TRANSFORMACIONES

Introducción

Aquí llamaremos transformación a una función que hace corresponder cada punto del espacio con otro punto del mismo espacio:

$\underline{P} = T(P)$

En computación gráfica modelamos el espacio euclídeo tridimensional, común y silvestre, pero en matemáticas los conceptos de punto y espacio tienen que ver con cuestiones más complicadas sobre conjuntos (espacios) y sus elementos (puntos), topología (entornos) y métrica (distancias). No debemos meternos aquí en la generalidad del tema, pero necesitamos aprovechar algunos desarrollos matemáticos.

Cada transformación cambia la posición de todos los puntos del espacio; nosotros las aplicaremos a los objetos, para moverlos o deformarlos y a las direcciones para alterarlas.

En principio, se podría realizar cualquier transformación sobre los puntos de un objeto, una continua (botella $\rightarrow$ nudo en el cuello) o una discontinua (botella $\rightarrow$ rota), pero aquí trataremos con un conjunto limitado, de unas pocas transformaciones continuas: girar, estirar, reflejar, cizallar, mover y proyectar. Esas operaciones tienen la ventaja de que se pueden representar con una matriz y que se pueden combinar multiplicando de antemano las matrices de transformación; con eso se evita la miríada de operaciones que llevaría hacerlas una por una. Es el método que han elegido los fabricantes de software y hardware gráfico y ya viene incorporado en los procesadores gráficos o GPU. Cualquier otra transformación que no sea representable mediante matrices de números constantes, también puede realizarse, pero en un procedimiento aparte, que deberá ser programado para tal fin.

Para comprender todas estas transformaciones y sus matrices, necesitamos revisar algunos conceptos conocidos y a partir de ellos analizar algunos que seguramente no se han visto previamente.

Notación:

Debería saltearse esta especificación y recurrir a ella cuando no se comprenda la notación utilizada o el motivo por el cual se utiliza así.

- Reales: (a) minúsculas o (α) griegas, normal o inclinada (*a*).
- Intervalo [a,b) entre corchetes o paréntesis según el estándar cerrado/abierto.
- Conjunto o coordenadas o componentes: {a,b,c} o {x,y,z} entre llaves y horizontal, aunque sea un vector columna.
- Base: en lugar de **i**, **j**, y **k**, utilizaremos genéricamente $\mathbf{e}_i$ con i= 0, 1 o 2.
- Puntos: (P) mayúsculas de coordenadas { P^x, P^y, P^z} o { P^0, P^1, P^2}, genéricamente: P^i; ordenadas en columna.
- Vectores: (**v**) minúsculas negritas, de componentes reales {v^x, v^y, v^z}, genéricamente: v^i; ordenadas en columna.
- Matrices: (**a** o *A*) minúsculas negritas o mayúsculas inclinadas, de componentes reales a^i_j en la fila i y columna j.
- Transformaciones: (*A*) mayúsculas inclinadas.
- Punto o vector transformado: ($\underline{P}$ o $\underline{\mathbf{v}}$) subrayado
- Delta de Kronecker: δ_{ij} un real que vale 1 si i = j y 0 si i $\neq$ j. Matriz identidad: δ^i_j ídem, pero con filas y columnas.

La $\Sigma_1^n a^i_j v^j$ i $\in$ [1,m], es la componente i del vector que resulta de multiplicar una matriz de m×n por un vector columna de n componentes; la sumatoria se hace para los n índices j variando de 1 a n o bien de 0 a n−1 (C++). Cuando resulta obvio se omite la n o la j o incluso la sumatoria: Según la convención de Einstein se considera que hay una sumatoria por cada par de índices iguales, uno subíndice y el otro superíndice: en $\mathbb{R}^n$, $a^i_j v^j \equiv \Sigma_{j=1}^n a^i_j v^j$. Los sumandos pueden no ser reales, ej.: $\mathbf{v} = v^i\mathbf{e}_i$ indica que **v** es un vector formado por la suma de los n productos de sus componentes reales v^i y los vectores de la base $\mathbf{e}_i$. En los productos matriciales, el primer factor aporta una fila y el segundo una columna, para dar como resultado la suma de los productos de las componentes en orden: en $c^i_j = a^i_k b^k_j \equiv \Sigma_1^n a^i_k b^k_j$ se multiplican en orden la fila i de **a** y la columna j de **b** para dar la componente c^i_j de la matriz **c**.

Excepción: Las potencias podrían confundirse con superíndices, por ejemplo: $\mathbb{R}^n$ es "$\mathbb{R}$ a la n", el espacio vectorial real de n dimensiones. Para evitar las confusiones, las potencias suelen ponerse después de encerrar la base entre paréntesis: $(v^j)^2$, pero eso se puede omitir cuando resulte obvio, ej.: $\mathbb{R}^n$, $\mathbf{v}^2$ ($\mathbf{v}\cdot\mathbf{v}$ o módulo de **v** al cuadrado), $\mathbf{a}^{-1}$ (matriz inversa de **a**).

La distinción entre subíndices (índice de columna o ítem) y superíndices (índice de fila o de componente cartesiana) es seguramente algo novedoso, complicado y evitable; pero aclara la forma en que deben programarse los *loops* y las operaciones y es un salvavidas para simplificar el álgebra de matrices, que usaremos aquí y la de tensores no−euclídeos, que pueden llegar a usar en materias con mayor contenido teórico o físico.

Con más rigor habría que distinguir los vectores columna estándar de sus duales: los vectores fila que son 1-formas lineales (funciones que dan un escalar para cada vector). Las normales, los productos vectoriales, los coeficientes del plano ({a,b,c,d} t.q. P{x,y,z} pertenece al plano si ax+by+cz+d=0) y otras entidades, como los gradientes, caen en ésta última denominación, a diferencia de las coordenadas de puntos, velocidades y fuerzas que son vectores comunes. Asimismo, los tensores con dos subíndices (formas bilineales o 2-formas) o dos superíndices, no deberían confundirse con matrices, aunque suelen identificarse con matrices por la forma en que operan. Cada superíndice hace que la entidad transforme como los vectores columna (contravariante) mientras que cada subíndice hace que transforme con la transformación inversa (covariante, como la base). Todo esto se suele pasar por alto en el espacio euclídeo, pero aquí trabajaremos con transformaciones que lo deforman y en ocasiones (normales y planos) chocaremos con ello lo más suavemente posible.

