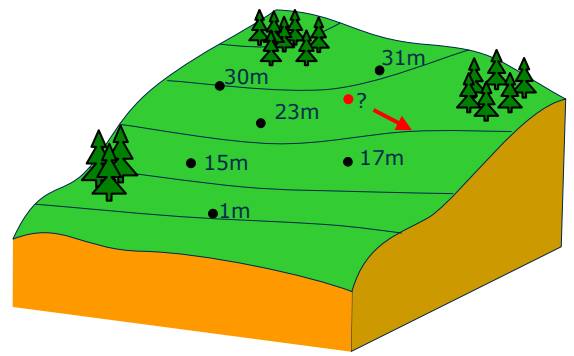


INTERPOLACIÓN

La interpolación es un proceso por el cual se define un valor en un punto cualquiera a partir de los valores conocidos en algunos puntos dados. Por ejemplo: tenemos temperatura, presión y viento en distintas localidades que cuentan con estaciones meteorológicas y queremos estimar el clima en un pueblo cualquiera. Suponiendo que las estaciones son razonablemente cercanas, surgen dos preguntas: ¿que estaciones considero? y ¿cómo calculo los valores desconocidos en función de los conocidos?



En general, tenemos un conjunto finito de nodos o puntos fijos con valores nodales asociados a cada uno de ellos: $\{(x_i, v_i)\}$, donde $x_i = \{x_i, y_i, z_i\}$ es la posición del i -ésimo punto y v_i el valor conocido en el punto fijo; se pretende encontrar el valor v asociado a un punto de coordenadas x cualquiera.

El proceso de interpolación define el valor asociado al punto variable. Podría, por ejemplo, asignarse el valor del punto más cercano o una combinación de valores cercanos o de todos los conocidos. En una sola dimensión es muy sencillo encontrar el punto más cercano o los dos nodos que “encierran” el punto buscado, pero en más dimensiones resulta más complicado.

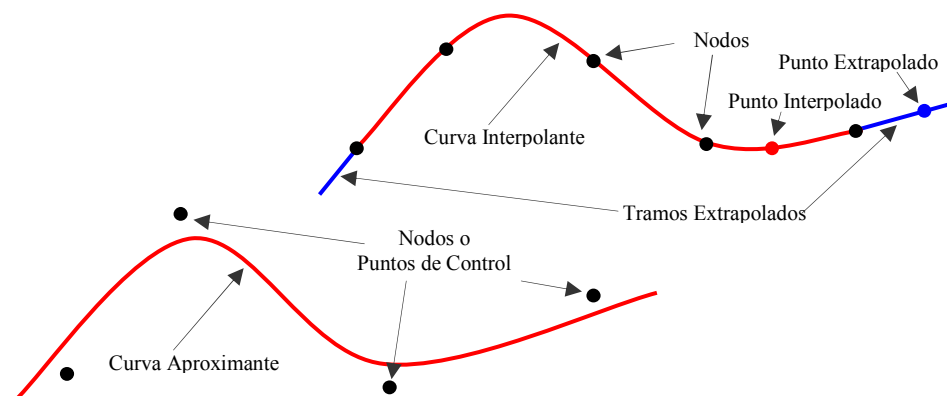
La forma estándar, para variables continuas, consiste en hacer una suma de valores ponderados:

$$v(\mathbf{x}) = \sum \alpha^i(\mathbf{x}) v_i \quad (\sum \alpha^i = 1)$$

El valor v en el punto variable x se calcula asignando, a cada punto fijo, un peso α^i que depende de la posición del punto variable. Normalmente se pretende que el peso de cada punto esté en relación inversa con la distancia al punto móvil, de ese modo el valor estará más influenciado por los puntos más cercanos. Pero hay muchas formas de definir los pesos. Es justamente, la selección de pesos la que define el método de interpolación.

Hay muchos métodos y no hay una única clasificación pero debemos mencionar algunas:

- Métodos Locales versus Globales:** Los métodos globales consideran todos los nodos. Un ejemplo es el método de [Kriging](#), muy utilizado en ciencias de la tierra, donde se construyen los pesos en relación inversa a las distancias con algunos supuestos de variabilidad. Para los métodos locales, en cambio, sólo se consideran los nodos cercanos, a los más lejanos se les asigna peso nulo, en general se utilizan estos últimos métodos pues son algorítmicamente más veloces, sobretodo cuando se dispone de un método rápido para encontrar el conjunto de nodos cercanos.
- Interpolación versus Extrapolación:** En la interpolación se asume que el punto, cuyo valor se busca, esta “entre los nodos”, en caso contrario es extrapolación. Este concepto (“entre”) es trivial en 1D: el punto está entre el nodo extremo izquierdo y el nodo extremo derecho, pero en más dimensiones es más complejo (envoltorio convexo) y en breve lo desarrollaremos.
- Interpolación versus Aproximación:** Supóngase móvil al punto incógnita, cuando el punto se mueve el valor cambia. Se denomina interpolación cuando el punto móvil adquiere el valor nodal cuando coincide con un nodo: $v(\mathbf{x}_j) = \sum \alpha^i(\mathbf{x}_j, \mathbf{x}_i) v_i = v_j$. El caso contrario recibe el nombre de aproximación y el ejemplo más conocido de aproximante es la regresión lineal o, en general, los procesos de mínimos cuadrados.



Interpolación afín

La interpolación afín es una extensión de la combinación afín de los puntos. Las coordenadas del punto variable se pueden obtener como combinación afín de los puntos fijos; una vez hallados los pesos, se aplican del mismo modo al valor asociado.

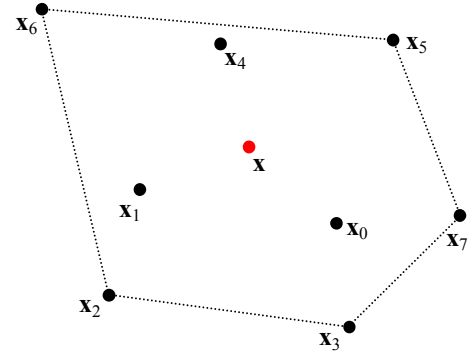
Del cálculo de la posición:

$$\mathbf{x} = \sum \alpha^i(\mathbf{x}) \mathbf{x}_i \quad (\sum \alpha^i = 1)$$

se obtienen los α^i y luego con esos mismos pesos se calcula:

$$v = \sum \alpha^i(\mathbf{x}) v_i$$

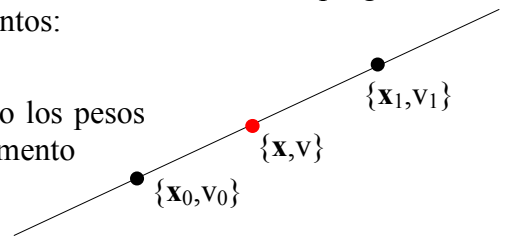
En la figura se representa un conjunto de puntos fijos \mathbf{x}_i y un punto variable \mathbf{x} , para el cual se pretende interpolar un valor, asignando pesos a los puntos fijos. Además se muestra también el envoltorio convexo o cápsula convexa o *convex-hull* del conjunto de puntos fijos. El envoltorio convexo se define como el menor convexo que contiene al conjunto de puntos; es la forma que tomaría una banda elástica envolviendo al conjunto. Enseguida veremos que ese es el límite de las posiciones que pueden obtenerse con pesos positivos.



