrightarrow{U}$ is a one dimensional subspace of $\mathbb{R}^2$. P_1 and P_2 are affine subspaces of $< U, \overrightarrow{U}, + >$ which equal $5 + \overrightarrow{U}$ and $2 + \overrightarrow{U}$ respectively. Affine subspaces P_1 and P_2 are parallel. P_1 is obtained from P_2 by translating 5-2 (i.e., 3) units.

Example 36



In $\mathbb{R}^3$, the upper plane (in blue) P_2 is not a vector subspace, since $0 \notin P_2$ and $a + b \notin P_2$; it is an affine subspace. Its direction (the linear subspace associated with this affine subspace) is the lower (green) plane P_1, which is a vector subspace. Although a and b are in P_2, their difference is a displacement vector, which does not belong to P_2, but belongs to vector space P_1.

We say that three points a,b,c are **collinear** if the vectors $\overrightarrow{ab}$ and $\overrightarrow{ac}$ are linearly dependent.
We say that four points a,b,c,d are **coplanar** if the vectors $\overrightarrow{ab}$, $\overrightarrow{ac}$ and $\overrightarrow{ad}$ are linearly dependent.

Lemma Given an affine space $< E, \overrightarrow{E}, + >$, for any family $(a_i)_{i \in I}$ of points in E:
The set V of barycenters $\sum_{i \in I} \lambda_i a_i$ is the smallest affine subspace containing $(a_i)_{i \in I}$.

Lemma Given an affine space $< E, \overrightarrow{E}, + >$, let $(a_i)_{i \in I}$ be a family of points in E. If the family $(\overrightarrow{a_i a_j})_{j \in (I - \{i\})}$ is linearly independent for some $i \in I$, then $(\overrightarrow{a_i a_j})_{j \in (I - \{i\})}$ is linearly independent for every $i \in I$.

Example 37

Let (a_1, a_2, a_3, a_4) be a family of points in $\mathbb{R}^3$ that equals

$$\left(\begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \right)$$

The family

$$(\overrightarrow{a_1 a_2}, \overrightarrow{a_1 a_3}, \overrightarrow{a_1 a_4}) = \left(\begin{bmatrix} 0 \\ -1 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} \right)$$

is linearly independent. According to the Lemma, so are $(\overrightarrow{a_2 a_1}, \overrightarrow{a_2 a_3}, \overrightarrow{a_2 a_4})$, $(\overrightarrow{a_3 a_1}, \overrightarrow{a_3 a_2}, \overrightarrow{a_3 a_4})$, and $(\overrightarrow{a_4 a_1}, \overrightarrow{a_4 a_2}, \overrightarrow{a_4 a_3})$.

Definition Given an affine space $< E, \overrightarrow{E}, + >$, a family $(a_i)_{i \in I}$ of points in E is **affinely independent** if the family $(\overrightarrow{a_i a_j})_{j \in (I - \{i\})}$ is linearly independent for some $i \in I$.

Example 38

Given the family of points $(a_1, a_2, a_3, a_4, a_5)$, to find out their affine independence it is sufficient to check linear independence of vectors using any point as a pivot. Below, a_1 is used as a pivot.

$$(\overrightarrow{a_1 a_2}, \overrightarrow{a_1 a_3}, \overrightarrow{a_1 a_4}, \overrightarrow{a_1 a_5})$$

Lemma

Given an affine space $< E, \overrightarrow{E}, + >$.
Let $(a_0, \ldots, a_m)$ be a family of $m+1$ points in E.
Let $x \in E$, and assume that $x = \sum_{i=0}^{m} \lambda_i a_i$, where $\sum_{i=0}^{m} \lambda_i = 1$.

Then, the family $(\lambda_0, \ldots, \lambda_m)$ such that $x = \sum_{i=0}^{m} \lambda_i a_i$ is unique iff the family $(\overrightarrow{a_0 a_1}, \ldots, \overrightarrow{a_0 a_m})$ is linearly independent.

Note that "*the family* $(\overrightarrow{a_0 a_1}, \ldots, \overrightarrow{a_0 a_m})$ *is linearly independent*" can alternatively be expressed as "*the family of points* $(a_0, \ldots, a_m)$ *is affinely independent*".

Definition Given two affine spaces $< E, \overrightarrow{E}, + >$ and $< E', \overrightarrow{E'}, +' >$, a function $f : E \to E'$ is an **affine map** iff for every family $((a_i, \lambda_i))_{i \in I}$ of weighted points in E such that $\sum_{i \in I} \lambda_i = 1$, we have

$$f\left(\sum_{i \in I} \lambda_i a_i\right) = \sum_{i \in I} \lambda_i f(a_i) \tag{6}$$

Example 39

Let $f(v) \triangleq Av + b$ where A and b are constant matrices of compatible sizes. We show that f is an affine map. We do this by calculating both sides of (6) for the function f, and showing that they are equal.

$$f\left(\sum_{i\in I}\lambda_i a_i\right)=\sum_{i\in I}\lambda_i f(a_i) \qquad \text{(cf. 6)}$$

Example 39 (cont.)

$$\text{LHS: } f(v)=f(\sum_{i\in I}\lambda_i a_i)=A(\sum_{i\in I}\lambda_i a_i)+b=\sum_{i\in I}\lambda_i Aa_i+b$$

$$\text{RHS: } \sum_{i\in I}\lambda_i f(a_i)=\sum_{i\in I}\lambda_i\,(Aa_i+b)=\sum_{i\in I}\lambda_i Aa_i+\sum_{i\in I}\lambda_i b$$

$$=\sum_{i\in I}\lambda_i Aa_i+\left(\sum_{i\in I}\lambda_i\right)b=\sum_{i\in I}\lambda_i Aa_i+1\cdot b=\sum_{i\in I}\lambda_i Aa_i+b$$

This shows that $Av+b$ is an affine map.

Affine map $Av + b$ is combination of linear transformation $f_1(v) = Av$ and a follow-up translation $f_2(v) = b$.

It can be shown that **every affine map T can be expressed in the form** $T(x) \triangleq Ax + T(0)$ for some matrix A and the vector $T(0)$. We next show this.

Affine Map definition -Revisit

Definition Given two affine spaces $< E, \overrightarrow{E}, + >$ and $< E', \overrightarrow{E'}, +' >$, a function $f: E \to E'$ is an **affine map** iff for every family $((a_i, \lambda_i))_{i \in I}$ of weighted points in E such that $\sum_{i \in I} \lambda_i = 1$, we have

$$f\left(\sum_{i \in I} \lambda_i a_i\right) = \sum_{i \in I} \lambda_i f(a_i)$$

$$f(\lambda_1 a_1 + \cdots + \lambda_m a_m) = \lambda_1 f(a_1) + \cdots + \lambda_m f(a_m)$$

Every affine map $T : \Omega \to R^k$ is expressed in the form $T(x) = Ax + b$, where A and b are constant matrices having compatible sizes.
Let $\{p_0, p_1, ..., p_m\}$ be an affinely independent family of points in R^n and let Ω denote the affine set it spans. We define a unique linear map L on $\mathbb{R}^n$. Note that $\{p_0, p_1, ..., p_m\}$ being affine independent means that the m vectors

$$b_1 \triangleq p_1 - p_0, \ldots, b_m \triangleq p_m - p_0$$

are linearly independent. Since every b_i is in R^n, we have $m \leq n$. If $m = n$, these vectors form a basis of R^n; if $m < n$, adjoin vectors $b_{m+1}, \ldots, b_n$ such that $b_1, \ldots, b_n$ becomes a basis of R^n.
Define the linear map L by

$$L(b_i) = \begin{cases} 0 & \text{for } m < i \leq n \\ T(p_i) - T(p_0) & \text{for } 1 < i \leq m \end{cases}$$

This determines a unique linear map $L : R^n \to R^k$.

Define the linear map L by

$$L(b_i) = \begin{cases} 0 & \text{for } m < i \leq n \\ T(p_i) - T(p_0) & \text{for } 1 < i \leq m \end{cases}$$

L is defined by its actions on a basis of $\mathbb{R}^n$. Its values are chosen in $\mathbb{R}^k$. Thus $L : \mathbb{R}^n \to \mathbb{R}^k$ is a unique linear map which we will analyze its action on a point in Ω. So,let $x \triangleq \sum_{i=0}^{m} t_i p_i \in \Omega$ (that is, $t_0 + ... + t_m = 1$) and analyze $L(x)$.

Let $x \triangleq \sum_{i=0}^{m} t_i p_i \in \Omega$ (that is, $t_0 + ... + t_m = 1$)

$$\begin{aligned}
L(x) = \quad & L(\textstyle\sum_{i=0}^{m} t_i p_i) \\
& = L\left(\textstyle\sum_{i=0}^{m} t_i (p_i - p_0) + \sum_{i=0}^{m} t_i p_0\right) \\
& = L\left(\textstyle\sum_{i=0}^{m} t_i (p_i - p_0) + (\sum_{i=0}^{m} t_i) p_0\right) \\
& = L\left(\textstyle\sum_{i=1}^{m} t_i (p_i - p_0) + p_0\right) \\
& = L\left(\textstyle\sum_{i=1}^{m} t_i b_i + p_0\right) \\
& = \textstyle\sum_{i=1}^{m} t_i L(b_i) + L(p_0) \\
& = \textstyle\sum_{i=1}^{m} t_i [T(p_i) - T(p_0)] + L(p_0) \\
& = \textstyle\sum_{i=0}^{m} t_i [T(p_i) - T(p_0)] + L(p_0) \\
& = \textstyle\sum_{i=0}^{m} t_i T(p_i) - (\sum_{i=0}^{m} t_i) T(p_0) + L(p_0) \\
& = \textstyle\sum_{i=0}^{m} t_i T(p_i) - T(p_0) + L(p_0), \quad \text{Define } y_0 \triangleq T(p_0) - L(p_0) \\
& = T\left(\textstyle\sum_{i=0}^{m} t_i p_i\right) - y_0 \\
& = T(x) - y_0
\end{aligned}$$

We have obtained

$$L(x) = T(x) - y_0$$

where $L : R^n \to R^k$ is a linear map, $T : \Omega \to R^k$ is an affine map, and $y_0 \in R^k$ is a constant vector.

We have obtained

$$L(x) = T(x) - y_0$$

where $L : R^n \to R^k$ is a linear map, $T : \Omega \to R^k$ is an affine map, and $y_0 \in R^k$ is a constant vector. This leads to the conclusion that every affine map can be expressed as

$$T(x) = L(x) + y_0$$

i.e., sum of a linear map and a translation.

Certain geometric properties are preserved, or invariant, under any affine transformation. If a geometric figure ϕ possesses a property that is invariant under affine transformations, then the image, $f(\phi)$, under any affine transformation f will also have that property.

Theorem 6

Let $f(x) = Ax + b$ be an affine transformation. Then f
(1) maps a line to a line,
(2) maps a line segment to a line segment,
(3) preserves the property of parallelism among lines and line segments,
(4) maps an n-gon to an n-gon,
(5) maps a parallelogram to a parallelogram,
(6) preserves the ratio of lengths of two parallel segments, and
(7) preserves the ratio of areas of two figures.

Affine transformations preserve parallelism

Affine transformations are transformations that preserve parallelism, meaning that if two lines are parallel before the transformation, they will remain parallel after the transformation.

Let's suppose we have two lines in a two-dimensional space, given by their direction vectors $\overrightarrow{u}$ and $\overrightarrow{v}$, respectively. These lines are parallel if and only if their direction vectors are linearly dependent, that is, if there exists a scalar λ such that $\overrightarrow{u} = \lambda\overrightarrow{v}$.

Let $l_1 = u_0 + t\overrightarrow{u}$ and $l_2 = v_0 + t\overrightarrow{v}$, $t \in \mathbb{R}$, be parallel lines. Then $\overrightarrow{u} = \lambda\overrightarrow{v}$ for some $\lambda \in \mathbb{R}$. Let's apply an affine transformation to these lines, given by a matrix A and a translation vector b. The transformed lines will be given by $A(l_1) + \overrightarrow{b}$ and $A(l_2) + \overrightarrow{b}$, respectively.

$$A(l_1) + \overrightarrow{b} = A(u_0 + t\overrightarrow{u}) + \overrightarrow{b} = Au_0 + \overrightarrow{b} + tA\overrightarrow{u} = p_1 + t\overrightarrow{u_1}$$

$$A(l_2) + \overrightarrow{b} = A(v_0 + t\overrightarrow{v}) + \overrightarrow{b} = Av_0 + \overrightarrow{b} + tA\lambda\overrightarrow{u} = p_2 + t\lambda\overrightarrow{u_1}$$

Direction vectors of the transformed lines are $t\overrightarrow{u_1}$ and $t\lambda\overrightarrow{u_1}$ which are multiple of each other. This shows that the images of l_1 and l_2 under an affine transformation are parallel.