Espacio Vectorial Lineal, Combinación Lineal y Transformación Lineal

En el espacio vectorial lineal $\mathbb{R}^n$ los elementos son vectores: n−uplas ordenadas de números reales y, mediante axiomas, se definen las conocidas operaciones de suma entre vectores y multiplicación por un escalar. Con esas dos operaciones se puede realizar una combinación lineal de m vectores:

$\mathbf{v} = \sum^m \alpha^i \, \mathbf{v}_i$. (Cada vector $\mathbf{v}_i$ se multiplica por un número α^i y se suman los m resultados en $\mathbf{v}$)

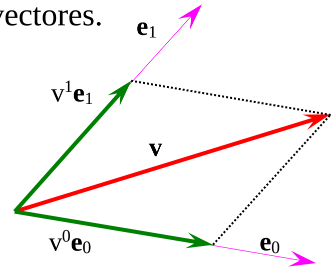
Un conjunto m de vectores es linealmente independiente (LI) si ningún elemento puede representarse como combinación lineal del resto:

$\{\mathbf{e}_i\}$, $i \in \mathbb{N}_0\,[0,m\text{-}1)$ es LI $\Leftrightarrow (\sum^m \alpha^i \, \mathbf{e}_i = \mathbf{0} \Rightarrow \alpha^i = 0)$. (No se puede despejar uno en función del resto)

En n dimensiones reales ($\mathbb{R}^n$) se pueden encontrar conjuntos LI de hasta n vectores. Cualquier otro vector se puede escribir como combinación lineal de ellos:

$\mathbf{v} = v^j \, \mathbf{e}_j$. (Se omitió la sumatoria obvia en j de 1 a n o de 0 a n−1)

Si todos los vectores pasan a representarse así, se dice que se efectúa un “cambio de base”, el conjunto LI:$\{\mathbf{e}_j\}$ es la nueva “base” y los coeficientes v^j de la combinación lineal son las nuevas “componentes en esa base”.



En cualquier base, el vector $\mathbf{e}_j$ de esa base, tiene componentes $(\mathbf{e}_j)^i = e_j^{\,i} = \delta_j^{\,i}$.

Una “transformación” es una función vectorial de variable vectorial; que hace corresponder un nuevo vector a cada vector del espacio:

$\underline{\mathbf{v}} = T(\mathbf{v})$.

También se les asignarán vectores transformados a los vectores de la base:

$\underline{\mathbf{e}}_j = T(\mathbf{e}_j)$.

Cada vector base transformado se puede representar, en la base original, mediante sus componentes:

$\boxed{\underline{\mathbf{e}}_j = \underline{e}_j^{\,i} \, \mathbf{e}_i}$. (otra vez omitida la suma obvia para i entre 0 y n-1, si no sería $\underline{\mathbf{e}}_j = \Sigma_{i=0}^{n-1} \underline{e}_j^{\,i} \, \mathbf{e}_i$)

Una transformación es lineal cuando preserva los coeficientes de toda combinación lineal:

$\underline{\mathbf{v}} = L(\mathbf{v}) = L(\sum^m \alpha^i \, \mathbf{v}_i)$ $\quad L$ es lineal $\Leftrightarrow L(\sum^m \alpha^i \, \mathbf{v}_i) = \sum^m \alpha^i \, L(\mathbf{v}_i)$. (los α^i no cambian)

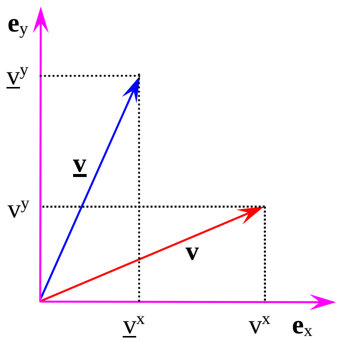
Como todo vector es combinación lineal de la base, al hacer una transformación lineal L, un vector cualquiera $\mathbf{v}$ se transforma en $\underline{\mathbf{v}}$ mediante:

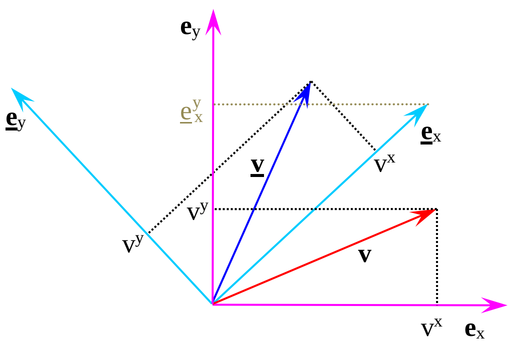
$\underline{\mathbf{v}} = L(\mathbf{v}) = L(v^j \, \mathbf{e}_j) = v^j \, L(\mathbf{e}_j) = v^j \, \underline{\mathbf{e}}_j = v^j \, \underline{e}_j^{\,i} \, \mathbf{e}_i$.

Como resultado: **En una transformación lineal basta conocer cómo se transforma la base para poder calcular como se transforman todos los demás vectores.** Basta conocer los n×n números $\{\underline{e}_j^{\,i}\}$.