El problema de interpolar un valor entre muchos puntos lo iremos resolviendo en forma progresiva. El caso más simple lo constituye la combinación afín 1D de dos puntos:

$$\mathbf{x} = \alpha^0 \mathbf{x}_0 + \alpha^1 \mathbf{x}_1, \text{ con } \alpha^0 + \alpha^1 = 1.$$

El resultado es un punto en la línea que une \mathbf{x}_0 con \mathbf{x}_1 . Cuando los pesos están limitados al rango $[0,1]$ el punto variable estará en el segmento que une los dos puntos y se garantiza que cualquier variable asociada asumirá también un valor intermedio. Si en cambio, algunos de los parámetros sale del intervalo $[0,1]$, el punto sale del segmento y el proceso se denomina extrapolación, el valor asociado también sale del intervalo $[v_0, v_1]$.



Para aclarar este último punto reescribimos la ecuación anterior, asumiendo que v representa cualquier coordenada (x, y o z) o cualquier valor asociado (temperatura, presión, humedad, color):

$$v = \alpha^0 v_0 + \alpha^1 v_1 = (1 - \alpha^1) v_0 + \alpha^1 v_1 = v_0 + (v_1 - v_0) \alpha^1.$$

Así vista la ecuación, cuando $\alpha^1 \in [0,1]$, queda claro que v adopta valores intermedios entre v_0 y v_1 , para cualquier variable que sea: coordenada o variable anexada. El punto se mantiene en el envoltorio convexo de los dos puntos fijos, que es simplemente el segmento $[\mathbf{x}_0, \mathbf{x}_1]$ y cualquier valor se mantiene en el rango de valores $[v_0, v_1]$. La combinación en este caso se denomina interpolación en sentido estricto o bien combinación convexa.

Si se pretende calcular una interpolación (o extrapolación) afín con dos puntos fijos, no pueden obtenerse puntos fuera de la recta que los une y por lo tanto la interpolación afín no está definida fuera de la recta; por definición utiliza los mismos pesos que la combinación afín, entonces está restringida a la expansión afín de los puntos fijos que es la recta que definen.

Pasemos a tres puntos fijos. En la figura está marcado \mathbf{x}_{01} como una combinación afín entre los puntos \mathbf{x}_0 y \mathbf{x}_1 . Ahora agregamos el punto \mathbf{x}_2 y hacemos $\mathbf{x} = \mathbf{x}_{012}$ combinando \mathbf{x}_{01} y el punto \mathbf{x}_2 .

La multiplicidad de índices (01, 012, ...) es sólo para indicar los pasos.

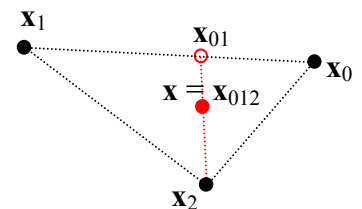
$$\mathbf{x}_{012} = \beta^{01} \mathbf{x}_{01} + \beta^2 \mathbf{x}_2; \quad \text{con } \beta^{01} + \beta^2 = 1$$

$$\mathbf{x}_{012} = \beta^{01} (\alpha^0 \mathbf{x}_0 + \alpha^1 \mathbf{x}_1) + \beta^2 \mathbf{x}_2 = \beta^{01} \alpha^0 \mathbf{x}_0 + \beta^{01} \alpha^1 \mathbf{x}_1 + \beta^2 \mathbf{x}_2 = \gamma^0 \mathbf{x}_0 + \gamma^1 \mathbf{x}_1 + \gamma^2 \mathbf{x}_2$$

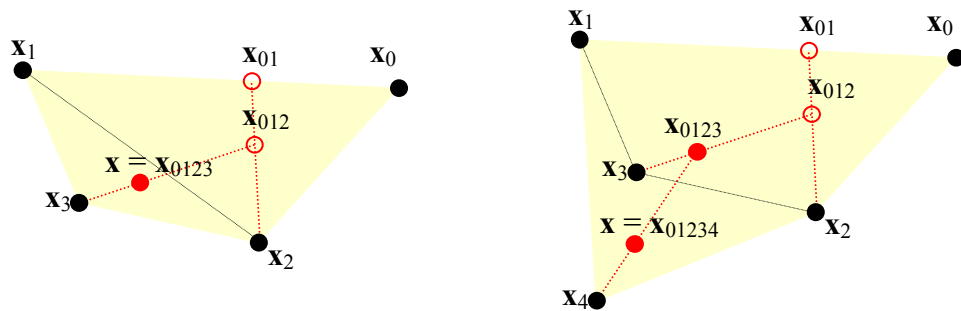
Podemos ver que se trata de una combinación afín porque los parámetros suman uno:

$$\gamma^0 + \gamma^1 + \gamma^2 = \beta^{01} \alpha^0 + \beta^{01} \alpha^1 + \beta^2 = \beta^{01} (\alpha^0 + \alpha^1) + \beta^2 = 1$$

Sin ser muy rigurosos podemos considerar demostrado que la combinación afín de combinaciones afines también es combinación afín de los puntos fijos. Además, si los coeficientes están entre cero y uno en cada paso, los coeficientes finales (γ) también, por ser productos de números entre cero y uno, por lo tanto: la combinación convexa de combinaciones convexas es una combinación convexa. Un sencillo análisis de los límites de cada combinación (hacerlo) nos revela que, en el caso convexo, el resultado estará dentro del triángulo marcado: el envoltorio convexo de los tres puntos.



Añadiendo al esquema un cuarto punto, quinto, etc. llegaremos fácilmente a la conclusión de que la combinación afín de un conjunto de puntos, cuando los parámetros están limitados al rango $[0,1]$, dará por resultado un punto entre los límites del envoltorio convexo del conjunto. Es por ello que éste tipo de combinación recibe el nombre de combinación convexa.

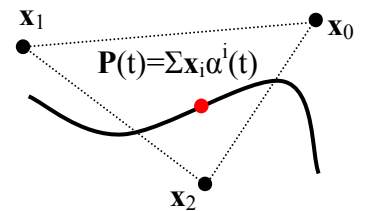


Obviamente, el resultado anterior es válido en cualquier cantidad de dimensiones y no solo en el plano. La extensión es trivial puesto que con x_i representamos cualquier coordenada o variable asociada.

Se puede pensar en el dibujo de arriba como la proyección plana de un resultado multidimensional. Con lo cual queda claro que la proyección ortogonal de una combinación afín es combinación afín de los puntos proyectados y por lo tanto está dentro del envoltorio convexo de los puntos proyectados (la proyección ortogonal es una transformación afín y por lo tanto preserva las combinaciones afines).