Affine transformations preserve lines

Let the line L be defined by $L(t) = u_0 + t\overrightarrow{u}$. Let the affine transformation be $T(x) = Ax + \overrightarrow{b}$. Then

$$\begin{aligned} T(L(t)) &= A(L(t)) + \overrightarrow{b} \\ &= A(u_0 + t\overrightarrow{u}) + \overrightarrow{b} \\ &= Au_0 + \overrightarrow{b} + t\overrightarrow{Au} \end{aligned}$$

The image above is a line through $Au_0 + \overrightarrow{b}$ with direction $\overrightarrow{Au}$.

Composition of Affine transformations is an Affine Transformation

Let the affine transformations $T_1(x) = A_1 x + \overrightarrow{b_1}$ and $T_2(x) = A_2 x + \overrightarrow{b_2}$ have compatible sizes for the composition $T(x) \triangleq T_1(T_2(x))$. Then

$$\begin{aligned} T(x) &= T_1(T_2(x)) \\ &= A_1(A_2 x + \overrightarrow{b_2}) + \overrightarrow{b_1} \\ &= A_1 A_2 x + A_1 \overrightarrow{b_2} + \overrightarrow{b_1} \\ &= Ax + \overrightarrow{b} \end{aligned}$$

where $A \triangleq A_1 A_2$ and $b \triangleq A_1 \overrightarrow{b_2} + \overrightarrow{b_1}$. This shows that T is also an affine transformation.

A single matrix representation of Affine Transformations

Example 40

Consider the affine transformation

$$P(v) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} v + \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

This could be written as

$$\begin{bmatrix} P_x \\ P_y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 5 \\ 3 & 4 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ 1 \end{bmatrix}$$

This representation allows one to combine sequential affine transformations by simply multiplying matricies.

Equation of the line $P(t)$ passing through the two points P_0 and P_1 in affine space:

$$P(t) = P_0 + t(P_1 - P_0)$$

The curve $P(t)$ passes through P_0 at $t = 0$ and P_1 at $t = 1$. Moreover, as t varies, the points on $P(t)$ extend in the direction along the vector from P_0 to P_1; thus, these points lie along the line in affine space generated by P_0 and P_1. Rearranging terms, we can rewrite this as

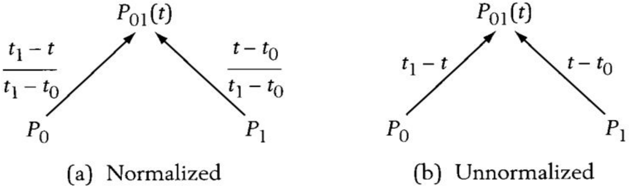
$$P(t) = (1 - t)P_0 + tP_1$$

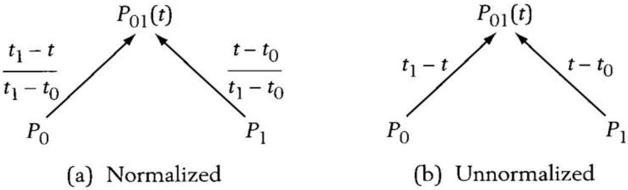
Equation above is called **linear interpolation**.

Now we want a line $P_{01}(t)$ to pass through P_0 at $t = t_0$ and through P_1 at $t = t_1$.

$$P_{01}(t) = \frac{t_1 - t}{t_1 - t_0} P_0 + \frac{t - t_0}{t_1 - t_0} P_1$$

Notice that the coefficients of P_0 and P_1 are precisely the barycentric coordinates of the point $P_{01}(t)$ with respect to the points P_0 and P_1, so linear interpolationis just another way of deriving barycentric coordinates along a line.





$$P_{01}(t) = \frac{t_1 - t}{t_1 - t_0} P_0 + \frac{t - t_0}{t_1 - t_0} P_1$$

$$P_{01}(t) = \frac{1}{t_1 - t_0} [(t_1 - t)P_0 + (t - t_0)P_1]$$

Apex term can be normalized at the end of the process.

Suppose we now have three points P_0, P_1, P_2 in affine space that we wish to interpolate at the parameters t_0, t_1, t_2.

$$P_{01}(t) = \frac{t_1 - t}{t_1 - t_0} P_0 + \frac{t - t_0}{t_1 - t_0} P_1$$

$$P_{12}(t) = \frac{t_2 - t}{t_2 - t_1} P_1 + \frac{t - t_1}{t_2 - t_1} P_2$$

$$P(t) = \begin{cases} P_{01}(t) & t \leq t_1 \\ P_{12}(t) & t \geq t_1 \end{cases}$$

To generate a smooth curve, apply linear interpolation to the two curves $P_{01}(t)$ and $P_{12}(t)$:

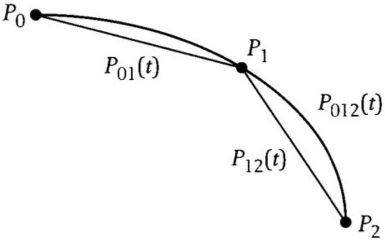
$$P_{012}(t) = \frac{t_2 - t}{t_2 - t_0} P_{01} + \frac{t - t_0}{t_2 - t_0} P_{12}$$

$$P_{012}(t) = \frac{t_2 - t}{t_2 - t_0} P_{01} + \frac{t - t_0}{t_2 - t_0} P_{12}$$

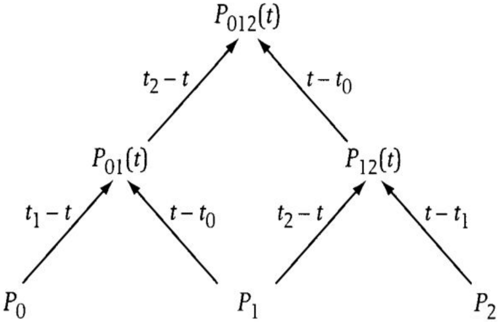
Note that

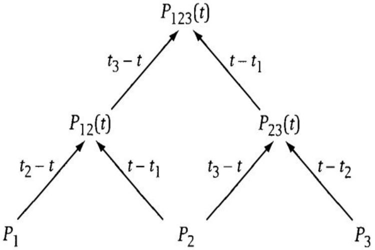
$$P_{012}(t_0) = P_{01}(t_0) = P_0$$

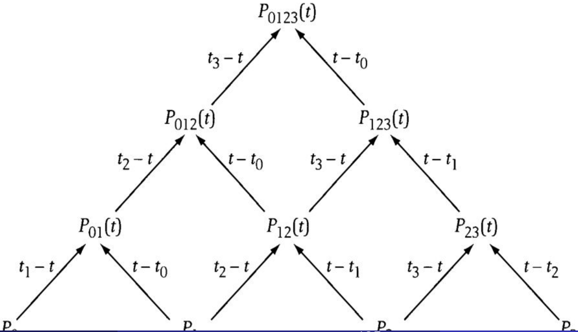
$$P_{012}(t_2) = P_{12}(t_2) = P_2$$



What if we want to interpolate four points P_0, P_1, P_2, P_3 at parameter values t_0, t_l, t_2, t_3? We already know how to build quadratic curves to interpolate portions of this data. We can construct $P_{012}(t)$ to interpolate P_0, P_1, P_2 at t_0, t_l, t_2 and $P_{l23}(t)$ to interpolate P_1, P_2, P_3 at t_l, t_2, t_3.




$P_{123}(t)$
$t_3 - t$
$t - t_1$
$P_{12}(t)$
$P_{23}(t)$
$t_2 - t$
$t - t_1$
$t_3 - t$
$t - t_2$
P_1
P_2
P_3


$P_{0123}(t)$
$t_3 - t$
$t - t_0$
$P_{012}(t)$
$P_{123}(t)$
$t_2 - t$
$t - t_0$
$t_3 - t$
$t - t_1$
$P_{01}(t)$
$P_{12}(t)$
$P_{23}(t)$
$t_1 - t$
$t - t_0$
$t_2 - t$
$t - t_1$
$t_3 - t$
$t - t_2$

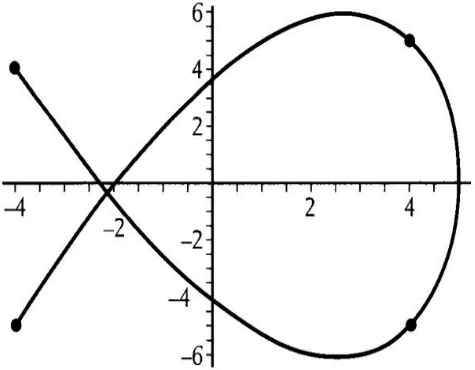


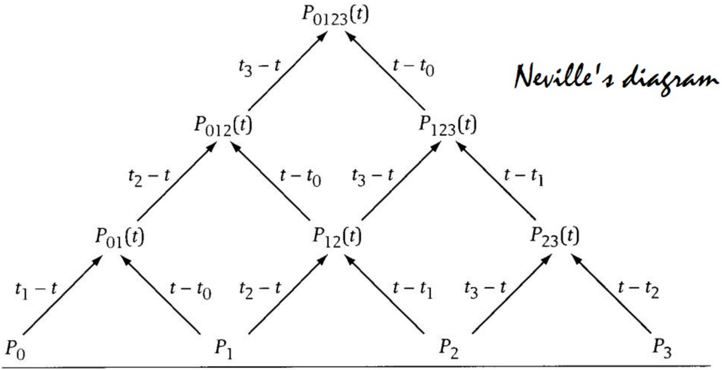
Figure 2.6: The cubic Lagrange polynomial for the control points $P_0 = (-4,4)$, $P_1 = (4,-5)$, $P_2 = (4,5)$, $P_3 = (-4,-5)$ (dots), interpolated at the nodes $t_k = k$, $k = 0,\ldots,3$.

The algorithm for computing $P_{0123}(t)$ is called Neville's algorithm. The curves generated by Neville's algorithm are called Lagrange interpolating polynomials.

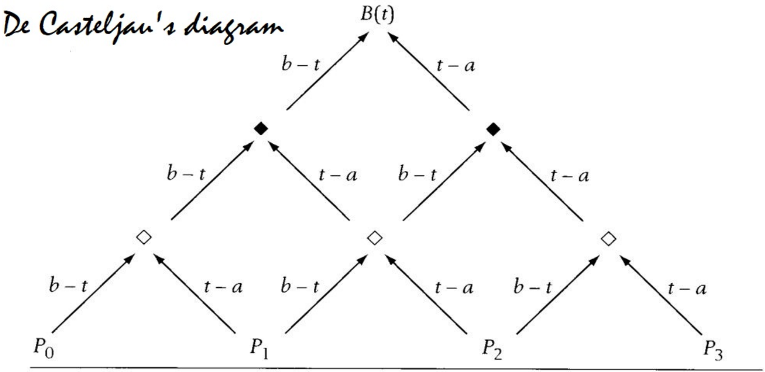
Theorem 7

Given affine points $P_0, \ldots, P_n$ and distinct parameters $t_0, \ldots, t_n$, there is a polynomial curve $P_{0\ldots n}(t)$ of degree n that interpolates the given points at the specified parameters. That is, $P_{0,\ldots,n}(t_k) = P_k, \; k = 0, \ldots, n$.

Neville's algorithm is a dynamic programming procedure for computing points along a polynomial interpolant.



The same triangular structure with easier evaluation algorithm as easy as possible, let us perform the same linear interpolation at each node. The algorithm represented below is called **de Casteljau's evaluation algorithm**, and the curves that emerge at the apex of this diagram are called **Bézier curves**.