Usando paréntesis, hay dos formas totalmente equivalentes de ver la misma transformación:

- El vector transformado tiene nuevas componentes en la base original:
  $\underline{\mathbf{v}} = (v^j \, \underline{e}_j^{\,i}) \, \mathbf{e}_i$: $\underline{\mathbf{v}} = \underline{v}^i \, \mathbf{e}_i$ $(\underline{v}^i = v^j \, \underline{e}_j^{\,i})$ (transformación de las componentes)
- El vector transformado tiene las mismas componentes, pero en la base transformada:
  $\underline{\mathbf{v}} = v^j \, (\underline{e}_j^{\,i} \, \mathbf{e}_i)$: $\underline{\mathbf{v}} = v^j \, \underline{\mathbf{e}}_j$ $(\underline{\mathbf{e}}_j = \underline{e}_j^{\,i} \, \mathbf{e}_i)$ (transformación de la base)



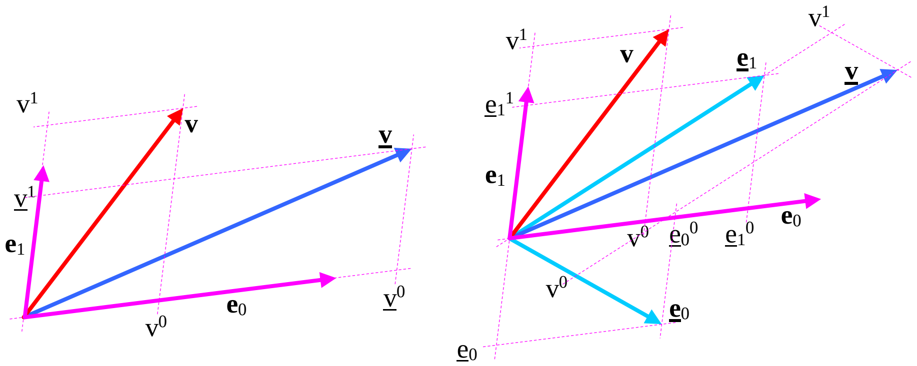


Los vectores $\underline{\mathbf{e}}_j$ de la nueva base tienen componentes $\{\underline{e}_j^{\,i}\}$ en la base original. En computación grafica se utiliza casi siempre ésta segunda interpretación: se acomoda un nuevo sistema de coordenadas en el actual y el objeto se define mediante sus coordenadas originales.[1]

[1] Es difícil evitar la confusión entre transformación y cambio de base. En una transformación los vectores cambian. En un cambio de base $\mathbf{e}_i$ a $\underline{\mathbf{e}}_i$ se conserva el vector $\mathbf{v} = v^i \, \mathbf{e}_i = \overline{v}^i \, \underline{\mathbf{e}}_i$. Las nuevas componentes $\overline{v}^i$ del mismo $\mathbf{v}$ se obtienen con la matriz inversa de $\underline{e}_j^{\,i}$. El cambio de base preserva el resultado de la combinación lineal y no los coeficientes (ver Apéndice I).

El conjunto de componentes $\{\underline{e}_j^i\}$ (de la base transformada vista desde la base original) opera como una matriz, la **matriz de la transformación lineal**. En la figura anterior, a la derecha, $\underline{\mathbf{e}}_x = \underline{e}_x^x\, \mathbf{e}_x + \underline{e}_x^y\, \mathbf{e}_y$, donde $\underline{e}_x^y$ (la única indicada) es la componente *y* del vector base transformado $\underline{\mathbf{e}}_x$, sobre el eje *y* original. Cada vector base $\underline{\mathbf{e}}_i$ es una columna, la matriz se arma (en 2D) con las dos columnas $\{\underline{\mathbf{e}}_x, \underline{\mathbf{e}}_y\}$.

En los dibujos de arriba, la transformación era un simple giro y las bases ortonormales, pero podría haber sido cualquier otra transformación lineal desde cualquier otra base, podrían haberse estirado los vectores y cambiado el ángulo entre ellos; el caso general se muestra en la siguiente figura.



Vimos que toda transformación lineal puede representarse mediante una matriz, cabe preguntarse si cualquier matriz nxn representa una transformación. La respuesta es sí: Tomamos las n columnas como n vectores y con ellos construimos la nueva base. Si las columnas no son LI, no pueden ser una base en sentido estricto, pero aun así podemos usarlas (ej.: proyecciones). Por lo tanto: **Toda matriz actúa sobre los vectores como una transformación lineal. La transformación lineal de un vector puede realizarse <u>pre</u>multiplicando el vector por una matriz de nxn <u>números</u> reales en n dimensiones**.

<u>Cuidado</u>: En $y = kx$, y es una función lineal de x, pero sólo cuando k es un número constante; si $k = 3x$ o $k = \sin(x)$, por ejemplo, ya deja de ser una función lineal. Lo mismo sucede con las matrices, estamos hablando de matrices de números y no de matrices cuyos elementos son funciones de las coordenadas. Los números sí podrían ser función de uno o más parámetros; por ejemplo: el tiempo en una animación; pero dado tiempo fijo, queda fija la transformación lineal. No pueden depender de las coordenadas.

<u>Remarcamos por su importancia</u>: **Si vemos una matriz, para entender como transforma a los vectores, tomamos sus columnas como una nueva base y así podemos deducir fácilmente su efecto sobre los objetos. Por otro lado, si necesitamos armar la matriz de una dada transformación, ponemos como columnas a los vectores base transformados**.

- Haga usted mismo un ejemplo bidimensional para entender cómo se arma la matriz de transformación con los vectores de la base transformada; luego multiplíquela por las componentes de un vector cualquiera para ver cómo se transforma.

- Verifique, en dos dimensiones, el deslizamiento (o cizalla o *shear*) de un cuadrado unitario, con una matriz que difiere de la unidad solo en el elemento $\underline{e}_2^1 = \tan(30°)$ (fila 1 columna 2) y otra con el mismo valor pero en $\underline{e}_1^2$.

<u>Propiedades:</u>

<u>Las transformaciones lineales conservan el vector nulo</u>: $\underline{\mathbf{0}} = L(\mathbf{0}) = \mathbf{0}$, (donde $\mathbf{0}$ es el vector de coordenadas nulas {0,0,0}, que pre-multiplicado por cualquier matriz da $\mathbf{0}$).

Las rectas y planos se generan por combinación lineal de vectores con parámetros variables. Dado que la transformación lineal preserva las combinaciones lineales de los vectores transformados, <u>las rectas y los planos se transforman en rectas y planos; cuyas intersecciones son las intersecciones originales transformadas. Si no hay intersecciones tampoco las habrá al transformar de modo que las paralelas se transforman en paralelas</u>.

<u>Invertibilidad e Inversa:</u>

Si la transformación tiene rango r < n, es que hay n-r columnas que son combinación lineal de las r restantes. Reemplazando en la ecuación de componentes ($\mathbf{v} = v^j\, \underline{\mathbf{e}}_j$) cada columna $\underline{\mathbf{e}}_j$ dependiente, por la combinación de las r independientes, se puede ver que todos los vectores transformados están en un subespacio r-dimensional. En ese caso la transformación "proyecta" (aplasta) y no admite inversa.