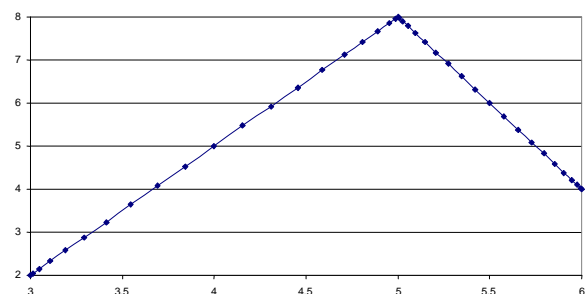
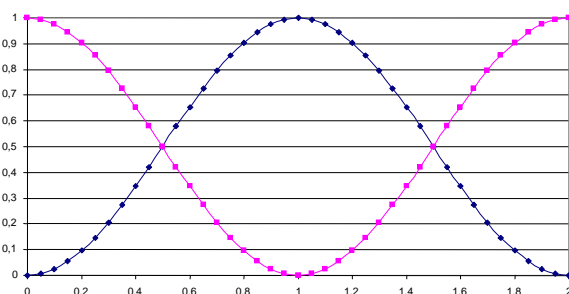
En el caso general de combinación afín (no-convexa), con al menos $d+1$ puntos independientes, se obtiene cualquier punto del espacio de soporte d dimensional. En el espacio tridimensional, con tres o más puntos coplanares (mas de tres coplanares tienen dependencia afín) se puede obtener cualquier punto del plano de los puntos fijos (el espacio soporte es el plano); con cuatro o más puntos no coplanares se puede obtener cualquier punto del espacio.

Si al conjunto de parámetros se le impone una restricción funcional se obtiene un subespacio. Por ejemplo, si cada parámetro α^i es función continua de otro parámetro t : $\alpha^i = \alpha^i(t)$, el resultado es una curva uniparamétrica como la de la figura. Si en 3D, $\alpha^i = \alpha^i(s,t)$ se obtiene una superficie biparamétrica. Este es el mecanismo que se utilizará mas adelante para construir curvas y superficies paramétricas (bilineales, Bézier, NURBS, etc.).



En general, los pesos pueden definirse mediante cualquier función de las coordenadas, mientras sumen uno. Por ejemplo, si tenemos dos puntos fijos x_0 y x_1 y hacemos $\alpha^0 = \sin^2(t)$ y $\alpha^1 = \cos^2(t)$, si permitimos que t tome cualquier valor real, tendremos un punto variable rebotando entre los nodos.

Ahora utilicemos el mismo estilo de funciones de peso no-lineales, pero para interpolar valores entre tres puntos en una recta 1D: $x_0 = 3$; $v_0 = 2$; $x_1 = 5$; $v_1 = 8$; $x_2 = 6$; $v_2 = 4$. Hacemos que t varíe entre 0 y 1 para el primer tramo y entre 1 y 2 para el segundo. Las funciones de forma serán $\alpha^1 = \sin^2(t \pi/2)$ que vale 1 en 1 y 0 en 0 y 2; $\alpha^0 = \alpha^2 = \cos^2(t \pi/2)$, que vale 1 para 0 y 2 y vale 0 para 1, tal como se muestra en la gráfica izquierda. Con esas funciones de forma, se calculan x y v , que se grafican a la derecha.



Por más que la “velocidad del recorrido” (derivada respecto de t) es variable, el valor asignado resulta lineal con las coordenadas de x . Esto es así, obviamente, pues ambos reciben el mismo tratamiento.

En general la interpolación afín será lineal cuando el valor dependa sólo de los nodos del simplejo (segmento, triángulo, tetraedro, ...) que rodea al punto y las funciones de peso de cada nodo valen uno en su nodo y cero en el resto. Ahora veremos el método estándar para construir funciones de peso lineales, que ellas mismas varían linealmente en el espacio.

Coordenadas Baricéntricas

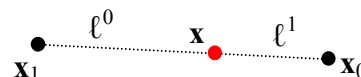
Para definir funciones de peso lineales, que recorren el espacio a velocidad constante, se utilizan las coordenadas baricéntricas o coordenadas de área o funciones de forma lineales, que son sólo una forma de denominar a los pesos lineales de un conjunto de nodos con independencia afín, por ejemplo: dos puntos, tres puntos no colineales o cuatro que no están en el mismo plano.

Los pesos de una combinación afín se denominan coordenadas porque la ecuación $\mathbf{x} = \sum \alpha^i \mathbf{x}_i$ define al punto variable como si los puntos fijos fuesen la base de un espacio vectorial. En particular, cuando los n pesos valen todos $1/n$, el punto que se obtiene es el baricentro (bari ~ peso o presión) o centro de gravedad o promedio de los puntos.

Partiremos del caso mas sencillo, el de dos puntos que definen un segmento $[\mathbf{x}_0, \mathbf{x}_1]$ y analicemos una variable en particular, que sabemos que varía en forma lineal entre los extremos: la longitud ℓ^1 del segmento $[\mathbf{x}_0, \mathbf{x}]$: en \mathbf{x}_0 vale $\ell^1_0 = \ell^1(\mathbf{x}_0) = 0$ y en \mathbf{x}_1 vale $\ell^1_1 = \ell^1(\mathbf{x}_1) = \ell$, entonces se puede interpolar:

$$\ell^1 = \alpha^0 \ell^1_0 + \alpha^1 \ell^1_1, \text{ con } \alpha^0 + \alpha^1 = 1$$

$$\ell^1 = \alpha^0 0 + \alpha^1 \ell; \Rightarrow \alpha^1 = \ell^1 / \ell$$



El peso de \mathbf{x}_1 es proporcional a la longitud ℓ^1 del segmento opuesto al punto.

Del mismo modo, o utilizando la suma unitaria, se obtiene que $\alpha^0 = \ell^0 / \ell$. Es decir que la relación (cociente) entre las longitudes de los segmentos opuestos es idéntica a la relación entre los pesos

Considerando longitudes con signo, la relación se cumplirá también para la extrapolación. Para definir el signo se debe establecer un orden de los puntos o una dirección preferencial, por ejemplo de cero a uno o el vector $\boldsymbol{\ell} = (\mathbf{x}_1 - \mathbf{x}_0)$, con $\ell = |\boldsymbol{\ell}|$; y entonces recurrimos a un producto escalar ordenado, que nos indica si la dirección es igual u opuesta:

$$\ell^1 = (\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\ell} / \ell$$

$$\alpha^1 = (\mathbf{x} - \mathbf{x}_0) \cdot \boldsymbol{\ell} / \ell^2$$

$$\ell^0 = (\mathbf{x}_1 - \mathbf{x}) \cdot \boldsymbol{\ell} / \ell$$

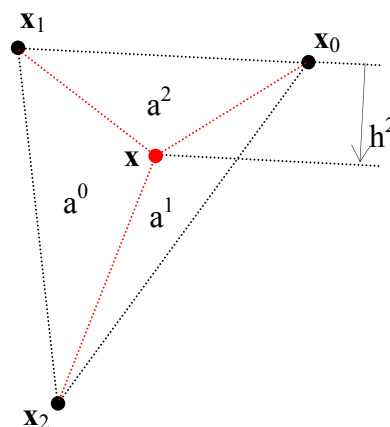
$$\alpha^0 = (\mathbf{x}_1 - \mathbf{x}) \cdot \boldsymbol{\ell} / \ell^2$$

En una extrapolación hay una longitud ℓ^i mayor que el total ℓ y una negativa, entonces habrá un peso α^i mayor que uno y el otro será negativo, pero los ℓ^i suman el total y los α^i suman uno.