Quadratic Bézier curve $B(t)$:

$$B(t) = B_0^2(t)P_0 + B_1^2(t)P_1 + B_2^2(t)P_2$$

Cubic Bézier curve $B(t)$:

$$B(t) = B_0^3(t)P_0 + B_1^3(t)P_1 + B_2^3(t)P_2 + B_3^3(t)P_3$$

where P_0, P_1, P_2, P_3 are the control points and $B_k^n(t)$ are the blending functions defined by

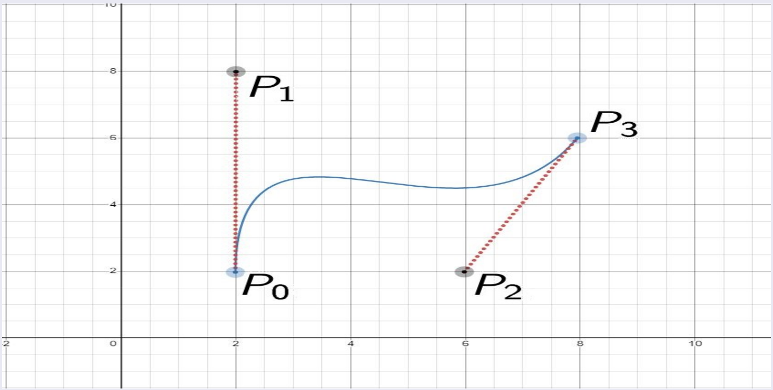
$$B_k^n(t) = \frac{n!}{k!(n-k)!} \frac{(t-a)^k(b-t)^{n-k}}{(b-a)^n}$$

t is the parameter specified between a and b. When $a = 0$ and $b = 1$, the cubic Bézier curve becomes

$$B(t) = (1-t)^3P_0 + 3t(1-t)^2P_1 + 3t^2(1-t)P_2 + t^3P_3 \qquad (7)$$

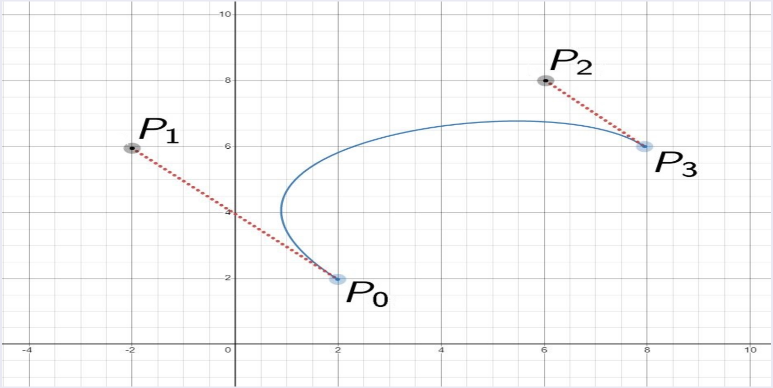
Example 41

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t) P_2 + t^3 P_3 \qquad \text{(cf. 7)}$$



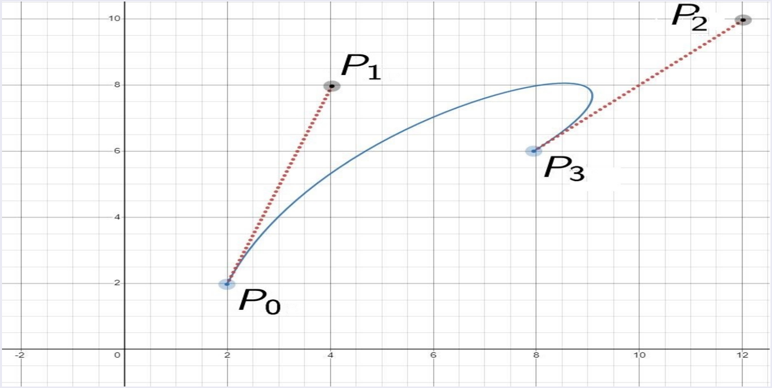
Example 42

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t) P_2 + t^3 P_3 \qquad \text{(cf. 7)}$$

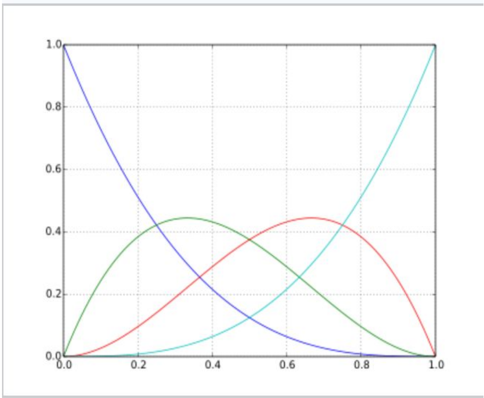


Example 43

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t) P_2 + t^3 P_3 \qquad \text{(cf. 7)}$$



$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t) P_2 + t^3 P_3 \qquad \text{(cf. 7)}$$

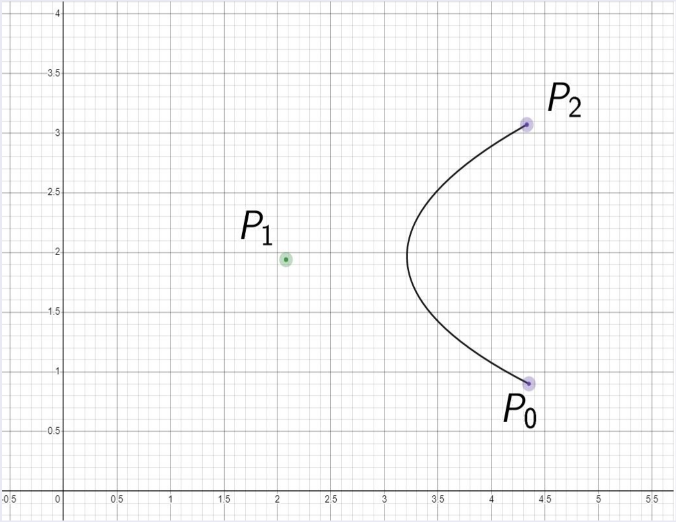


The basis functions on the range t in [0,1] for cubic Bézier curves: blue: $y = (1 - t)^3$, green: $y = 3(1 - t)^2 t$, red: $y = 3(1 - t)t^2$, and cyan: $y = t^3$.

Quadratic Bézier curve $B(t)$:

$$B(t) = B_0^2(t)P_0 + B_1^2(t)P_1 + B_2^2(t)P_2$$

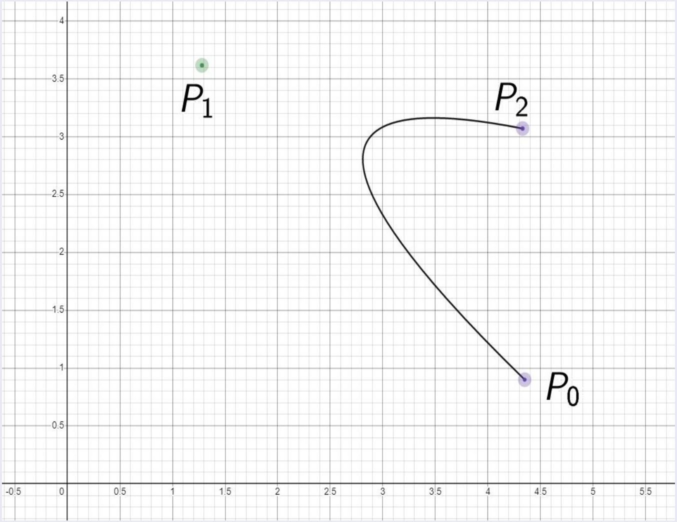
Example 44



Quadratic Bézier curve $B(t)$:

$$B(t) = B_0^2(t)P_0 + B_1^2(t)P_1 + B_2^2(t)P_2$$

Example 45



Example 46

Let $P_0 = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$ and $P_1 = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$.

Their linear interpolation $(1-t)P_0 + tP_1$ in matrix notation

$$(1-t)\begin{bmatrix} -4 \\ 4 \end{bmatrix} + t\begin{bmatrix} 4 \\ -5 \end{bmatrix}, \quad 0 \leq t \leq 1$$

is equivalently viewed as two scalar equations:

$$\begin{array}{rl} -4(1-t) & +4t \\ 4(1-t) & -5t \end{array}$$

where $0 \leq t \leq 1$

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t)P_2 + t^3 P_3$$

Example 47

Let $P_0 = \begin{bmatrix} -4 \\ 4 \end{bmatrix}$, $P_1 = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$, $P_2 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$, and $P_3 = \begin{bmatrix} -4 \\ -5 \end{bmatrix}$.

This yields the cubic Bezier curve

$$\begin{bmatrix} B_x(t) \\ B_y(t) \end{bmatrix} = (1-t)^3 \begin{bmatrix} -4 \\ 4 \end{bmatrix} + 3t(1-t)^2 \begin{bmatrix} 4 \\ -5 \end{bmatrix} + 3t^2(1-t) \begin{bmatrix} 4 \\ 5 \end{bmatrix} + t^3 \begin{bmatrix} -4 \\ -5 \end{bmatrix}$$

$$B_x(t) = (1-t)^3(-4) + 3t(1-t)^2 4 + 3t^2(1-t))4 + t^3(-4)$$

$$B_y(t) = (1-t)^3 4 + 3t(1-t)^2(-5) + 3t^2(1-t))5 + t^3(-5)$$

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t)P_2 + t^3 P_3$$

In the formula above, blending functions are multiplied by the control points P_0, P_1, P_2, P_3. Thus, we can write expressions for x and y components of the Bézier curve B as:

$$\begin{bmatrix} B_x(t) \\ B_y(t) \end{bmatrix} = (1-t)^3 \begin{bmatrix} P_{0,x} \\ P_{0,y} \end{bmatrix} + 3t(1-t)^2 \begin{bmatrix} P_{1,x} \\ P_{1,y} \end{bmatrix} + 3t^2(1-t) \begin{bmatrix} P_{2,x} \\ P_{2,y} \end{bmatrix} + t^3 \begin{bmatrix} P_{3,x} \\ P_{3,y} \end{bmatrix}$$

$$B_x(t) = (1-t)^3 P_{0,x} + 3t(1-t)^2 P_{1,x} + 3t^2(1-t))P_{2,x} + t^3 P_{3,x}$$
$$B_y(t) = (1-t)^3 P_{0,y} + 3t(1-t)^2 P_{1,y} + 3t^2(1-t))P_{2,y} + t^3 P_{3,y}$$

$$B_x(t) = (1-t)^3 P_{0,x} + 3t(1-t)^2 P_{1,x} + 3t^2(1-t))P_{2,x} + t^3 P_{3,x}$$

$$B_y(t) = (1-t)^3 P_{0,y} + 3t(1-t)^2 P_{1,y} + 3t^2(1-t))P_{2,y} + t^3 P_{3,y}$$

In a matric notation, we have

$$B_x(t) = \begin{bmatrix} 1 & t & t^2 & t^3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix} \begin{bmatrix} P_{0,x} \\ P_{1,x} \\ P_{2,x} \\ P_{3,x} \end{bmatrix}$$

$$B_y(t) = \begin{bmatrix} 1 & t & t^2 & t^3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix} \begin{bmatrix} P_{0,y} \\ P_{1,y} \\ P_{2,y} \\ P_{3,y} \end{bmatrix}$$

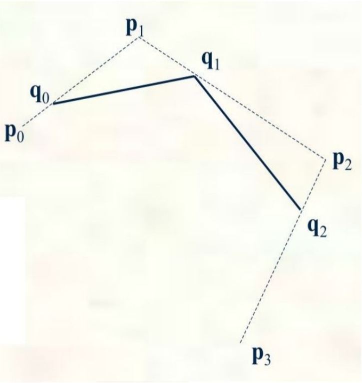
Quadratic Bézier curve

$$B(t) = (1-t)^2 P_0 + 2(1-t)tP_1 + t^2 P_2 = (1-2t+t^2)P_0 + (2t-2t^2)P_1 + t^2 P_2$$

$$B_x(t) = \begin{bmatrix} 1 & t & t^2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 2 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} P_{0,x} \\ P_{1,x} \\ P_{2,x} \end{bmatrix}$$

$$B_y(t) = \begin{bmatrix} 1 & t & t^2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 2 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} P_{0,y} \\ P_{1,y} \\ P_{2,y} \end{bmatrix}$$