En caso contrario, una matriz nxn de rango n se puede invertir y su inversa está conformada por las componentes de la base original, vistas desde la nueva base. La inversa de la matriz es lo mismo que la transformación (función) inversa.

Composición de transformaciones y combinación:

Llamaremos composición a la aplicación de sucesivas transformaciones (A, B, C, …) de los vectores y combinación (Z) a una sola transformación que produce el mismo resultado que la composición:

$$\underline{\mathbf{v}} = C(B(A(\mathbf{v}))) = Z(\mathbf{v})$$

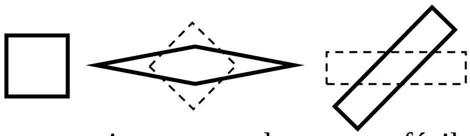
Si todas las transformaciones son lineales podemos representarlas por medio de matrices:

$$\underline{v}^h = (\textstyle\sum_k c^h_k (\sum_j b^k_j (\sum_i a^j_i v^i))) \equiv c^h_k (b^k_j (a^j_i v^i)) = c^h_k b^k_j a^j_i v^i = z^h_i v^i$$

$$z^h_i = c^h_k b^k_j a^j_i$$

En forma matricial:

$$\mathbf{z} = \mathbf{c}\ \mathbf{b}\ \mathbf{a}$$



La multiplicación de matrices no es conmutativa y se puede ver muy fácilmente que las transformaciones tampoco: la rotación de un cuadrado centrado en el origen, seguida de una escala no uniforme da un rombo, mientras que si se invierte el orden da un rectángulo girado (hacerlo analíticamente).

Para determinar el orden de las transformaciones, notar que se aplican primero "las más cercanas a **v**" y que en la multiplicación se pueden hacer asociaciones, pero manteniendo el **orden** de aplicación:

$z^h_i = c^h_k b^k_j a^j_i = c^h_k (b^k_j a^j_i) = (c^h_k b^k_j) a^j_i$ [2] $\qquad (\neq (c^h_k b^j_i a^k_j = c^h_k a^k_j b^j_i = q^h_i))$

$\mathbf{z} = \mathbf{c}\ \mathbf{b}\ \mathbf{a} = \mathbf{c}\ (\mathbf{b}\ \mathbf{a}) = (\mathbf{c}\ \mathbf{b})\ \mathbf{a} \qquad (\neq (\mathbf{c}\ \mathbf{a}\ \mathbf{b} = \mathbf{q}))$

La composición se puede mirar de dos maneras, como ya se explicó: una es la que acabamos de ver, una serie se transformaciones en orden a, b, c, sobre las coordenadas del vector original. La otra es una serie de transformaciones de la base en orden inverso: c, b, a.

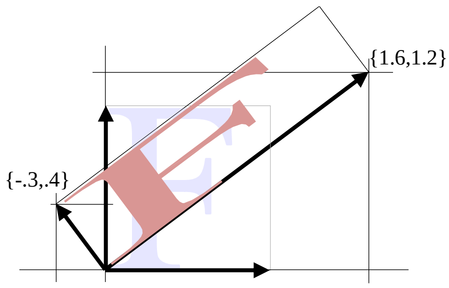
$\underline{\mathbf{v}} = \underline{v}^h \mathbf{e}_h$	$= \mathbf{e}_h (c^h_k b^k_j a^j_i v^i)$	:	$\underline{v}^h = (c^h_k (b^k_j (a^j_i v^i)))$	←
$\underline{\mathbf{v}} = v^i \underline{\mathbf{e}}_i$	$= (\mathbf{e}_h c^h_k b^k_j a^j_i) v^i$	:	$\underline{\mathbf{e}}_i = (((\mathbf{e}_h c^h_k) b^k_j) a^j_i)$	→

En la práctica, la segunda forma se corresponde con el orden de escritura de las transformaciones en un programa de Computación Gráfica y por eso la usaremos más.

Ejemplo de Transformación Lineal:

$$\begin{array}{cc} \underline{\mathbf{e}}_1 & \underline{\mathbf{e}}_2 \\ \Downarrow & \Downarrow \end{array} \quad \begin{vmatrix} 0.1 \\ 0.2 \end{vmatrix} \begin{vmatrix} 0.3 \\ 0.5 \end{vmatrix} \begin{vmatrix} 0.6 \\ 0.3 \end{vmatrix} \begin{vmatrix} 0.75 \\ 0.5 \end{vmatrix} \Leftarrow \mathbf{v}$$

$$\begin{vmatrix} 1.6 & -0.3 \\ 1.2 & 0.4 \end{vmatrix} \begin{vmatrix} 0.1 \\ 0.2 \end{vmatrix} \begin{vmatrix} 0.33 \\ 0.56 \end{vmatrix} \begin{vmatrix} 0.87 \\ 0.84 \end{vmatrix} \begin{vmatrix} 1.05 \\ 1.1 \end{vmatrix} \Leftarrow \underline{\mathbf{v}}$$



Esta transformación consiste en una rotación y escalado no-uniforme, sin cizallamiento. Normalmente, en la práctica, uno conoce el resultado al que desea llegar; pero este ejemplo muestra que es igualmente sencillo definir la matriz que logra un resultado, como deducir el resultado de aplicar una matriz dada.

La praxis básica que se enseña en los libros y en la web, consiste en definir una secuencia no conmutativa de transformaciones simples; en este caso, una escala no-uniforme y una rotación; pero el orden de las operaciones no es evidente. Del modo en que lo hemos presentado, la matriz de transformación se fija de una sola vez; usando como columnas a las componentes de la base transformada.

A la dificultad de elegir el orden de las transformaciones individuales, se contrapone la sencillez de deducirlas a partir de la matriz obtenida. Los módulos de los nuevos vectores base $|\underline{\mathbf{e}}_x| = 4$ y $|\underline{\mathbf{e}}_y| = .25$, son los factores de escala. El ángulo de rotación se desprende de la rotación del nuevo versor $\underline{\mathbf{e}}_x$ respecto del original: $\alpha = \mathrm{atan2}(1.2,1.6) \cong 37º$. Si hubiese cizallamiento, la medida del mismo se deduciría del ángulo entre los nuevos vectores base transformados, que dejaría de ser 90º.