En un algoritmo, se implementa la última ecuación, de modo que no se necesita invocar ningún condicional para averiguar si el punto móvil está entremedio de los nodos; además, dado que $\ell^2 = \boldsymbol{\ell} \cdot \boldsymbol{\ell}$, no es necesario calcular la longitud del segmento que es la raíz cuadrada del denominador.

Pasamos ahora a tres puntos no colineales. A la derecha se pueden ver los triángulos que define el punto interpolado \mathbf{x} con cada uno de los lados del triángulo que lo contiene, el área total es $a = a^0 + a^1 + a^2$, la denominación de cada uno corresponde al punto fijo opuesto.

El área a^2 (léase: “a2” y no: “a al cuadrado”) del triángulo opuesto a \mathbf{x}_2 : $(\mathbf{x}, \mathbf{x}_0, \mathbf{x}_1)$ varía linealmente: vale 0 cuando el punto variable está en la línea $(\mathbf{x}_0, \mathbf{x}_1)$, en particular cuando coincide con \mathbf{x}_0 ($a^2_0 = 0$) o con \mathbf{x}_1 ($a^2_1 = 0$) y es igual al área total a , cuando el punto variable está en una línea paralela a la anterior y que pasa por \mathbf{x}_2 , en particular cuando coincide con ese punto ($a^2_2 = a$). Es lineal porque tiene base fija $|\mathbf{x}_1 - \mathbf{x}_0|$ y varía igual que su altura h^2 (léase: “h2” y no: “h al cuadrado”). Interpolando, entonces, para calcular el área a^2 :



$$a^2 = \alpha^0 a^2_0 + \alpha^1 a^2_1 + \alpha^2 a^2_2 = \alpha^0 0 + \alpha^1 0 + \alpha^2 a \Rightarrow \alpha^2 = a^2/a.$$

Del mismo modo, $\alpha^i = a^i/a$ para cualquier vértice. Esta es la razón por la que también se denominan coordenadas de área.

Considerando áreas orientadas (con signo) hacemos que la relación también resulte válida para la extrapolación. Pero aquí también hay que definir un ordenamiento preferencial que, en este caso, se puede obtener a partir de un producto vectorial realizado con cierto orden. El producto vectorial es el área orientada del paralelogramo que definen los vectores, el área de un triángulo es la mitad, pero los cocientes entre áreas hacen innecesarias las divisiones por dos.

Definimos:

$$\mathbf{a} = (\mathbf{x}_1 - \mathbf{x}_0) \times (\mathbf{x}_2 - \mathbf{x}_0)$$

$$\mathbf{a}^0 = (\mathbf{x}_1 - \mathbf{x}) \times (\mathbf{x}_2 - \mathbf{x})$$

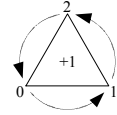
$$\mathbf{a}^1 = (\mathbf{x}_2 - \mathbf{x}) \times (\mathbf{x}_0 - \mathbf{x})$$

$$\mathbf{a}^2 = (\mathbf{x}_0 - \mathbf{x}) \times (\mathbf{x}_1 - \mathbf{x})$$

Notar el orden, que puede (suele) generalizarse como:

$$\mathbf{a}^i = (\mathbf{x}_{i+1} - \mathbf{x}) \times (\mathbf{x}_{i+2} - \mathbf{x})$$

donde las sumas de índices se hacen módulo tres ((i+1)%3; ej:(2+1)%3=0)



La dirección del vector \mathbf{a} será la que define las áreas positivas. Para concluir definimos:

$$\alpha^i = \mathbf{a}^i \cdot \mathbf{a} / \mathbf{a}^2$$

Lo mismo sucede con volúmenes de tetraedros. Para ese caso, los volúmenes orientados de los tetraedros son productos triples (volumen del paralelepípedo dividido por seis, pero el seis se cancela en las divisiones):

$$v = ((\mathbf{x}_1 - \mathbf{x}_0) \times (\mathbf{x}_2 - \mathbf{x}_0)) \cdot (\mathbf{x}_3 - \mathbf{x}_0) \quad \alpha^i = ((\mathbf{x}_{i+1} - \mathbf{x}) \times (\mathbf{x}_{i+2} - \mathbf{x})) \cdot (\mathbf{x}_{i+3} - \mathbf{x}) / v$$

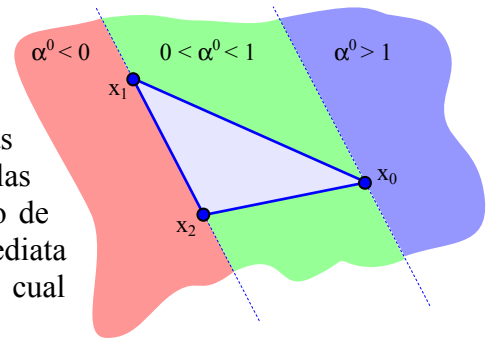
con las sumas de índices en módulo cuatro.

Es muy sencillo hacer los algoritmos de modo que realicen estas operaciones en forma ordenada.

Se puede recordar así: el peso de un nodo en un punto variable es igual al cociente entre el área opuesta y la total. (O longitudes en 1D o volúmenes en 3D)

En el caso de tres puntos no alineados, las líneas en las cuales una coordenada de área resulta constante se denominan líneas isoparamétricas; dado que la base del triángulo es siempre el segmento opuesto, éstas son las líneas de altura constante, las paralelas a dicho segmento. Para cuatro puntos no coplanares, las superficies isoparamétricas serán los planos paralelos a la cara opuesta a un nodo.

En la figura, La recta $(\mathbf{x}_1, \mathbf{x}_2)$ divide al plano en dos, podemos ver que α^0 es positivo para cualquier posición de punto variable en el semiplano que contiene a \mathbf{x}_0 , es nula en la recta y negativa en el semiplano opuesto y, como ya vimos, es constante en las rectas paralelas a la mencionada. La simpleza del cálculo de las áreas permite averiguar fácilmente si un punto está o no dentro de un triángulo (¡sin hacer las divisiones!) y nos da una pista inmediata de cómo movernos para averiguar, en una triangulación, cual triángulo contiene al punto variable (lo veremos más adelante).



En 3D las superficies isoparamétricas serán los planos paralelos a la cara opuesta a un nodo.

Las aplicaciones más interesantes no son nunca las más simples, normalmente se cuenta con varios puntos definidos (sin independencia afín). Por ejemplo, se suelen obtener datos del clima (más de un valor: presión, temperatura, dirección del viento, velocidad, etc.) en distintos nodos o puntos de muestreo, que en este caso son estaciones meteorológicas. Para obtener en forma aproximada un valor en una localidad cualquiera se puede realizar una interpolación lineal, pero no queda claro cuales estaciones meteorológicas deben tenerse en cuenta, cuales tienen peso no-nulo y que peso tiene cada una. Un método estándar consiste en realizar una división del envoltorio convexo en triángulos no solapados y cuyos vértices son los nodos, esto es: una triangulación del dominio. Para calcular el valor en un punto cualquiera se pueden utilizar sólo las estaciones del triángulo que encierra al punto. Ésta técnica se conoce como interpolación poli-lineal, las variantes están dadas por los distintos métodos para realizar la triangulación.

Hay muchas técnicas (lineales o no) y casi todas se basan en triangulaciones u otras divisiones estratégicamente simples del dominio.