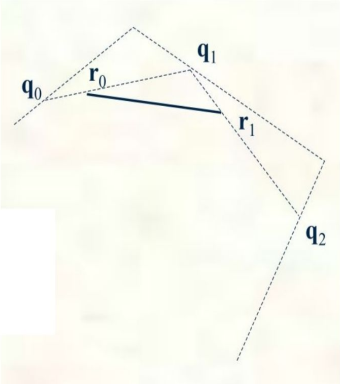
Casteljau Algorithm



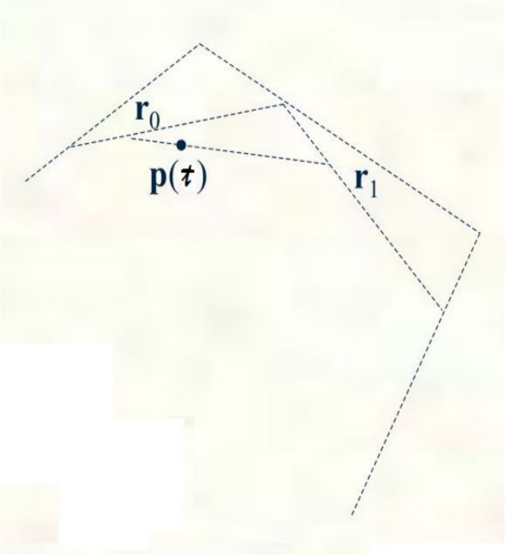
$$q_0(t) = (1 - t)p_0 + tp_1$$

$$q_1(t) = (1 - t)p_1 + tp_2$$

$$q_2(t) = (1 - t)p_2 + tp_3$$

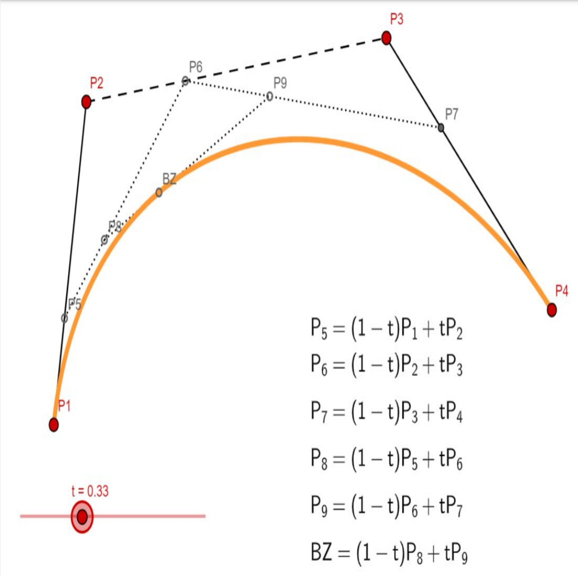


$$r_0(t) = (1-t)q_0(t) + tq_1(t)$$
$$r_1(t) = (1-t)q_1(t) + tq_2(t)$$



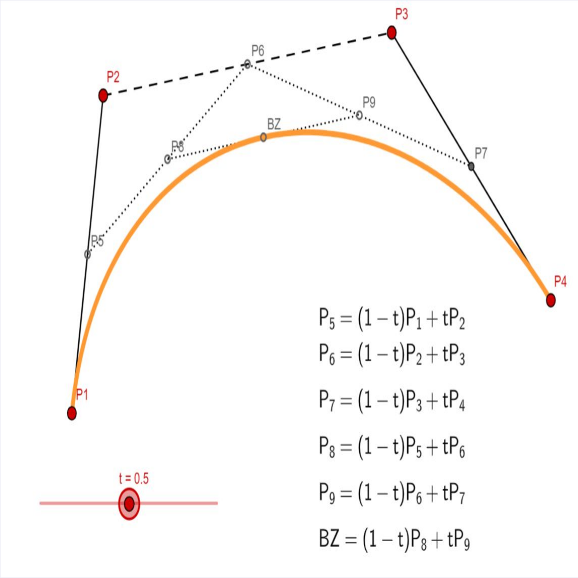
$$p(t) = (1-t)r_0(t) + tr_1(t)$$

Example 48



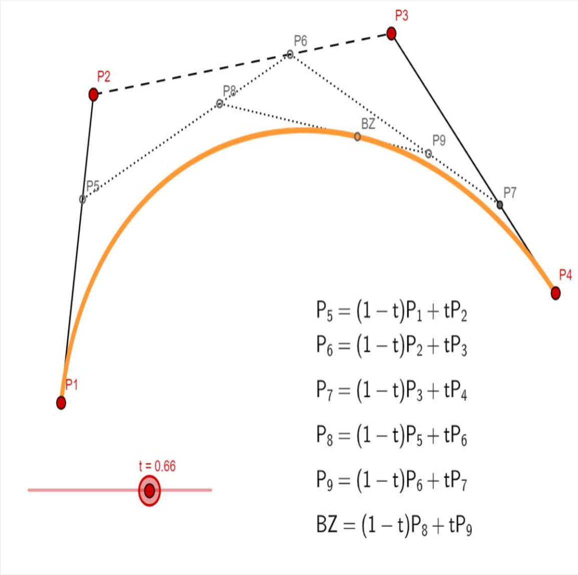
P1
P2
P3
P4
P5
P6
P7
P8
P9
BZ
t = 0.33
$P_5 = (1-t)P_1 + tP_2$
$P_6 = (1-t)P_2 + tP_3$
$P_7 = (1-t)P_3 + tP_4$
$P_8 = (1-t)P_5 + tP_6$
$P_9 = (1-t)P_6 + tP_7$
$BZ = (1-t)P_8 + tP_9$

Example 48 (cont.)

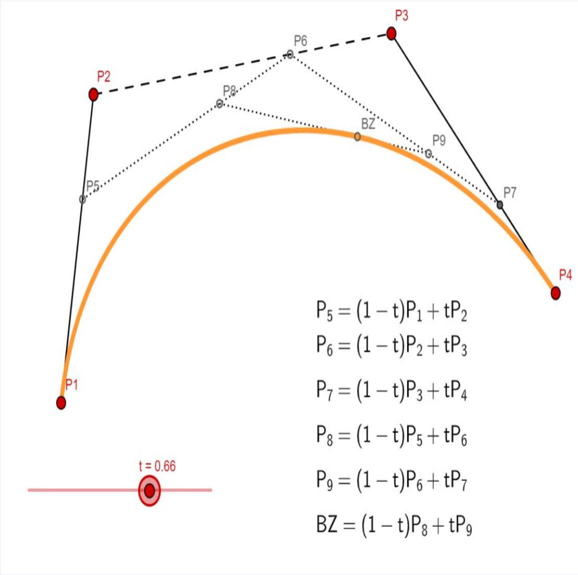


$$P_5 = (1-t)P_1 + tP_2$$
$$P_6 = (1-t)P_2 + tP_3$$
$$P_7 = (1-t)P_3 + tP_4$$
$$P_8 = (1-t)P_5 + tP_6$$
$$P_9 = (1-t)P_6 + tP_7$$
$$BZ = (1-t)P_8 + tP_9$$

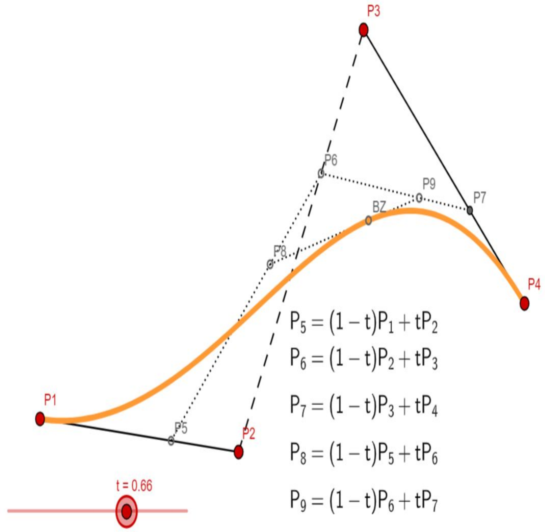
Example 48 (cont.)



Example 48 (cont.)



Example 49



$$P_5 = (1-t)P_1 + tP_2$$

$$P_6 = (1-t)P_2 + tP_3$$

$$P_7 = (1-t)P_3 + tP_4$$

$$P_8 = (1-t)P_5 + tP_6$$

$$P_9 = (1-t)P_6 + tP_7$$

$$BZ = (1-t)P_8 + tP_9$$

Recap

If T is an affine transformation, then you can compute a point on a transformed Bezier curve in two ways:
(1) calculate a point $P(t)$ on the original curve, and then transform this point get a new point T(P(t);
(2) transform the control points P_i to get new ones $T(P_i,)$ construct a Bézier curve from these, and calculate the point at parameter value t on this curve
The result of these two calculations will be the same point.
In short, you can transform the curve just by transforming its control points.
In symbols

$$T\left(\sum_i \phi_i^m(t)P_i\right) = \sum_i \phi_i^m(t)T(P_i)$$

Here ϕ_i^m is the i-th Bernstein polynomial of degree m.

A Remark

Let quadratic Bezier curves be

$$f(t) = A(1-t)^2 + B(1-t)t + Ct^2$$

$$g(t) = D(1-t)^2 + E(1-t)t + Ft^2$$

where A, B, C, D, E, F are control points.
Let's say that you evaluate f and g for some value t and linearly interpolate between those two values by some amount u. The result is always the same as the function below:

$$h(t) = G(1-t)^2 + H(1-t)t + It^2$$

where

$$G = A(1-u) + D(u),\ H = B(1-u) + E(u),\ I = C(1-u) + F(u)$$

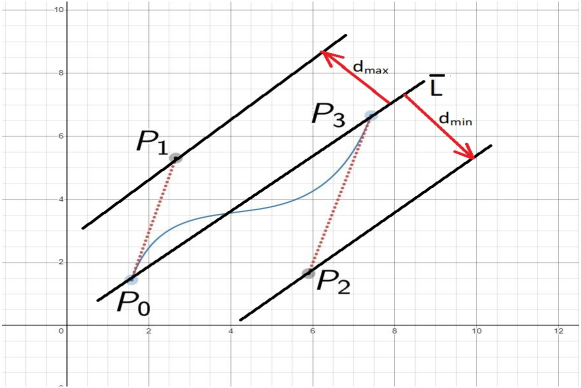
Bezier Clipping Algorithm

Given two Bezier curves it is desired to find their intersection position if it exists. Except some special cases Bezier Clipping algorithm provides a solution to this problem.

Source: Curve intersection using Bézier clipping, T. Sederberg, and T. Nishita, Comput. Aided Des., 22 (9): 538-549 (1990)

Define **fatline** as the region between two parallel lines. Denote by $\overline{L}$ the line passing through P_0 and P_3. Let the normalized equation for this line be

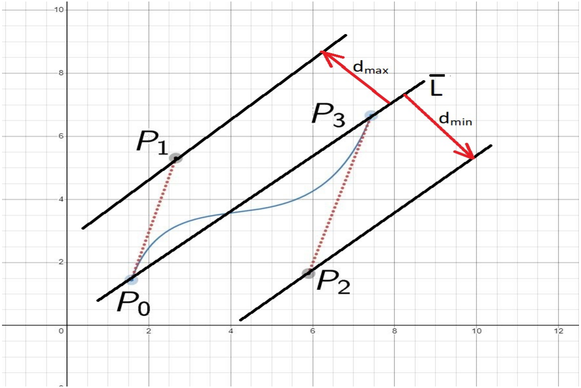
$$ax + by + c = 0, \qquad a^2 + b^2 = 1$$



Distance $d(x, y)$ from (x, y) to the line $\overline{L}$ is an affine transformation:

$$d\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + c = ax + by + c \quad (8)$$

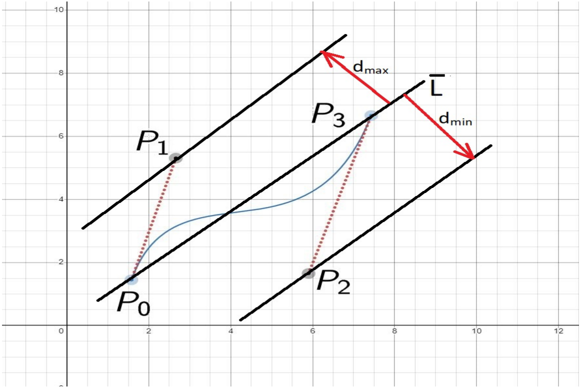
Denote by $d_i = d(x_i, y_i)$ the signed distance from $P_i = (x_i, y_i)$ to $\overline{L}$.



Fatline may more precisely be defined as

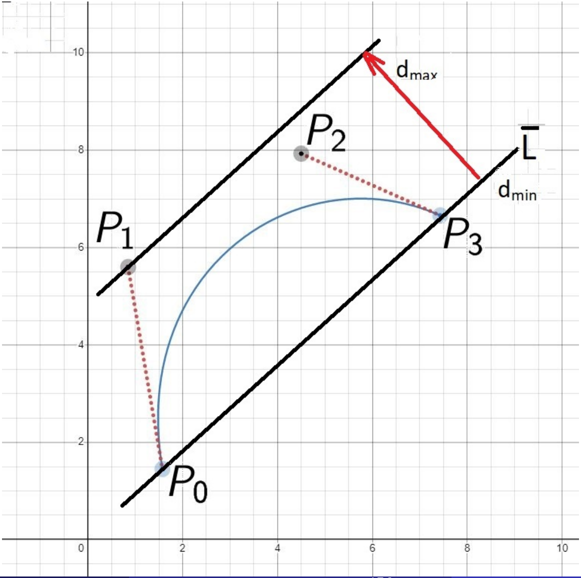
$$\{(x, y):\ d_{min} \leq d(x, y) \leq d_{max}\}$$

where $d_{min} = \min\{d_0, d_1, d_2, d_3\}$ and $d_{max} = \max\{d_0, d_1, d_2, d_3\}$.