En el ejemplo (y como suele suceder en la práctica) se define la posición deseada del objeto (la F) y se define la matriz de transformación que logra ese resultado. Pero aún si la ubicación final que se requiere, viene definida como escalas s_x, s_y y rotación α, la matriz que lo logra se puede definir de una sola vez, haciendo las obvias operaciones: $\underline{\mathbf{e}}_x = \{s_x \cos(\alpha), s_x \operatorname{sen}(\alpha)\}$; $\underline{\mathbf{e}}_y = \{s_y \cos(\alpha+90º), s_y \operatorname{sen}(\alpha+90º)\}$.

---

[2] En un programa, el tercer miembro es un *loop* en j dentro de un *loop* en k; mientras que para el cuarto es k dentro de j. Con "asociación" nos referimos al orden de los lazos o sumatorias y no al orden de los factores; que, siendo números fijos, obviamente no importa. Lo que no puede cambiase es la correspondencia de índices, el orden de los productos de matrices.

Espacio afín, combinación afín y transformación afín

Introducción:

Es muy difícil olvidar lo aprendido: lo que ya está incorporado a la forma de pensar. Pero hagamos el esfuerzo de desligar los conceptos de **punto** y **vector**. Un punto será un lugar en el espacio y un vector será una dirección o un desplazamiento.

Podemos decir que "el bar está dos cuadras al norte de la plaza". El bar y la plaza son puntos, lugares o posiciones; mientras que "dos cuadras al norte" es un vector, recorrido o desplazamiento. Podemos sumar recorridos (vectores) o multiplicarlos por números, pero no podemos sumar ni multiplicar posiciones (puntos) entre sí ni con números. No tiene sentido decir: "El doble de calle San Martín 4321" o "El bar más la plaza"; pero sí se puede decir: "El doble del recorrido desde la plaza al bar". Sí pueden hacerse algunas operaciones que dan cosas como: "R se encuentra a un tercio del camino de P a Q".

Para definir un vector en n dimensiones se necesitan n componentes y una base de n vectores. La base no está ubicada en ningún "lugar". **Los vectores no "están" en ningún punto**. Un fluido, por ejemplo, tiene una velocidad en cada punto; pero eso no debe entenderse como que hay un vector "en" un punto, sino que hay un vector "asignado" a un punto. Un campo vectorial es una función que asigna un vector (velocidad) a cada punto, así como un campo escalar asigna un número (temperatura) a cada punto.

Para ubicar un punto necesitamos un sistema de **coordenadas**. El método estándar, en el plano, lo formalizó Descartes: consiste en definir un origen o punto de referencia al cual se asignan coordenadas {0,0} (vector nulo) y dos vectores ortonormales que conforman una base de vectores. La base y el origen forman lo que en la física se conoce como marco o sistema de referencia y en matemáticas como un sistema de coordenadas cartesianas. También suele asignarse un significado físico o práctico a la base, tanto en dirección (norte, derecha) como en magnitud (metros, píxeles).

Ahora hagamos de cuenta que no sabemos nada de todo esto, sólo conocemos el espacio vectorial $\mathbb{R}^n$.

Teoría:

El espacio afín $\mathbb{A}^n$ tiene vectores y además tiene puntos. Para definirlo, se parte del espacio vectorial lineal $\mathbb{R}^n$ y se define una operación más: la diferencia entre puntos, que da un vector; o, en forma equivalente, la suma de un punto y un vector, que da otro punto:

$P - O = \mathbf{p}$

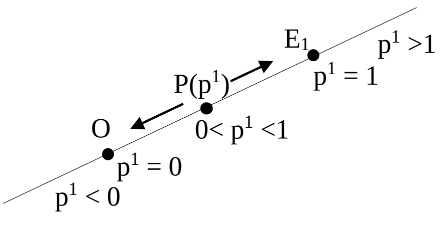
$P = O + \mathbf{p}$ (Vale restar un vector, es lo mismo que es sumar el opuesto: $O = P - \mathbf{p} \equiv O = P + (-\mathbf{p})$)

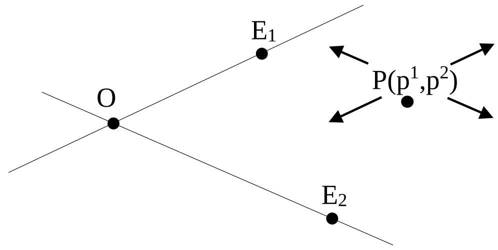
No se pueden sumar puntos, ni restar un punto a un vector, ni multiplicar puntos por escalares

El vector (bar – plaza) = (200m al N) indica el camino de la plaza al bar: bar = plaza + (200m al N). En cambio, bar + plaza no tiene sentido, el doble de la plaza tampoco. El vector no está en el bar ni en la plaza: 200m al norte del club hay un almacén, (almacén − club) es un vector y es "exactamente el mismo" que (bar − plaza), en ambos casos el vector es (200m al N). **Los vectores no tienen punto de aplicación y los puntos no tienen componentes (aunque en breve les asignaremos coordenadas).**

El espacio afín tiene una combinación lineal de vectores, idéntica a la del espacio vectorial; pero además se puede realizar una **combinación afín de puntos**. La combinación afín de m+1 puntos es un punto obtenido a través de una combinación lineal de m vectores: Sean O y $E_1 \ldots E_m$ los m+1 puntos dados; P es el punto resultante de la combinación afín; cada vector es la diferencia con O:

$(P - O) = \sum^m p^i (E_i - O)$ (CL de m vectores $(E_i - O)$ mediante m números p^i)





(Los p^i son aquí los coeficientes arbitrarios de una CL, pero en breve conformarán las coordenadas de P o su "vector posición")

La independencia afín (AI) de m+1 puntos es la independencia lineal (LI) de m vectores:

$\{E_i,O\}$ AI $\Leftrightarrow$ $\{(E_i - O)\}$ LI $\Leftrightarrow \left(\sum^m p^i (E_i - O) = \mathbf{0} \Rightarrow p^i = 0 \right)$

En un espacio afín con d dimensiones no puede haber más de d+1 puntos con independencia afín. Por ejemplo, en $\mathbb{A}^3$, tienen independencia afín: dos puntos distintos, tres puntos no alineados (no colineales) o cuatro puntos que no no están en el mismo plano (no coplanares).