Interpolación afín en coordenadas homogéneas e interpolación hiperbólica

La interpolación afín en coordenadas homogéneas no tiene nada de extraño, la coordenada w también resulta interpolada, como cualquier variable más. Lo extraño aparece al hacer la división por w , es decir la proyección. Por simplicidad, llamemos $\{x, w\}$ a las coordenadas homogéneas e $y = x/w$ a las coordenadas espaciales del punto.*

$$\begin{aligned} \{x, w\} &= \alpha^0 \{x_0, w_0\} + \alpha^1 \{x_1, w_1\} = \{\alpha^0 x_0, \alpha^0 w_0\} + \{\alpha^1 x_1, \alpha^1 w_1\} & (\alpha^0 + \alpha^1 = 1) \\ y &= (\alpha^0 x_0 + \alpha^1 x_1) / (\alpha^0 w_0 + \alpha^1 w_1) \\ &= (\alpha^0 w_0 x_0 / w_0 + \alpha^1 w_1 x_1 / w_1) / (\alpha^0 w_0 + \alpha^1 w_1) \\ &= (\alpha^0 w_0 y_0 + \alpha^1 w_1 y_1) / (\alpha^0 w_0 + \alpha^1 w_1) \\ &= \beta^0 y_0 + \beta^1 y_1 \end{aligned}$$

Los pesos “proyectados” (β) son distintos que los pesos en el espacio homogéneo (α):

$$\beta^i = \alpha^i w_i / \sum_j \alpha^j w_j = \alpha^i w_i / w \quad (\text{notar que los } \beta_i \text{ también suman uno})$$

La combinación que se hizo aquí entre dos puntos se extiende en forma trivial a más puntos.

Hagamos un par de ejemplos. Tenemos dos puntos en coordenadas homogéneas: $x_0 = \{0,0,0,1\}$ y $x_1 = \{1,0,0,1\}$; supongamos que los pesos son $\alpha^0 = \alpha^1 = 1/2$; el resultado es el punto $x = \{1/2,0,0,1\}$ que equivale a $y = \{1/2,0,0\}$; en este caso, ambos w valen uno y el resultado no es sorprendente. Ahora, si el segundo punto es $x_1 = \{2,0,0,2\}$, es el mismo que antes pero con distintas coordenadas homogéneas (mismo y , distinto x), el resultado ahora es $x = \{1,0,0,3/2\}$ con $y = \{2/3,0,0\}$. Es como que en el segundo caso, el segundo punto tiene un mayor peso cuanto mayor sea su w .

Puede pensarse que los w_i son pesos adicionales en el cálculo del promedio ponderado. De hecho, en algunos textos (y en particular para curvas y superficies NURBS) se introducen como pesos en lugar de coordenadas homogéneas.

La relación entre los pesos proyectados β^i y originales α^i está alterada con la relación de los w_i :

$$\beta^i / \beta^j = \alpha^i w_i / \alpha^j w_j$$

Aquí se ve que **la transformación proyectiva no preserva la combinación afín.**

El mismo problema se presenta al interpolar los resultados en la pantalla (posición, color, normal o textura) cuando en realidad lo que se pretende es interpolar en el espacio, antes de la división perspectiva. La interpolación hiperbólica[†] es la interpolación en el espacio proyectivo que corresponde a una lineal en coordenadas homogéneas.

En el caso de la proyección perspectiva, en el monitor por ejemplo, la coordenada visual z cumple un papel similar al w del espacio proyectivo, pero se debe recordar que el plano de la imagen se identifica con el plano near, de donde z_{near} es como la “unidad” $w=1$ del plano proyectivo; entonces, en las fórmulas anteriores z/z_{near} cumple exactamente el mismo papel que w pero, dado que son todos cocientes, z_{near} se cancela y se usa solo z . En la web ese tema se encuentra como “perspective correct” (interpolation o texture mapping).

Ejemplo: se quiere averiguar donde se proyecta el punto medio entre x_0 y x_1 ($\alpha^0 = \alpha^1 = 1/2$), se conocen y_0 e y_1 ; se calcula $\beta^1 = (w_1/2) / [(w_0+w_1)/2] = w_1 / (w_0+w_1)$; y luego $y = (1-\beta^1)y_0 + \beta^1 y_1 = (z_0 y_0 + z_1 y_1) / (z_0 + z_1)$.

Si el problema es el inverso: se está operando píxel a píxel en la pantalla, para ello se barre entre un punto y el otro con β^1 variable entre cero y uno y se necesita un valor correspondiente, interpolado con α^1 en el espacio real. Bastará con aplicar las relaciones anteriores, considerando suma uno:

$$\alpha^0 / \alpha^1 = (\beta^0 / z_0) / (\beta^1 / z_1) \quad \Rightarrow \quad \alpha^0 = (\beta^0 / z_0) / (\beta^0 / z_0 + \beta^1 / z_1); \quad \alpha^1 = (\beta^1 / z_1) / (\beta^0 / z_0 + \beta^1 / z_1)$$

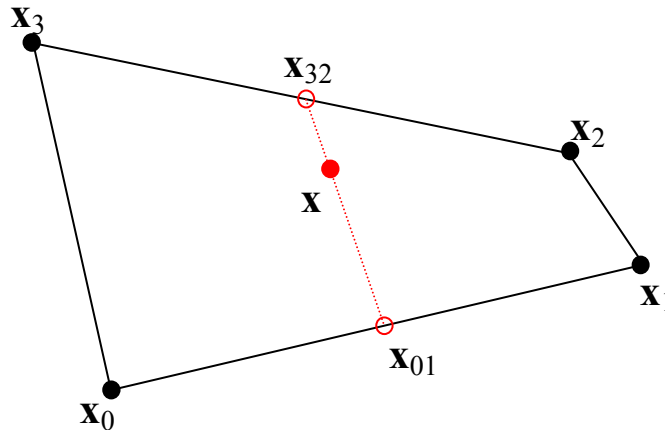
Esas ecuaciones son las que se utilizan para interpolar o pegar texturas con la perspectiva corregida.

* Aquí no utilizamos la convención de Einstein, no hay suma en i y la suma en j está explícita.

† En la geometría plana de Euclides las paralelas no se cortan, en la hiperbólica (proyectiva) todo par de rectas distintas se corta en un punto y en la esférica, todo par de rectas distintas se corta en dos puntos

Interpolación bilineal

Se denomina así a una forma particular de expresar, en función de sólo dos parámetros, los pesos de la combinación afín de cuatro puntos que forman un cuadrángulo, ya sea en el plano o en el espacio.