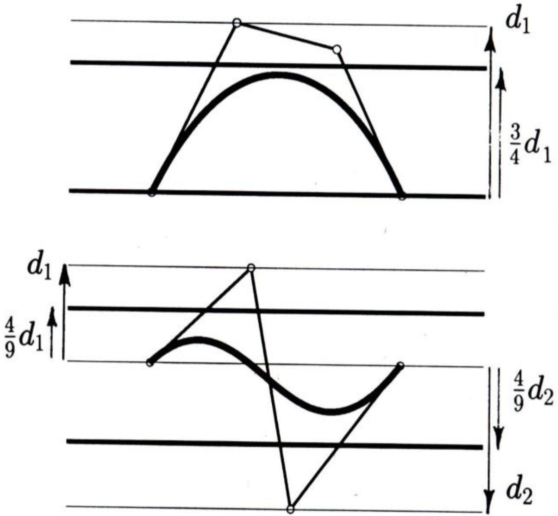


$$\{(x, y) : \; d_{min} \leq d(x, y) \leq d_{max}\}$$

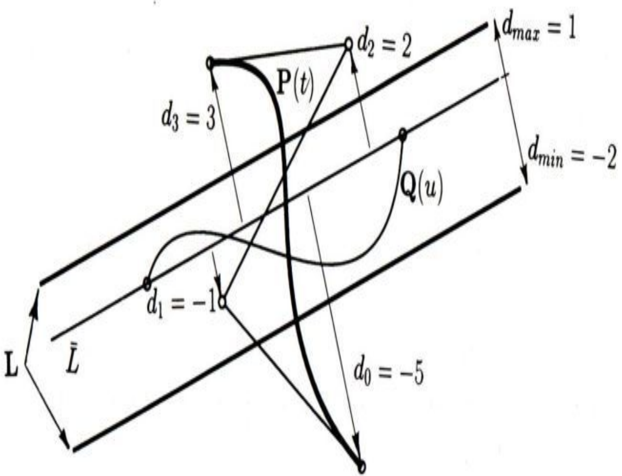
where $d_{min} = \min\{d_0, d_1, d_2, d_3\}$ and $d_{max} = \max\{d_0, d_1, d_2, d_3\}$.



For a tighter fatline,
if $d_1 d_2 > 0$ use $d_{min} = \frac{3}{4} \min\{0, d_1, d_2\}$, $d_{max} = \frac{3}{4} \max\{0, d_1, d_2\}$.
If $d_1 d_2 < 0$, use $d_{min} = \frac{4}{9} \min\{0, d_1, d_2\}$, $d_{max} = \frac{4}{9} \max\{0, d_1, d_2\}$.

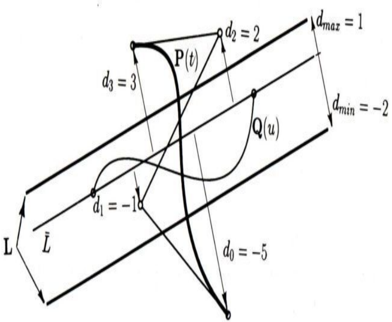


Now, we have a tight fatline containing the cubic Bezier curve Q, and the cubic Bezier curve P. The curves P and Q intersect each other as shown below.



Parts of P that lie outside the fatline are clipped out. Because those parts cannot intersect Q. We want to determine parts of P that lie outside the fatline. Control points of P have distance to $\overline{L}$ calculated by (8):

$$d_i \triangleq d(P_i) = \begin{bmatrix} a & b \end{bmatrix} P_i + c$$



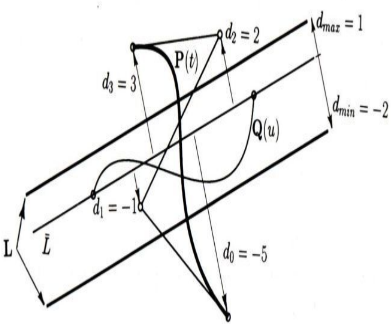
Noting that $P(t)$ has the form

$$P(t) = \sum_{i=0}^{3} P_i B_i^3(t)$$

using (8), any point on $P(t)$ has distance to $\overline{L}$ that equals

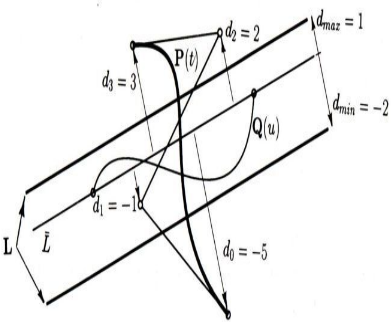
$$d(t) = \sum_{i=0}^{3} d_i B_i^3(t)$$

This has the form of cubic Bezier curve.



Now form a cubic Bezier curve having the control points $D_i = (t_i, d_i)$ such that $t_i = 0, \frac{1}{3}, \frac{2}{3}, 1$. For the data in the figure they are $(0, -5), (\frac{1}{3}, -1), (\frac{2}{3}, 2), (1, 3)$. Using these control points the corresponding nonparametric Bezier curve is

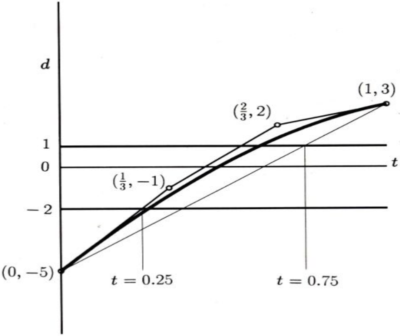
$$D(t) = (t, d(t)) = \sum_{i=0}^{3} D_i B_i^n(t)$$



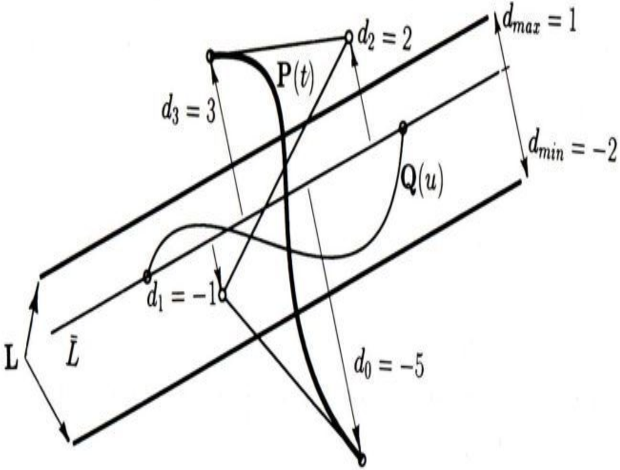
Nonparametric Bezier curve

$$D(t) = (t, d(t)) = \sum_{i=0}^{3} D_i B_i^n(t)$$

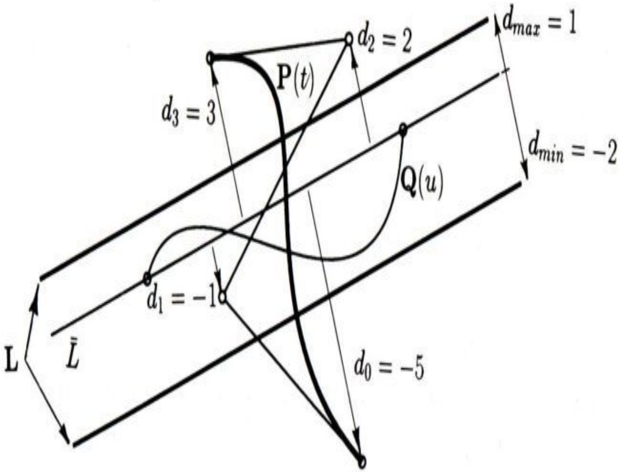
In this example, it is certain that $P(t)$ lies outside the fatline for $t < 0.25$ and $t > 0.75$



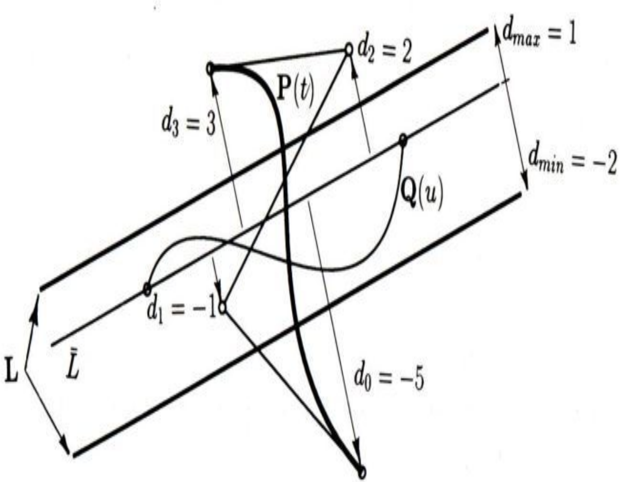
Parts of P that lie outside the fatline are clipped out. Because those parts cannot intersect Q. This makes surviving curve P' shorter. We construct a fatline for P' and clip parts of Q that lie outside the fatline.



Parts of P where $d(t) \geq d_{\max}$ and $d(t) \leq d_{\min}$ are clipped out.



We alternate the roles of P and Q and repeat the process. In few steps, a satisfactorily accurate intersection point is obtained.



Bézier curve and the control points

Cubic Bézier curve points and the control points are related through a linear algebraic equation set. This equation set yields the control points if four distinct parameters and corresponding points on the cubic Bezier curve are known. For convenience in calculations, the parameter values are chosen as $t = 0, \frac{1}{3}, \frac{2}{3}, 1$.

Note that, the curve points corresponding to the parameter values $t = 0$ and $t = 1$ directly present the control points without any computation: $B(0) = P_0$ and $B(1) = P_3$, where B is the Bezier curve function, and P_i is the i-th control point. This reduces problem to solving for two control points P_1 and P_2 only.

Cubic Bézier curve has the form

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t)P_2 + t^3 P_3$$

In terms of components of the curve and control points in two dimensions, it is

$$(B_x(t), B_y(t)) = (1-t)^3(P_{0,x}, P_{0,y}) + 3t(1-t)^2(P_{1,x}, P_{1,y}) +$$

$$3t^2(1-t)(P_{2,x}, P_{2,y}) + t^3(P_{3,x}, P_{3,y})$$

Let some points on the curve be $B(0) = (b_{0,x}, b_{0,y})$, $B(\frac{1}{3}) = (b_{1,x}, b_{1,y})$, $B(\frac{2}{3}) = (b_{2,x}, b_{2,y})$, $B(1) = (b_{3,x}, b_{3,y})$ Writing an equation for each point we obtain four equation sets. Consider first components of each set:

$$\begin{bmatrix} b_{0,x} \\ b_{1,x} \\ b_{2,x} \\ b_{3,x} \end{bmatrix} = \frac{1}{27} \begin{bmatrix} 27 & 0 & 0 & 0 \\ 8 & 12 & 6 & 1 \\ 1 & 6 & 12 & 8 \\ 0 & 0 & 0 & 27 \end{bmatrix} \begin{bmatrix} P_{0,x} \\ P_{1,x} \\ P_{2,x} \\ P_{3,x} \end{bmatrix}$$

$$\begin{bmatrix} b_{0,x} \\ b_{1,x} \\ b_{2,x} \\ b_{3,x} \end{bmatrix} = \frac{1}{27} \begin{bmatrix} 27 & 0 & 0 & 0 \\ 8 & 12 & 6 & 1 \\ 1 & 6 & 12 & 8 \\ 0 & 0 & 0 & 27 \end{bmatrix} \begin{bmatrix} P_{0,x} \\ P_{1,x} \\ P_{2,x} \\ P_{3,x} \end{bmatrix}$$

Using the above equation, first components of the control points are calculated as

$$\begin{bmatrix} P_{0,x} \\ P_{1,x} \\ P_{2,x} \\ P_{3,x} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 6 & 0 & 0 & 0 \\ -5 & 18 & -9 & 2 \\ 2 & -9 & 18 & -5 \\ 0 & 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} b_{0,x} \\ b_{1,x} \\ b_{2,x} \\ b_{3,x} \end{bmatrix}$$

We redo the same work the second components:

$$\begin{bmatrix} b_{0,y} \\ b_{1,y} \\ b_{2,y} \\ b_{3,y} \end{bmatrix} = \frac{1}{27} \begin{bmatrix} 27 & 0 & 0 & 0 \\ 8 & 12 & 6 & 1 \\ 1 & 6 & 12 & 8 \\ 0 & 0 & 0 & 27 \end{bmatrix} \begin{bmatrix} P_{0,y} \\ P_{1,y} \\ P_{2,y} \\ P_{3,y} \end{bmatrix}$$

$$\begin{bmatrix} P_{0,y} \\ P_{1,y} \\ P_{2,y} \\ P_{3,y} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 6 & 0 & 0 & 0 \\ -5 & 18 & -9 & 2 \\ 2 & -9 & 18 & -5 \\ 0 & 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} b_{0,y} \\ b_{1,y} \\ b_{2,y} \\ b_{3,y} \end{bmatrix}$$

If it is in the 3D space, we can also do it the third components.