En $\mathbb{A}^3$, dos puntos “definen” una recta: El “conjunto de combinaciones afines de dos puntos fijos” distintos o “expansión afín de dos puntos” (figura de arriba a la izquierda) es un espacio afín unidimensional, más conocido como recta. Del mismo modo, tres puntos “definen” un plano: La expansión afín de tres puntos fijos y no alineados es un espacio afín bidimensional o plano (figura de la derecha) y en general: **En $\mathbb{A}^d$, la expansión afín de n+1 puntos con independencia afín ($\Rightarrow n \leq d$) define un subespacio afín n-dimensional**.

El papel de los vectores $\{E_i – O\}$ en el espacio afín es exactamente el mismo que el de la base $\mathbf{e}_i$ en el espacio vectorial. Para establecer la ubicación de un punto cualquiera en $\mathbb{A}^n$, se utiliza un punto O que llamaremos **origen** y una **base** $\{\mathbf{e}_i = E_i – O\}$ de n vectores LI. El conjunto $SR = \{O, \mathbf{e}_i\} \equiv \{O, E_i\}$ es el **sistema de referencia**. Un punto queda definido por el origen y el vector posición $\mathbf{p} = (P – O)$, cuyas componentes son las **coordenadas** de P en el SR elegido:

$$P = O + \mathbf{p} = O + \sum^n p^i\mathbf{e}_i = O + \sum^n p^i (E_i – O)$$

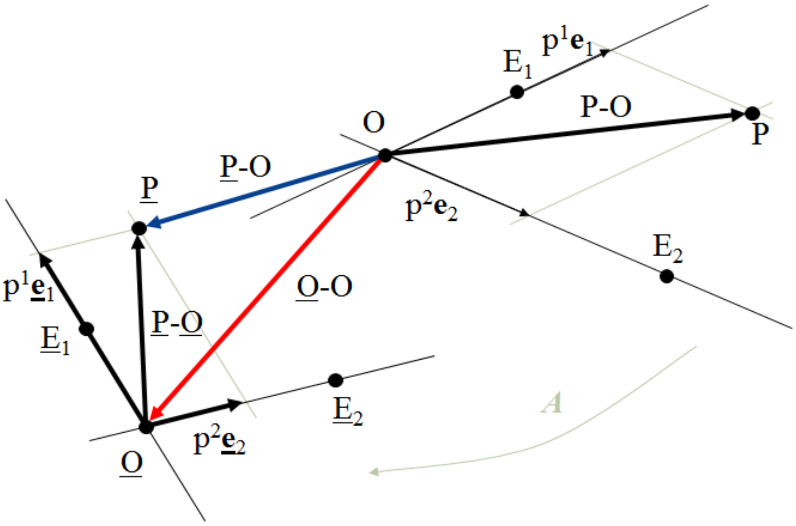
Las transformaciones afines de los puntos son aquellas que preservan los coeficientes de las combinaciones afines; es decir, la combinación lineal de (vectores diferencia con uno de los puntos):

$$\underline{P} – \underline{O} = A(P) – A(O) = A(P – O) = A[\Sigma^n p^i \mathbf{e}_i] = \Sigma^n p^i A(\mathbf{e}_i) = \Sigma^n p^i \underline{\mathbf{e}}_i \;\Rightarrow\; \underline{P} = A(P) = \Sigma^n p^i \underline{\mathbf{e}}_i + \underline{O}$$

La transformación afín de un punto es una transformación lineal del vector posición y una reubicación del origen o traslación. La transformación afín de un vector es, por definición, una transformación lineal; porque para los vectores, no importa el origen.

Si representamos los nuevos puntos $\underline{P}$ en el viejo SR (restando el viejo O):

$$(\underline{P} – O) = \Sigma^n p^i (\underline{E}_i – \underline{O}) + (\underline{O} – O) \quad \text{o} \quad \underline{p}^j = p^i \underline{e}_i^j + \underline{o}^j \qquad \text{(Notar: Lineal: } y = k\,x\text{; Afín: } y = k\,x + b\text{)}$$



La transformación afín combina una transformación lineal y una translación, pero no puede representarse sólo con la matriz $\{\underline{e}_i^j = (\underline{E}_i–\underline{O})^j\}$ sino que requiere la ubicación $\{\underline{o}^j = (\underline{O}–O)^j\}$del nuevo origen. Más adelante veremos la forma de incorporar todo en una única matriz, pero de una dimensión más.

La transformación afín conserva los coeficientes de toda combinación afín, **conserva las proporciones**: Un punto a un 10% del camino de un punto a otro, pasa a un 10% del camino entre los transformados; en particular: el centroide o el punto medio de varios puntos dados se transforma en el centroide de los transformados. Las rectas y planos son expansión afín de dos o tres puntos y por lo tanto se transforman en rectas y planos; la misma expansión afín, pero de los puntos transformados. Transformando un paralelogramo o un paralelepípedo podemos ver que las paralelas siguen siendo paralelas y las intersecciones se conservan. Sólo agregamos una traslación a las transformaciones lineales.

Síntesis de ideas útiles:
- Una combinación afín de vectores es lineal. Una combinación afín de puntos es la lineal de sus diferencias, todas con uno cualquiera de los puntos. Lo mismo sucede para la independencia afín.
- Mediante una combinación afín de dos puntos distintos podemos representar cualquier punto de la recta que definen. La combinación afín de tres puntos no alineados define los puntos de su plano.
- Dado un origen y una base, cada punto tiene asignado un vector posición, el origen tiene asignado el vector nulo.
- Las transformaciones afines consisten en una reubicación del origen y una transformación lineal de los vectores, incluyendo los vectores posición de los puntos; que, linealmente transformados, se suman al nuevo origen.
- Las transformaciones afines preservan los coeficientes de las combinaciones afines: el punto R a 1/3 del camino de P a Q, se transforma en $\underline{R}$ que está a 1/3 del camino de $\underline{P}$ a $\underline{Q}$ transformados.