La combinación bilineal consiste en tomar la misma combinación afín en dos líneas opuestas, con uno de los parámetros y luego interpolar los dos puntos resultantes con el otro parámetro:

$$x_{01} = (1-u)x_0 + ux_1$$

$$x_{32} = (1-u)x_3 + ux_2$$

$$x = (1-v)x_{01} + vx_{32} = (1-v)(1-u)x_0 + (1-v)ux_1 + v ux_2 + v(1-u)x_3$$

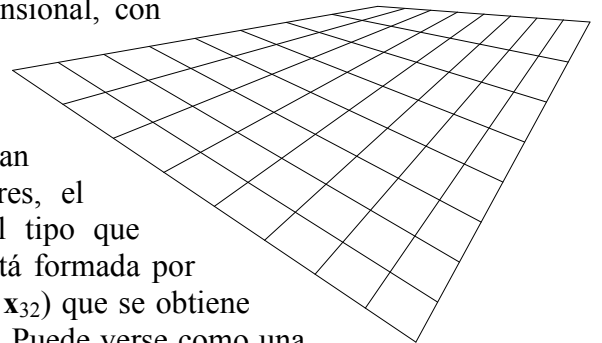
El resultado es idéntico si primero se hace la combinación con v en los segmentos (x_0, x_3) y (x_1, x_2) , para luego combinar con u los puntos resultantes (probar).

La denominación de bilineal proviene del uso de los parámetros, normalmente identificados con las coordenadas de un cuadrado unitario en el espacio de parámetros. Un polinomio formado así, con productos de monomios lineales, sin ningún término cuadrático en u ni en v , se denomina bilineal.

Los cuatro parámetros de la combinación afín resultante: $\{(1-v)(1-u), (1-v)u, v u, v(1-u)\}$ suman uno y, si $u, v \in [0,1]^2$ también estarán en el rango $[0,1]$. El resultado es una interpolación afín no-lineal de los cuatro puntos y valores; con una receta para la asignación de los pesos.

En computación gráfica se utiliza este tipo de interpolación para mapear texturas o interpolar colores en cuadriláteros, el resultado es más suave que el obtenido dividiendo en dos triángulos.

Es muy importante recalcar que en el espacio tridimensional, con cuatro puntos no coplanares, se obtiene una superficie reglada (bilineal), que tiene por bordes a los cuatro segmentos. Las coordenadas baricéntricas de cuatro puntos tienen tres grados de libertad porque suman uno; aquí, en cambio, hay sólo dos parámetros libres, el conjunto tiene una restricción paramétrica $\alpha^i(u,v)$ del tipo que mencionamos antes. La superficie es reglada porque está formada por una sucesión infinita de segmentos rectos como el (x_{01}, x_{32}) que se obtiene barriendo u entre cero y uno (lo mismo sucederá con v). Puede verse como una pompa de jabón entre cuatro alambres rectos no coplanares.

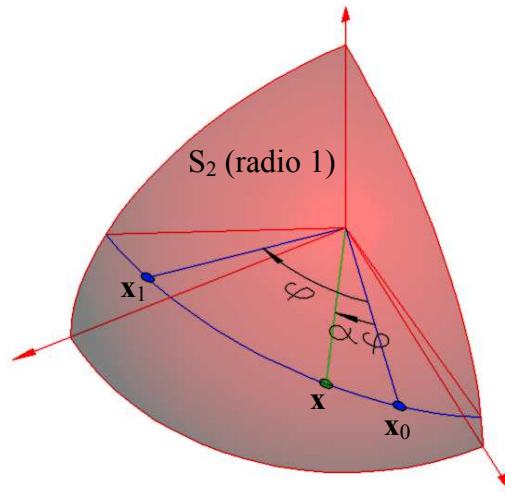


La función es biunívoca en el interior, excepto cuando los cuatro puntos son coplanares y el cuadrilátero cóncavo. No analizaremos en detalle este caso, basta con decir que se obtienen, como siempre, puntos del *convex-hull*, pero fuera del cuadrilátero cóncavo se pueden obtener con dos conjuntos de parámetros distintos. (Pensarlo como una superficie bilineal 3D aplastada o proyectada)

El mismo concepto se extiende a un "cubo" y se utiliza la interpolación trilineal para mapear el interior de un cubo unitario en un hexaedro o cuboide de seis caras bilineales en el espacio. Pero eso no suele usarse en Computación Gráfica sino en Cálculo Numérico (Elementos Finitos). En CG podría utilizarse para mapear texturas tridimensionales.

Interpolación esférica lineal o *Slerp* (*Spherical Linear Interpolation*)

Se trata de un método de interpolación entre dos puntos de una esfera unitaria y se utiliza para interpolar rotaciones, orientaciones o normales, que se describen mediante vectores unitarios o, según veremos más adelante, mediante cuaterniones. Se interpolan coordenadas pero no variables asociadas.



Se denomina S_n (o n -esfera) a la generalización dimensional del concepto de esfera unitaria. En $n+1$ dimensiones, es una superficie n -dimensional, cuyos puntos distan una unidad del origen. La 2-esfera S_2 es la superficie de una esfera común de radio 1 y centrada en el origen en 3D.

En cualquier superficie, las curvas que unen puntos, minimizando la longitud del recorrido, se llaman geodésicas. Son el equivalente de las rectas en el plano.

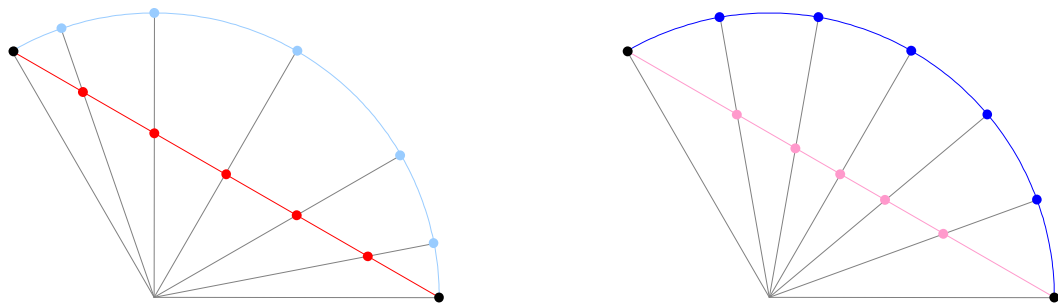
El camino más corto entre dos puntos de una esfera, pero recorriendo puntos de la superficie, es un arco de circunferencia máxima. Las circunferencias máximas son las que comparten el centro y el radio con la esfera. Cualquier otra circunferencia en la superficie tendrá un radio menor. Por ejemplo: de los paralelos de la tierra, solo el ecuador es un círculo máximo; mientras que todos los meridianos lo son. En S_n las geodésicas están centradas en el origen y tienen radio unitario.

Si dos puntos de la esfera, x_0 y x_1 , no son antipodales (no están alineados con el origen), habrá una sola circunferencia máxima que los une. Si son antipodales, habrá infinitas. De Buenos Aires a Sydney hay un único camino de mínima longitud, pero del Polo Norte al Polo Sur hay infinitos meridianos con la misma longitud. Lo mismo sucede entre Formosa y Taiwán; lo cual es doblemente raro: 1) La mayor parte de la tierra es antípoda del océano y 2) La isla de Taiwán también se llama Formosa.