1. Polynomial Parametrization

In the de Casteljau algorithm, the only operations we perform involving the functions along the edges are addition and multiplication. Since the functions along the edges are linear polynomials, it follows that a Bézier curve with $n+1$ control points is a polynomial curve of degree n because there are n levels from the control points at the base to the curve at the apex of the triangle. Since Bézier curves are polynomial curves, all the tools we know for polynomials apply.

2. Affine Invariance

Let T be an affine map. And let P_i be the i-th control point of n-th degree bezier curve. Then

$$T\left(\sum_{i=0}^{n} P_i B_i^n(t)\right) = \sum_{i=0}^{n} T(P_i) B_i^n(t)$$

In particular, for the cubic Bezier curve with $a = 0$ and $b = 1$, we express this as

$$T\left[(1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t)P_2 + t^3 P_3\right] =$$
$$(1-t)^3 T(P_0) + 3t(1-t)^2 T(P_1) + 3t^2(1-t)T(P_2) + t^3 T(P_3)$$

Recall the definition that given two affine spaces $< E, \overrightarrow{E}, + >$ and $< E', \overrightarrow{E'}, +' >$, a function $f : E \to E'$ is an **affine map** iff for every family $((a_i, \lambda_i))_{i \in I}$ of weighted points in E such that $\sum_{i \in I} \lambda_i = 1$, we have

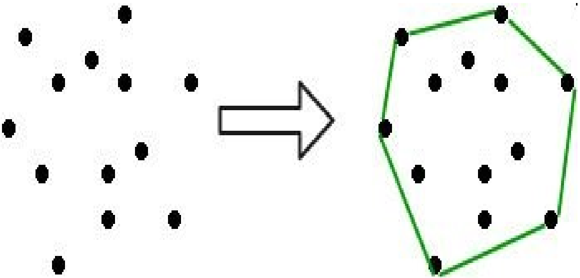
$$f\left(\sum_{i \in I} \lambda_i a_i\right) = \sum_{i \in I} \lambda_i f(a_i) \qquad \text{(cf. 6)}$$

3. Convex Hull Property

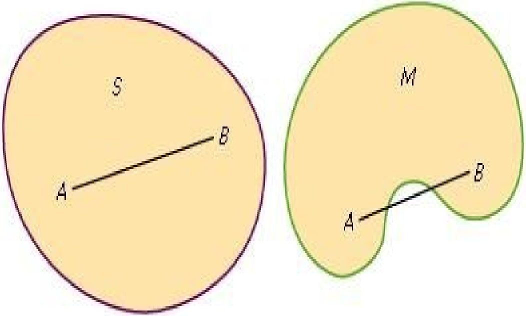
A set S of points in affine space is said to be convex if, whenever P and Q are points in S, the entire line segment from P to Q lies in S.

The intersection S of a collection of convex sets $\{S_i\}$ is a convex set because if P and Q are points in S, they must also be points in each of the sets S_i. Since, by assumption, the sets S_i are convex, the entire line segment from P to Q lies in each set S_i. Hence the entire line segment from P to Q lies in the intersection S, so S too is convex.

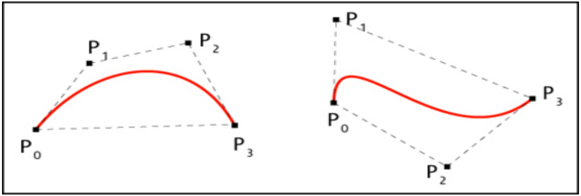
The convex hull of a set of points is defined as the smallest convex polygon, that encloses all of the points in the set. Convex means that the polygon has no corner that is bent inwards.



A convex set is a set of points such that, given any two points A, B in that set, the line AB joining them lies entirely within that set.



A Bézier curve will always be completely contained inside of the Convex Hull of the control points.



4. Symmetry

Replacing t by $a+b-t$ reverses the order of the parameter domain. As the parameter t varies from a to b, the curve B[Po,...,P n](a + b- t) traverses the same points as $B[P_0, \ldots, P_n](t)$ but in the direction from b to a rather than from a to b. Thus $B[P_0, \ldots, P_n](a+b-t)$ is essentially the same curve as $B[P_0, \ldots, P_n](t)$ but with opposite orientation. Similarly, reversing the order of the control points of a Bézier curve generates the same Bézier curve but with opposite orientation. Analytically this means that $B[P_n, \ldots, P_0](t) = B[P_0, \ldots, P_n](a+b-t)$, $a < t < b$.

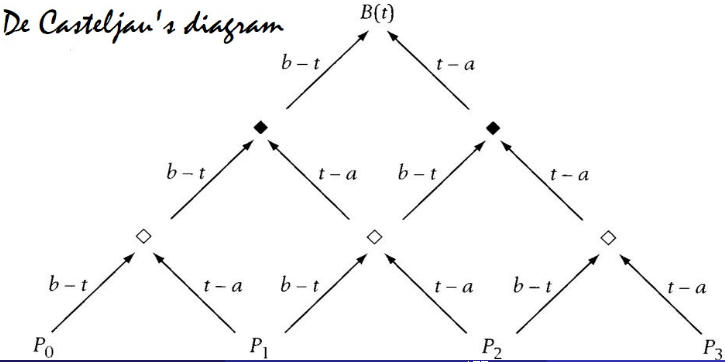
To prove the symmetry, simply replace t by $a+b-t$ in the de Casteljau diagram and observe that the new diagram is the mirror image of the de Casteljau diagram for $B[P_n, \ldots, P_0](t)$.

5. Interpolation of End Points

Unlike Lagrange polynomials, Bézier curves generally do not interpolate all their control points. But Bézier curves always interpolate their first and last control points. In fact,

$$B[P_0, \ldots, P_n](a) = P_0 \text{ and } B[P_0, \ldots, P_n](b) = P_n$$

Set $t = a$ in de Casteljau's algorithm and observe that all the labels on left-pointing arrows become zero while all the labels on right-pointing arrows become one.

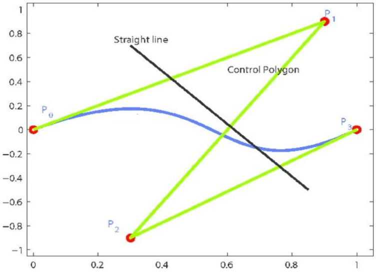


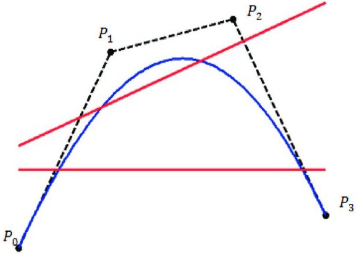
If $k \neq 0$, then any path from P_k to the apex of the triangle must traverse at least one left-pointing arrow, so there is no contribution from P_k to the value of the curve at $t = a$. On the other hand, when $t = a$ all the labels on the single path from P_0 to the apex of the triangle are one. Hence $B[P_0, \ldots, P_n](a) = P_0$. A similar argument for $t = b$ shows that $B[P_0, \ldots, P_n](b) = P_n$.

6. Variation Diminishing

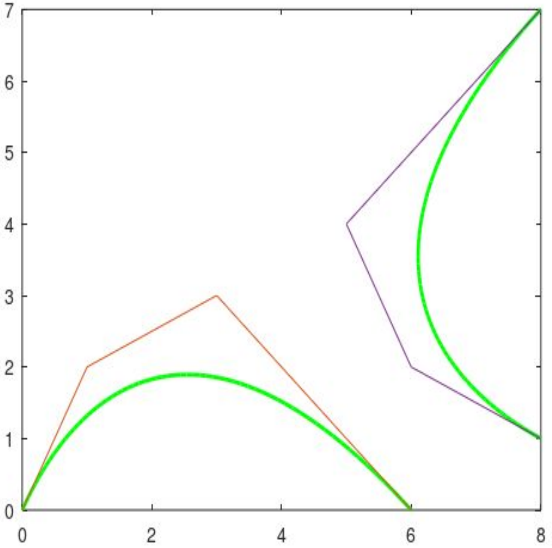
Bezier curves exhibit a variation diminishing property. Informally this means that the Bezier curve will not "wiggle" any more than the control polygon does. In other words, the curve will not wiggle unless the designer specifically introduces wiggling in the control polygon.

If a line is drawn through the curve, the number of intersections with the curve will be less than or equal to the number of intersections with the control polygon.

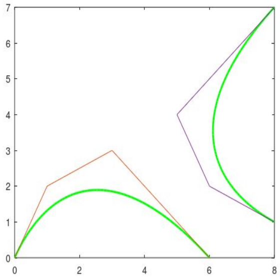

Straight line
Control Polygon
P 0
P 1
P 2
P 3


P_0
P_1
P_2
P_3

7. Control points determine the curve



New orientation and translation do not change the curve shape.




```
>>drawBezierCurve([0 0;1 2;3 3;6 0], 'linewidth', 2, 'color', 'g');
>>hold on
>>plot([0 1 3 6],[0 2 3 0])
>>hold on
>>drawBezierCurve([8 1 ;6 2;5 4;8 7], 'linewidth', 2, 'color', 'g')
>>hold on
>>plot([8 6 5 8],[1 2 4 7])
```

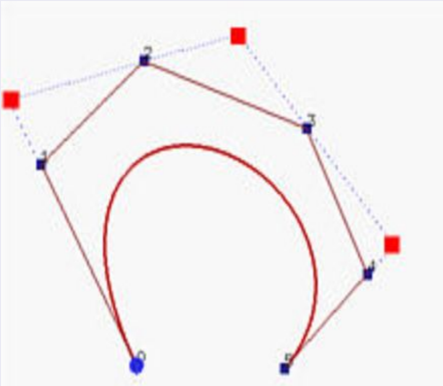
Octave codes for the Bezier curves in the figure above.

8. Degree Elevation

Increase the degree of a Bézier curve without changing its shape. Suppose we have a Bézier curve of degree n defined by $n+1$ control points $P_0, P_1, P_2, ..., P_n$ and we want to increase the degree of this curve to $n+1$ without changing its shape. Since a degree $n+1$ Bézier curve is defined by $n+2$ control points, we need to find such a new set of control points. Obviously, P_0 and P_n must be in the new set because the new curve also passes through them. Therefore, what we need is only n new control points. Let the new set of control points be $Q_0, Q_1, Q_2, ..., Q_{n+1}$. Noting that $Q_0 = P_0$ and $Q_{n+1} = P_n$, the other control points are computed as follows:

$$Q_i = \frac{i}{n+1} P_{i-1} + \left(1 - \frac{i}{n+1}\right) P_i, \quad 1 \leq i \leq n$$

Example 50



To elevate the degree, note the following:

$$B(t) = (1-t)B(t) + tB(t)$$

Shape preserving linear maps

Rotation, translation, and reflection preserve shape. We next show that, nonzero multiples of rotation matrices preserve angles. By such matrices, when control points are transformed, their new positions result in the same curve shape.
A matrix $A \in R^{nn}$ is orthogonal if $A^T A = AA^T = I^n$. Let $T : R^n \to R^n$ and $k \in R$ such that kT is orthogonal. Let $x, y \in R^n$. Then,

$$\cos^{-1}\left(\frac{<T(x), T(y)>}{||T(x)|| \cdot ||T(y)||}\right) = \cos^{-1}\left(\frac{<kT(x), kT(y)>}{||kT(x)|| \cdot ||kT(y)||}\right) = \cos^{-1}\left(\frac{<x, y>}{||x|| \cdot ||y||}\right)$$

because $||kT(x)||^2 = <kT(x), kT(x)> = <x, x> = ||x||^2$, so $||kT(x)|| = ||x||$ for all x. Hence, T is angle-preserving. It can also be shown that, as a linear transformation, an orthogonal matrix preserves the inner product of vectors.

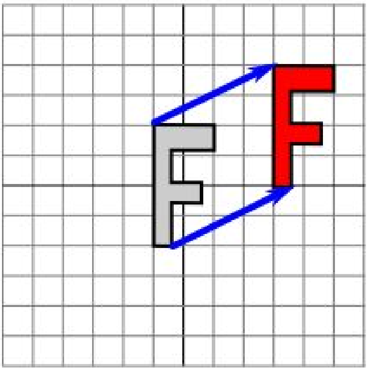
Note that $||x|| \cdot ||y|| \cos\theta = < x, y >$ where θ is the angle between the vectors x and y.