Recetas:

1) La transformación lineal $\mathbf{v} \rightarrow \underline{\mathbf{v}}$ se logra premultiplicando a $\mathbf{v}$ con la matriz de transformación lineal, cuyas columnas son los vectores base transformados, pero representados en la base original de partida: $\underline{v}^j = \underline{e}_i^j\, v^i$. Para obtener la transformación afín de un punto $P \rightarrow \underline{P}$ la situación es similar, excepto por la traslación del origen: $\underline{p}^j = \underline{e}_i^j\, p^i + \underline{o}^j$; que, expandido para x, queda: $\underline{p}^x = \underline{e}_x{}^x\, p^x + \underline{e}_y{}^x\, p^x + \underline{e}_z{}^x\, p^z + \underline{o}^x\, 1$. El 1 que multiplica a $\underline{o}^x$ puede ser considerado como una cuarta componente de P y el vector $\underline{\mathbf{o}}$ puede ponerse como cuarta columna dentro de la matriz. Con una cuarta fila, se puede hacer un producto más prolijo, que puede combinarse multiplicando matrices de transformación.

$$\begin{array}{cc} & \begin{vmatrix} p^x \\ p^y \\ p^z \\ 1 \end{vmatrix} \\ \begin{vmatrix} \underline{e}_x{}^x & \underline{e}_y{}^x & \underline{e}_z{}^x & \underline{o}^x \\ \underline{e}_x{}^y & \underline{e}_y{}^y & \underline{e}_z{}^y & \underline{o}^y \\ \underline{e}_x{}^z & \underline{e}_y{}^z & \underline{e}_z{}^z & \underline{o}^z \\ 0 & 0 & 0 & 1 \end{vmatrix} = & \begin{vmatrix} \underline{p}^x \\ \underline{p}^y \\ \underline{p}^z \\ 1 \end{vmatrix} \end{array}$$

El espacio proyectivo, que veremos más adelante, dará sustento geométrico a esta receta.

CUIDADO: Para inferir la matriz una transformación afín, a partir del resultado deseado, el punto $\underline{O}$ y sus coordenadas $\underline{o}^j$ son muy fáciles de identificar en el SR original; pero para las primeras columnas suelen confundirse las nuevas bases con las coordenadas de los nuevos $\underline{E}_i$. No, no son los vectores posición $\underline{E}_i$-O, son los vectores $\underline{\mathbf{e}}_i = \underline{E}_i - \underline{O}$ de componentes $\underline{e}_i{}^j = (\underline{E}_i - \underline{O})^j = \underline{E}_i{}^j - \underline{O}^j$.

2) Ya se dijo que no se pueden sumar puntos ni multiplicarlos por escalares. Lo que sí puede hacerse es una combinación afín; por ejemplo entre P y Q se puede definir un punto R = P + (Q-P)/4, a ¼ del camino entre P y Q, sumando un punto P y una fracción del vector (P-Q). Analicemos como depende ese punto R de la base y el origen del sistema de referencia 1D:

$$O + r^1 (E_1 - O) = O + p^1 (E_1 - O) + \tfrac{1}{4} (q^1 - p^1) (E_1 - O)$$

Haciendo operaciones válidas (restar puntos a puntos, sumar vectores, etc.):

$$r^1 (E_1 - O) = (\tfrac{3}{4} p^1 + \tfrac{1}{4} q^1) (E_1 - O) \;\Rightarrow\; [r^1 - (\tfrac{3}{4} p^1 + \tfrac{1}{4} q^1)] (E_1 - O) = \mathbf{0}$$

$$r^1 = \tfrac{3}{4} p^1 + \tfrac{1}{4} q^1$$

Esta última igualdad es porque el vector **0** tiene componentes nulas. Haciendo cualquier otro ejemplo 1D, se puede ver que siempre se puede operar con las coordenadas y el resultado es independiente del SR, pero es un promedio ponderado: una suma pesada cuyos pesos suman uno.

Veamos una combinación afín 2D: S = P + ½ (Q-P) + ⅜ (R-P). Operando del mismo modo que antes:

$$O + s^i (E_i - O) = O + p^i (E_i - O) + \tfrac{1}{2} (q^i - p^i) (E_i - O) + \tfrac{3}{8} (r^i - p^i) (E_i - O)$$

$$s^i (E_i - O) = (\tfrac{1}{8} p^i + \tfrac{1}{2} q^i + \tfrac{3}{8} r^i) (E_i - O) \;\Rightarrow [s^i - (\tfrac{1}{8} p^i + \tfrac{1}{2} q^i + \tfrac{3}{8} r^i)] (E_i - O) = \mathbf{0}$$

$$s^i = \tfrac{1}{8} p^i + \tfrac{1}{2} q^i + \tfrac{3}{8} r^i$$

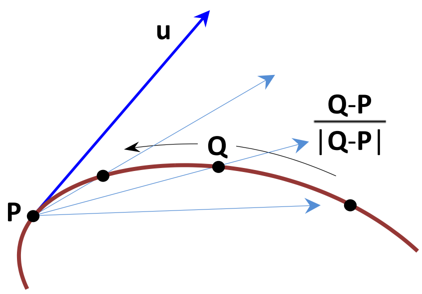
Otra vez, haciendo un promedio ponderado de las coordenadas se obtiene el mismo resultado y es independiente del sistema de referencia.

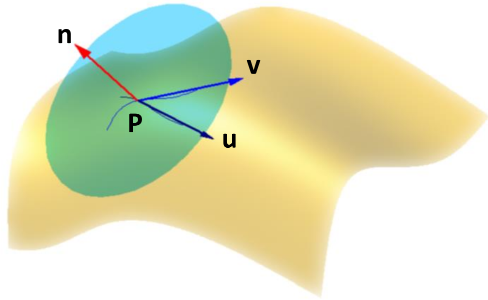
Esta receta es, entonces: Se permite hacer un promedio ponderado de puntos y representarlo como una suma de puntos, multiplicados por pesos, pero siempre que los pesos sumen uno. El resultado debe interpretarse como un promedio ponderado de las coordenadas de los puntos en cualquier SR. La notación permitida sería: R = ¾ P + ¼ Q en una dimensión, o S = ⅛ P + ½ Q + ⅜ R en dos. Cualquier transformación afín hará que: $\underline{R} = \tfrac{3}{4} \underline{P} + \tfrac{1}{4} \underline{Q}$ o $\underline{S} = \tfrac{1}{8} \underline{P} + \tfrac{1}{2} \underline{Q} + \tfrac{3}{8} \underline{R}$.