La interpolación esférica entre dos puntos es un punto intermedio, pero en la esfera. Lo que se hace es una interpolación lineal de los ángulos centrales: Si $\varphi \in [0,180)$ es el ángulo central entre los puntos, con vértice en el origen y $\alpha \in [0,1]$ el parámetro de la interpolación; queremos un punto, en el plano del ángulo, pero angularmente distanciado $\alpha\varphi$ del primer punto y en dirección al segundo punto:

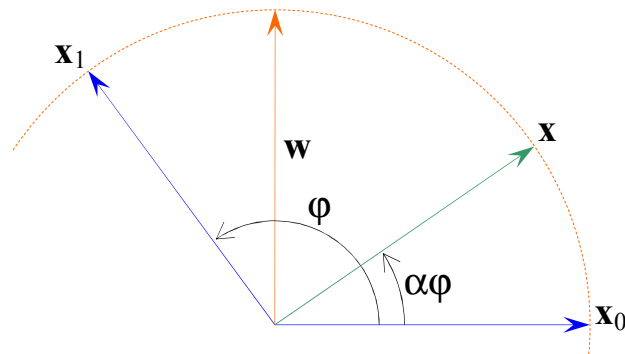
$$\mathbf{x} = \text{slerp}(\mathbf{x}_0, \mathbf{x}_1, \alpha)$$

La interpolación esférica es lineal en la superficie de la esfera pero no en el espacio, el camino recorrido por el punto móvil es un arco de longitud $\alpha\varphi$ (porque el radio es uno). El camino recorrido por unidad de "tiempo" ($d/d\alpha$) es constante, pero no el vector velocidad, que no cambia su módulo pero sí su dirección (aceleración centrípeta).



En las figuras de arriba se hace evidente la diferencia entre una interpolación lineal de direcciones y una esférica. La diferencia se puede apreciar visualmente al interpolar normales para iluminación.

Todo el proceso de interpolación descansa gráficamente en el plano del ángulo, definido por el origen y los dos puntos dados; por lo tanto haremos los cálculos en ese plano.



Partimos de las coordenadas x_0 y x_1 de los puntos dados y del parámetro variable α . Debemos calcular el ángulo φ entre los puntos y luego el punto interpolado x .

En primer lugar, dado que los módulos son unitarios, notemos que:

$$\varphi = \text{acos}(x_0 \cdot x_1)$$

Podemos representar cualquier punto del plano mediante una combinación lineal del par ortonormal $\{x_0, w\}$, donde w es un vector perpendicular a x_0 en la dirección de x_1 , que usaremos de intermediario.

$$x = \cos(\alpha\varphi) x_0 + \text{sen}(\alpha\varphi) w$$

$$x_1 = \cos(\varphi) x_0 + \text{sen}(\varphi) w$$

De la segunda ecuación podemos extraer w y usarlo en la primera.

$$x = \cos(\alpha\varphi) x_0 + \frac{\text{sen}(\alpha\varphi)}{\text{sen}(\varphi)} (x_1 - \cos(\varphi) x_0)$$

$$\begin{aligned} \text{sen}(\varphi) x &= \cos(\alpha\varphi) \text{sen}(\varphi) x_0 + \text{sen}(\alpha\varphi) (x_1 - \cos(\varphi) x_0) \\ &= [\cos(\alpha\varphi) \text{sen}(\varphi) - \text{sen}(\alpha\varphi) \cos(\varphi)] x_0 + \text{sen}(\alpha\varphi) x_1 \\ &= \text{sen}(\varphi - \alpha\varphi) x_0 + \text{sen}(\alpha\varphi) x_1 \\ &= \text{sen}[(1 - \alpha)\varphi] x_0 + \text{sen}(\alpha\varphi) x_1 \end{aligned}$$

$$x = \frac{\text{sen}[(1 - \alpha)\varphi]}{\text{sen}(\varphi)} x_0 + \frac{\text{sen}(\alpha\varphi)}{\text{sen}(\varphi)} x_1$$

Para evitar la posible división por cero o la inestabilidad, cuando φ sea muy pequeño, hay que hacer una aproximación lineal. Cuando φ tiende a cero, la ecuación anterior tiende a la simple interpolación lineal (el seno tiende al ángulo o el arco tiende a la cuerda):

$$x = (1 - \alpha) x_0 + \alpha x_1$$

A nivel de implementación, para evitar la inestabilidad bastará saber si se pueden equiparar el ángulo y su seno. Dado que el mayor ángulo involucrado es φ , basta saber si $|\varphi - \text{sen}(\varphi)| < \varepsilon$, donde ε es un pequeño valor que depende de la precisión numérica utilizada. Haciendo la expansión en serie de Taylor, se ve fácilmente que:

$$|\varphi - \text{sen}(\varphi)| \sim \varphi^3 / 6,$$

podemos entonces comparar φ con 10^{-3} o 10^{-6} en simple o doble precisión respectivamente.

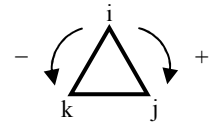
Cuaterniones

Si bien tienen múltiples aplicaciones, la única utilización práctica de los cuaterniones en CG es para representar (e interpolar) rotaciones y efectivamente son muy utilizados para eso en videojuegos.

Un cuaternión se forma como un vector de cuatro dimensiones cuya base son los números $\{1, i, j, k\}$. El número 1 es la unidad real estándar; i, j y k son extensiones del imaginario puro i , de los números complejos. El cuadrado de cualquiera de ellos es -1 . Los productos mutuos (más allá del trivial 1) son:

$$i^2 = j^2 = k^2 = -1; \quad ij = -ji = k; \quad jk = -kj = i; \quad ki = -ik = j$$

Notar la antisimetría. La regla mnemotécnica es: ijk en orden circular, cada par sucesivo da el siguiente y en orden inverso dan negativos.



Un cuaternión es entonces una combinación lineal $q = ai + bj + ck + d$, donde a, b, c y d son números reales. Tiene una componente real: d y tres imaginarias: a, b y c . También podemos representarlo como $q = \langle d; \mathbf{u} \rangle$ donde \mathbf{u} es el vector tridimensional $\{a, b, c\}$ que engloba las tres componentes imaginarias, en este caso se habla de una componente escalar y una vectorial o cuaternión pura.

Con esa notación podemos representar escalares y vectores como cuaterniones haciendo la distinción:

$\langle d, \mathbf{0} \rangle$ representa el escalar d ;

$\langle 0, \mathbf{u} \rangle$ representa el vector o cuaternión puro \mathbf{u} .

La operatoria de los cuaterniones es idéntica a la de los complejos. Para el producto hay que tener en cuenta los productos mutuos de los números que conforman la base. Para poder usar la notación estándar con los vectores (\mathbf{u}) reservaremos el término “producto escalar” y el símbolo “ \cdot ” para los vectores estándar y definimos el producto (a secas) entre cuaterniones, que se desarrolla distribuyendo los productos de a pares (todos contra todos) que, ya agrupados, queda:

$$q_1 q_2 = \langle (d_1 d_2 - \mathbf{u}_1 \cdot \mathbf{u}_2); (d_1 \mathbf{u}_2 + d_2 \mathbf{u}_1 + \mathbf{u}_1 \times \mathbf{u}_2) \rangle$$

Se puede ver que la no-commutatividad del producto está sólo en el producto vectorial, pues si se invierten los factores, resulta opuesto. La conmutatividad sólo se da, entonces, cuando las partes vectoriales son paralelas (\mathbf{u} opuestas) y el producto vectorial es nulo.