Translation preserves shape.

If (x, y) is the original point and (x_1, y_1) is the transformed point, then the formula for a translation **translate(e, f)** is

$$\text{translate}\ (e, f) = \begin{cases} x_1 & = x + e \\ y_1 & = y + f \end{cases}$$

where e is the number of units by which the point is moved horizontally and f is the amount by which it is moved vertically.



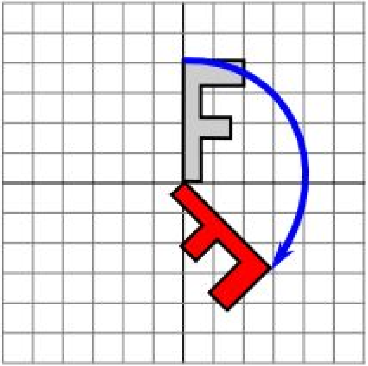
translate(4, 2)

Rotation preserves shape.

If (x, y) is the original point and (x_1, y_1) is the transformed point, then the formula for a rotation **rotate(r)** is

$$\text{rotate}\ (r) = \begin{cases} x_1 & = x\cos(r) - y\sin(r) \\ y_1 & = x\sin(r) + y\cos(r) \end{cases}$$

where r is the amount of rotation.



rotate(-135^0)

A Bézier curve is an affine combination of its control points.

Any affine transformation of a curve is the curve of the transformed control points.

Any shape preserving transformation of te control points, transforms the curve while preserving its shape.

Modifying the parameter interval

If the parameter starting value is 0 and ending value is 1,and the desired values are t_k and t_{k+1} respectively, then we do change of variable:

$$t \leftarrow \frac{t - t_k}{t_{k+1} - t_k}$$

Derivative of a Cubic Bézier Curve

Cubic Bézier Curve

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t)P_2 + t^3 P_3$$

Its derivative $B'(t)$ is

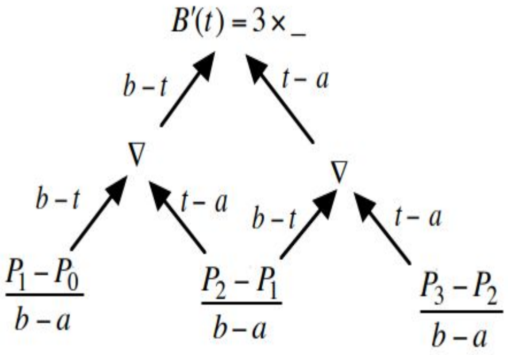
$$-3(1-t)^2 P_0 + 3(1-t)^2 P_1 - 6t(1-t)P_1 + 6t(1-t)P_2 - 3t^2 P_2 + 3t^2 P_3$$

Its value at $t = 0$ is $3(P_1 - P_0)$

Its value at $t = 1$ is $3(P_3 - P_2)$

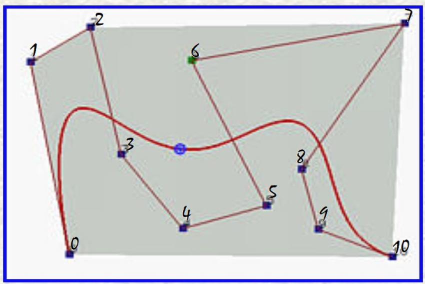
Note that derivative of a cubic Bezier curve is a quadratic Bezier curve

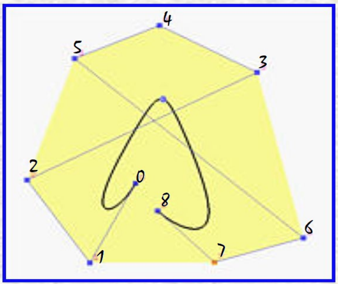
We can compute derivative of a Bezier curve using de Casteljau algorithm: Place the coefficients $\frac{P_1 - P_0}{b-a}, \frac{P_2 - P_1}{b-a}, \ldots, \frac{P_n - P_{n-1}}{b-a}$ at the base diagram, run $n-1$ levels of the algorithm, and multiply the output by n.

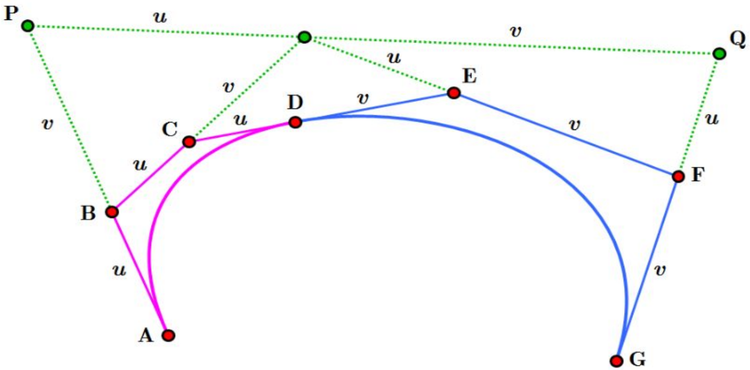


Derivative of a cubic Bezier curve

Bézier Curves Are Tangent to Their First and Last Legs.
Letting $u = 0$ and $u = 1$ gives $C'(0) = n(P_1 - P_0)$ and $C'(1) = n(P_n - P_{n-1})$ The first means that the tangent vector at $u = 0$ is in the direction of $P_1 - P_0$ multiplied by n. Therefore, the first leg in the indicated direction is tangent to the Bézier curve. The second means that the tangent vector at $u = 1$ is in the direction of $P_n - P_{n-1}$ multiplied by n. Therefore, the last leg in the indicated direction is tangent to the Bézier curve. The following figures show this property.



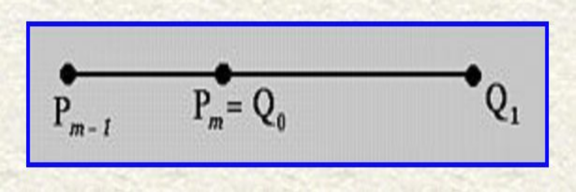




Bezier curve above is split at $t = \frac{u}{u+v}$.
Newly obtained pink Bezier curve has the control points A, B, C, and D.
The second curve (in blue) has the control points D, E, F, and G.

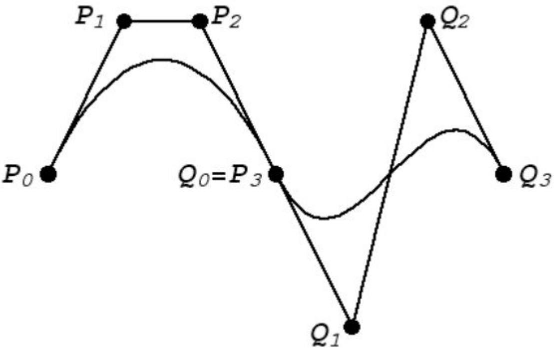
Joining Two Bézier Curves with C^1-Continuity

That a Bézier curve being tangent to its first and last legs provides us with a technique for joining two or more Bézier curves together for designing a desired shape. Let the first curve $C(u)$ be defined by $m+1$ control points $P_0, P_1, P_2, ..., P_m$. Let the second curve $D(u)$ be defined by $n+1$ control points $Q_0, Q_1, Q_2, \ldots, Q_n$. If we want to join these two Bézier curves together, then P_m must be equal to Q_0. This guarantees a C^0 continuous join. Recall that the first curve is tangent to its last leg and the second curve is tangent to its first leg. Consequently, to achieve a smooth transition, $P_{m-1}, P_m = Q_0$, and Q_1 must be on the same line such that the directions from P_{m-1} to P_m and the direction from Q_0 to Q_1 are the same. This is shown below.



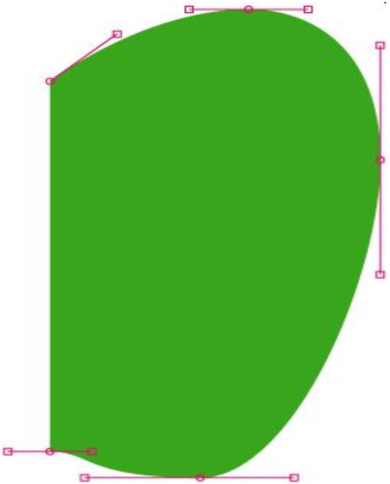
While joining two Bézier curves in this way looks smooth, it is still a C^0 join and is not yet C^1. However, it is G^1, because they have the same tangent vector directions. To achieve C^1 continuity, we have to make sure that the tangent vector at $u = 1$ of the first curve, $C'(1)$, and the tangent vector at $u = 0$ of the second curve, $D'(0)$, are identical. That is, the following must hold:

$$C'(1) = m(P_m - P_{m-1}) = D'(0) = n(Q_1 - Q_0)$$

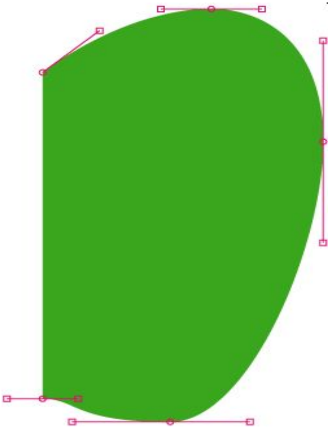


Two cubic Bezier curves – one with control points P_0, P_1, P_2, P_3 and the other with control points Q_0, Q_1, Q_2, Q_3 – that meet with matching first derivatives at their join. Here $Q_0 = P_3$ and $Q_1 - Q_0 = P_3 - P_2$.

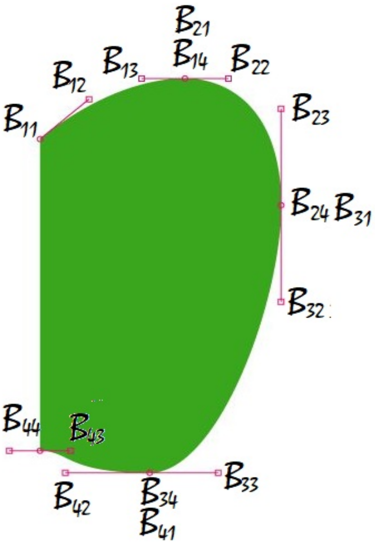
An application: Borders of a half apple



On the half apple borders, we first determine the anchor points and control points. In the figure above there are four sets of control points where each set represents a cubic Bezier curve.

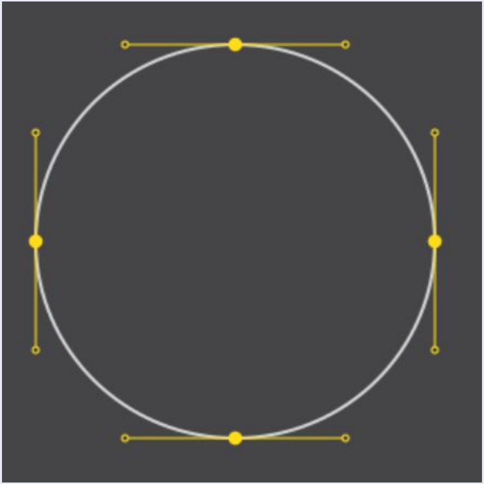


Note that the anchor points are the endpoints of the successive cubic Bezier curves. There are two control points associated with each anchor point such that 1) they are collinear with the anchor point, and 2) their distances to the anchor point is equal. These properties ensure the C^1 continuity. If only the collinearity is satisfied, however, the distances to the anchor point is not equal; it is called G_1 continuity.



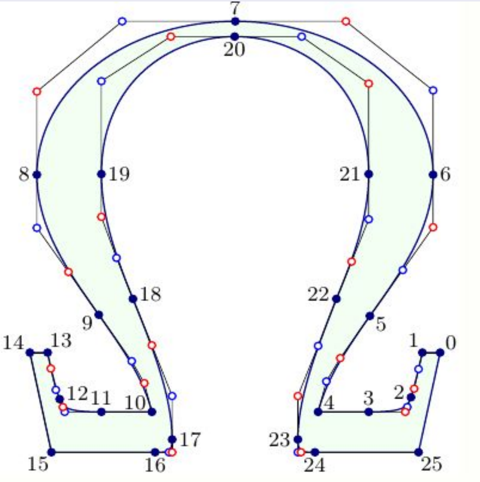
Bezier control points of the successive Bezier curves are labeled in the figure above.