El conjugado de $q = \langle d; \mathbf{u} \rangle$ es otro cuaternión con la parte vectorial opuesta:

$$q^* = \langle d; -\mathbf{u} \rangle$$

Se puede ver muy fácilmente que:

$$qq^* = q^*q = a^2 + b^2 + c^2 + d^2 = d^2 + \mathbf{u}^2 \quad (\text{es un escalar})$$

La norma o magnitud de un cuaternión se define como en el caso complejo:

$$\|q\| = \sqrt{qq^*} = \sqrt{a^2 + b^2 + c^2 + d^2} = \sqrt{d^2 + \mathbf{u}^2}$$

Mirando la fórmula anterior es fácil deducir que:

$$q^{-1} = \frac{q^*}{\|q\|^2}; \quad \text{siendo: } qq^{-1} = q^{-1}q = 1. \quad (\text{Conmutativo por ser antiparalelos los vectores})$$

Un cuaternión se puede normalizar, igual que un vector, dividiéndolo por su módulo:

$$q_1 = \frac{q}{\|q\|}$$

Para un cuaternión unitario: $q^{-1} = q^*$.

Todo cuaternión unitario se puede escribir como: $q_1 = \langle \cos(\alpha); \mathbf{u}_1 \sin(\alpha) \rangle$; donde \mathbf{u}_1 es, a su vez, un vector unitario:

$$q_1 = \langle \cos(\alpha); \mathbf{u}_1 \sin(\alpha) \rangle; \quad \mathbf{u}_1^2 = 1 \Rightarrow \|q_1\| = \sqrt{\cos^2(\alpha) + \sin^2(\alpha) \mathbf{u}_1^2} = 1$$

Rotación con cuaterniones:

Analicemos la siguiente expresión que combina el vector \mathbf{v} con el cuaternión r :

$$\underline{\mathbf{v}} = r \mathbf{v} r^{-1} = r \langle 0, \mathbf{v} \rangle r^{-1}$$

En primer lugar vemos que el resultado es el mismo si normalizamos el cuaternión:

$$r^{-1} = (\|r\| r_1)^{-1} = \|r\|^{-1} r_1^{-1} \Rightarrow (\|r\| r_1) \mathbf{v} (\|r\| r_1)^{-1} = r_1 \mathbf{v} r_1^{-1},$$

ésto simplifica la tarea, pues $r_1^{-1} = r_1^*$. Usaremos directamente cuaterniones unitarios.

En segundo lugar, notemos que nos hemos adelantado al llamar \underline{v} al resultado, pues aun no sabemos que tipo de cuaternión es. Enseguida veremos que, efectivamente, es un vector.

Apliquemos esa expresión a un vector $\underline{v}_{//}$ paralelo a \underline{u} ($\underline{v}_{//} = \gamma \underline{u}$). Usando la fórmula del producto y siendo $\underline{u}^2 = 1$:

$$\begin{aligned} \underline{v}_{//} &= \langle \cos(\alpha); \underline{u} \sin(\alpha) \rangle \langle 0; \gamma \underline{u} \rangle \langle \cos(\alpha); -\underline{u} \sin(\alpha) \rangle = \\ &= \langle -\gamma \sin(\alpha); \gamma \cos(\alpha) \underline{u} \rangle \langle \cos(\alpha); -\underline{u} \sin(\alpha) \rangle = \\ &= \langle [\gamma \sin(\alpha) \cos(\alpha) - \gamma \sin(\alpha) \cos(\alpha)]; \gamma \underline{u} [\cos^2(\alpha) + \sin^2(\alpha)] \rangle = \langle 0; \gamma \underline{u} \rangle = \\ &= \underline{v}_{//} \end{aligned}$$

Es decir que un vector paralelo a \underline{u} no cambia con esa expresión.

Ahora apliquemos la misma expresión a un vector \underline{v}_{\perp} perpendicular a \underline{u} . Llamando \underline{w} al vector $\underline{u} \times \underline{v}_{\perp}$ que es perpendicular tanto a \underline{u} como a \underline{v}_{\perp} y notando además que $\underline{u} \times \underline{w} = -\underline{v}_{\perp}$:

$$\begin{aligned} \underline{v}_{\perp} &= \langle \cos(\alpha); \underline{u} \sin(\alpha) \rangle \langle 0; \underline{v}_{\perp} \rangle \langle \cos(\alpha); -\underline{u} \sin(\alpha) \rangle = \\ &= \langle 0; \cos(\alpha) \underline{v}_{\perp} + \sin(\alpha) \underline{w} \rangle \langle \cos(\alpha); -\underline{u} \sin(\alpha) \rangle = \\ &= \langle 0; \cos^2(\alpha) \underline{v}_{\perp} + \sin(\alpha) \cos(\alpha) \underline{w} + \sin(\alpha) \cos(\alpha) \underline{w} - \sin^2(\alpha) \underline{v}_{\perp} \rangle = \\ &= \langle 0; \{[\cos^2(\alpha) - \sin^2(\alpha)] \underline{v}_{\perp} + 2 \sin(\alpha) \cos(\alpha) \underline{w}\} \rangle = \\ &= \langle 0; \cos(2\alpha) \underline{v}_{\perp} + \sin(2\alpha) \underline{w} \rangle \end{aligned}$$

Esto es: la expresión aplicada a un vector perpendicular a \underline{u} , es un giro de ángulo 2α y eje \underline{u} .

Dado que a todo vector lo podemos representar como suma de un vector paralelo a \underline{u} y uno perpendicular, vemos que la fórmula anterior efectivamente es una transformación de vectores en vectores y en particular es un giro de ángulo 2α alrededor de \underline{u} .

Dado que el producto es asociativo:

$$\begin{aligned} q_1 q_2 q_2^{-1} q_1^{-1} &= 1 \quad \Rightarrow \quad (q_1 q_2)^{-1} = q_2^{-1} q_1^{-1} \\ q_1 (q_2 \underline{v} q_2^{-1}) q_1^{-1} &= (q_1 q_2) \underline{v} (q_1 q_2)^{-1} \end{aligned}$$

La composición de rotaciones se puede hacer con el producto de cuaterniones.

Para armar la matriz de rotación que corresponde a un cuaternión, basta recordar que sus columnas son los versores base transformados: $\{q \underline{e}_x q^{-1}, q \underline{e}_y q^{-1}, q \underline{e}_z q^{-1}\}$. Encontrar el cuaternión correspondiente a una matriz de rotación es más difícil pero se puede deducir resolviendo un sistema de ecuaciones o utilizando las fórmulas que aparecen en la bibliografía y en *web*. El problema es la falta de unicidad, puesto que dos cuaterniones unitarios opuestos producen la misma rotación (analizar).

Las cuatro componentes de un cuaternión unitario forman un vector unitario en \mathbb{R}^4 , la identidad de los opuestos implica que los cuaterniones en \mathbb{R}^4 se comportan como coordenadas homogéneas.

La interpolación de rotaciones se realiza mediante *slerping* en S_3 (la esfera unitaria de \mathbb{R}^4) con la única advertencia de que para calcular el ángulo entre los vectores se utiliza el producto escalar estándar de \mathbb{R}^4 y no el producto de cuaterniones.

Con estos métodos se puede trazar una curva cualquiera en la superficie de S_3 (una spline definida por secuencias de interpolaciones) y realizar una animación suave de los movimientos de un objeto con giros complejos (trompo, helicóptero) y de hecho así se utiliza en la práctica, fundamentalmente en juegos interactivos y animaciones.