Example 51



Anchor and control points for approximating a circle

Example 52



Anchor and control points for representing the letter Omega

Fréchet distance

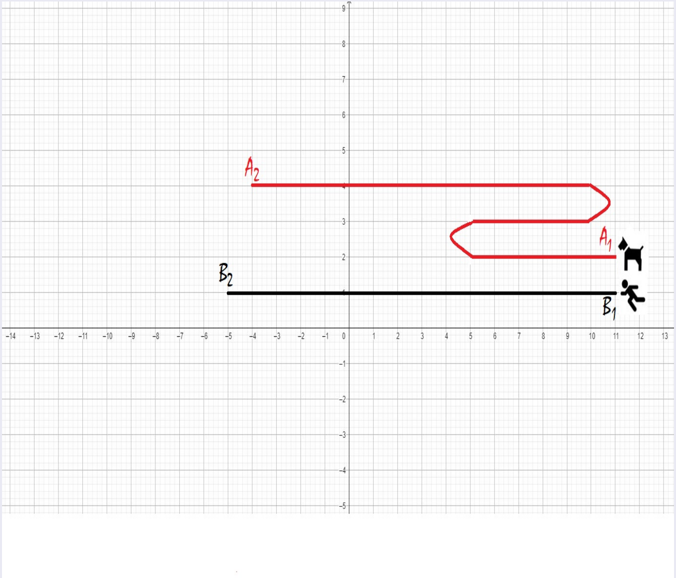
One of the fundamental problems when working with curves is curve fitting, or determining the Bézier that's closest to some source curve. How do we measure the distance between two curves? One meaningful metric is the Fréchet distance:
The **Fréchet distance** between two curves, sometimes also called the dog-leash distance, is defined as the minimum length of a leash required to connect a dog and its owner as they walk without backtracking along their respective curves from one endpoint to the other.

Formal definition of curves in $\mathbb{R}^2$:
Let $\alpha(t) : [0,1] \to [0,1]$ and $\beta(t) : [0,1] \to [0,1]$ be continuous parametrizations. Also let the curves A and B be continuous mappings from $[0,1]$ to $\mathbb{R}^2$.

$$F(A,B) = \inf_{\alpha,\beta} \max_{t\in[0,1]} \left\{ d\Big(A(\alpha(t)),\ B(\beta(t))\Big) \right\}$$

where d is the distance function.

Example 53

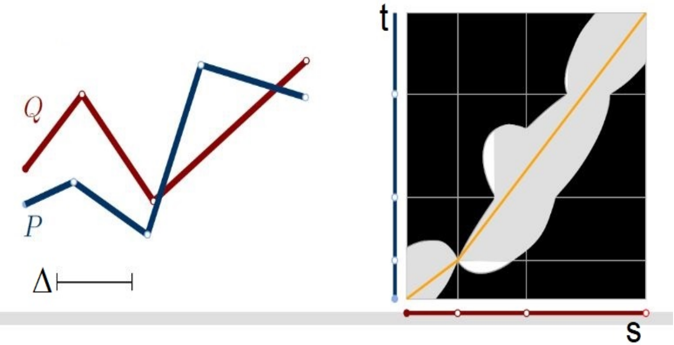


The Fréchet problem can be posed in two ways:
(1) what is the minimum leash length Δ for a given pair of trajectories.
(2) Given a leash of length Δ, can the man and the dog complete their paths.

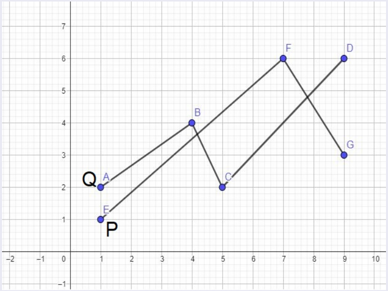
Second one is called the **decision problem**, which is simpler to compute.

The decision problem can be solved using a so called freespace diagram. A freespace diagram is a two-dimensional figure which given two trajectories P and Q and a Fréchet distance Δ, contains for each combination of points on P and Q whether $d(P, Q) \leq \Delta$, and if so, marks it as free (white). The freespace relates to the Fréchet decision problem as follows: if a monotonically increasing path in x and y can be found through the freespace, such that this path starts in the lower left corner and end in the top right corner, then the Fréchet distance between P and Q is lower than Δ, and vice versa.

Given two curves (i.e., two polygonal paths P and Q from [0,1] to the plane) the "free space" is the points (s, t) in the square $[0, 1] \times [0, 1]$ such that the distance from $P(s)$ to $Q(t)$ is less or equal than a prefixed Δ. The free space is the area not colored in black.

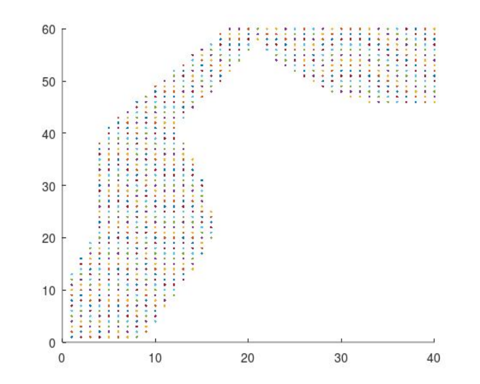


Example 54



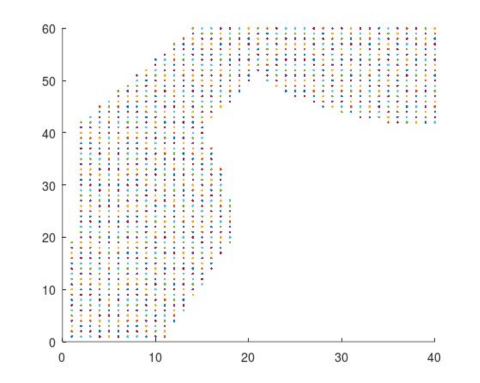
Taking 20 equally spaced samples from each line segment, P is represented by 40 sampling points: $P(1), \ldots, P(40)$. Likewise Q is represented by 60 points: $Q(1), \ldots, Q(60)$. A test for 3 unit length leash generates the following figure, which shows that 3 unit length leash is sufficient. A test for 4 unit length leash also results in a sufficient length. However, 2 unit length is not good.

Example 54 (cont.)



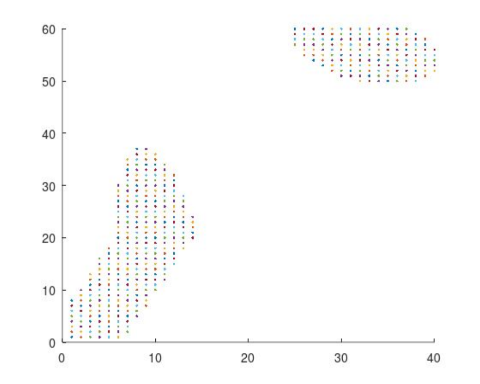
Free space diagram for 3 unit length leash

Example 54 (cont.)



Free space diagram for 4 unit length leash

Example 54 (cont.)



Free space diagram for 2 unit length leash

Codes in the OCTAVE environment

```
A=[1 2];
B=[4 4];
C=[5 2];
D=[9 6];
for k=1:20
Q{k}=(1-0.05*(k-1))*A+0.05*(k-1)*B;
end
for k=1:20
Q{k+20}=(1-0.05*(k-1))*B+0.05*(k-1)*C;
end
for k=1:20
Q{k+40}=(1-0.05*(k-1))*C+0.05*(k-1)*D;
end
```

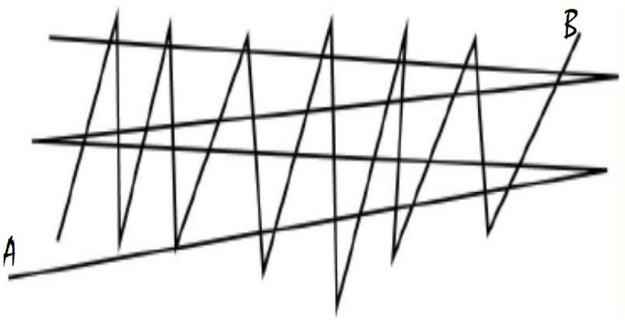
```
E=[1 1];
F=[7 8];
G=[9 3];
for k=1:20
P{k}=(1-0.05*(k-1))*E+0.05*(k-1)*F;
end
for k=1:20
P{k+20}=(1-0.05*(k-1))*F+0.05*(k-1)*G;
end

hold on
for n=1:40
for k=1:60
if norm(P{n}-Q{k})<2
plot(n,k)
end
end
end
```

Hausdorff Distance

We are given two point sets $A = \{a_1, a_2, ..., a_n\}$ and $B = \{b_1, b_2, ..., b_m\}$ in R^2. The one-sided Hausdorff distance from A to B is defined as:

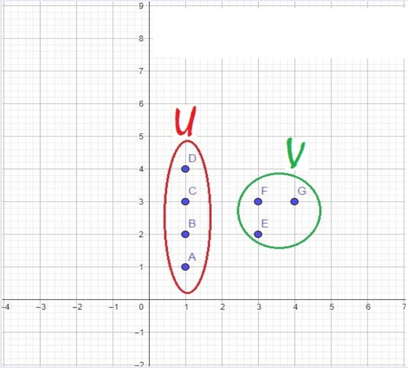
$$d_H(A, B) = \max_{a \in A} \min_{b \in B} ||a - b||$$

Hausdorff distance is not commutative. Two-sided Hausdorff distance between A and B is defined as maximum of $d_H(A, B)$ and $d_H(B, A)$. One downside of the Hausdorff distance is that it may call things similar which don't seem alike. For example, in next figure, the two shapes are not alike, but since any point in one is very close to some point in the other, the Hausdorff distance will be small.



Two curves not alike with small Hausdorff distance

Example 55



$$d_H(U,V) = \max_{a \in U} \min_{b \in V} ||a - b||$$

$$U = \{A, B, C, D\}, \quad V = \{E, F, G\}$$

Distance from A to the closest point in V: $\sqrt{5}$
Distance from B to the closest point in V: 2
Distance from C to the closest point in V: 2
Distance from D to the closest point in V: $\sqrt{5}$
Hausdorff distance from U to V: $d_H(U,V) = \max\{\sqrt{5}, 2\} = \sqrt{5}$

A given Bézier surface of degree (n, m) is defined by a set of $(n+1)(m+1)$ control points $k_{i,j}$ where $i = 0, \ldots, n$ and $j = 0, \ldots, m$. It maps the unit square into a smooth-continuous surface embedded within the space containing the $k_{i,j}$s.
A two-dimensional Bézier surface can be defined as a parametric surface where the position of a point p as a function of the parametric coordinates u, v is given by

$$p(u, v) = \sum_{i=0}^{n} \sum_{j=0}^{m} B_i^n(u)\, B_j^m(v)\, k_{i,j}$$

evaluated over the unit square, where

$$B_i^n(u) = \binom{n}{i} u^i (1-u)^{n-i}$$

is a basis Bernstein polynomial, and

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

Some properties of Bézier surfaces

∗ A Bézier surface will transform in the same way as its control points under all linear transformations and translations.

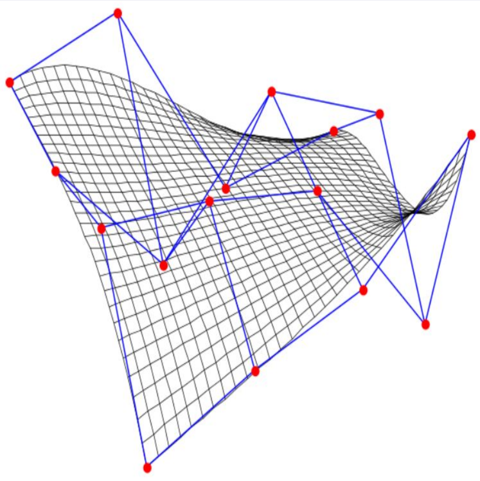
∗ All u=constant and v=constant lines in the (u, v) space, and - in particular - all four edges of the deformed (u, v) unit square are Bézier curves.

∗ A Bézier surface will lie completely within the convex hull of its control points, and therefore also completely within the bounding box of its control points in any given Cartesian coordinate system.

∗ The points in the patch corresponding to the corners of the deformed unit square coincide with four of the control points.

∗ However, a Bézier surface does not generally pass through its other control points.

Example 56



A 2d Bezier surface

Geogebra: A Tool for Bezier Curve Construction

GeoGebra is an interactive mathematics software for teaching mathematics and science, including algebra, geometry, calculus, and statistics. It is composed of an algebra window, a graphics window (2D and 3D graphics), an input bar, and includes a built-in environment spreadsheet, CAS (an advanced calculator), and statistics and calculus tools.

Geogebra is a freely-available, open source, multi-platform, and composed of easy-to-handle tools.

It support dynamic scenes.

It is an interactive platform to design Bezier curves.

Term projects of the course

Given the curve S, find a sequence of cubic Bezier curves connected en to end, call it B, such that the curve B is C^1 continuous and the Frechet distance between B and S is minimum.