Normales y transformación de las normales:

La normal a una superficie en un punto es un vector perpendicular al plano tangente en ese punto. Parece trivial, pero la superficie es curva. Podemos definir vectores tangentes en un punto P de una superficie por medio del límite de una diferencia de puntos, normalizada (dividida por el módulo o distancia):

$$\mathbf{u}(\mathrm{P}) = \lim_{\mathrm{Q}\to\mathrm{P}} \frac{\mathrm{Q}-\mathrm{P}}{|\mathrm{Q}-\mathrm{P}|}$$

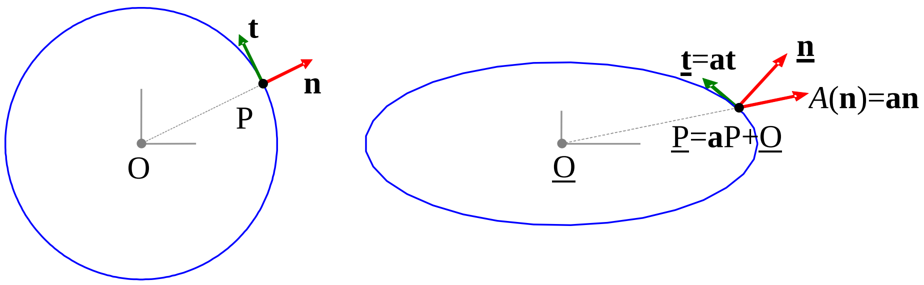




Para definir un vector tangente **u** a una superficie en un punto P, utilizamos otro punto Q que se acerca a P siguiendo una curva sobre la superficie. Por cada dirección de llegada habrá una tangente distinta, límite de la secante Q−P. Con determinadas condiciones de suavidad (G^1) el límite existe y todas las tangentes serán coplanares, estarán en el mismo plano tangente. En tal caso, con dos tangentes distintas, **u** y **v**, alcanza para definir el plano tangente: P+α**u**+β**v** que pasa por el punto P y cuya normal **n** = **u**×**v**.

Analicemos como se transforma la normal en una transformación afín *A*. Siendo **n** un vector no nos importa la traslación de los puntos y nos concentraremos en la matriz lineal **a**: $a_i^{\,j} = \underline{e}_i^{\,j} = (\underline{E}_i-\underline{O})^j$. Mediante la transformación *A*, la superficie se transforma en otra, posiblemente deformada y trasladada, pero el vector *A*(**n**) = **an** (que no representaremos mediante $\underline{\mathbf{n}}$) no es necesariamente perpendicular a la superficie transformada. El plano tangente, en cambio, sí se transforma en un plano tangente; porque la transformación afín conserva la combinación afín que utilizamos para definir las tangentes como límite.

Si la transformación aplasta, no habrá tangente ni normal; por lo tanto, hablemos de transformaciones que no aplastan y se pueden revertir: dada la transformación *A* existe A^{-1} y su matriz $\mathbf{a}^{-1}$.



Queremos un vector que sea normal en $\underline{\mathrm{P}}$ a la superficie transformada por *A*. Buscamos un nuevo vector perpendicular al plano tangente, que llamaremos $\underline{\mathbf{n}}$, pero es distinto del vector transformado *A*(**n**) = **an**. Para conseguirlo en forma simple, partimos de una base ortonormal, donde se define el producto punto o escalar $\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\,\mathbf{v} = \Sigma\, u^i\, v^i$, que resulta nulo si los vectores son perpendiculares.

Si **t** es cualquier vector del plano tangente, es el anti-transformado de $\underline{\mathbf{t}} = A(\mathbf{t})$. A la izquierda se harán las expresiones indiciales y a la derecha las equivalentes matriciales:

$$0 = \Sigma_i\, n^i\, t^i = \Sigma_i\, n^i\, [\Sigma_j\, (a^{-1})^i{}_j\, \underline{t}^j] \qquad\qquad 0 = \mathbf{n}^T\,\mathbf{t} = \mathbf{n}^T\,(\mathbf{a}^{-1}\,\underline{\mathbf{t}})$$

El segundo miembro es el producto escalar en el espacio euclídeo. En el tercero, se reemplaza a **t** por el vector $\mathbf{a}^{-1}\underline{\mathbf{t}}$, anti-transformado de $\underline{\mathbf{t}}$ (el origen no afecta a los vectores).

Reordenando las sumas y los números, pero sin alterar la correspondencia de índices:

$$0 = \Sigma_j\, [\Sigma_i\, (a^{-1})^i{}_j\, n^i]\, \underline{t}^j = \Sigma_j\, [\Sigma_i\, ((a^{-1})^T)^j{}_i\, n^i]\, \underline{t}^j = \Sigma_j\, \underline{n}^j\, \underline{t}^j \qquad\qquad 0 = (\mathbf{n}^T\,\mathbf{a}^{-1})\,\underline{\mathbf{t}} = ((\mathbf{a}^{-1})^T\,\mathbf{n})^T\,\underline{\mathbf{t}} = \underline{\mathbf{n}}^T\,\underline{\mathbf{t}}$$

Por lo tanto, el vector:

$$\underline{n}^j = ((\mathbf{a}^{-1})^T)^j{}_i\, n^i \qquad\qquad \underline{\mathbf{n}} = (\mathbf{a}^{-1})^T\,\mathbf{n}$$

es normal a la superficie transformada en el punto transformado.

La normal a la superficie transformada se obtiene transformando la normal original mediante la transpuesta de la inversa de la matriz de transformación.

Se puede ver que "hay algo raro": la normal no transforma como cualquier vector; podemos tranquilizarnos al verlo como que la perpendicularidad no se preserva y la tangencia sí; pero, en realidad, la normal es un vector fila y transforma como tal.

Espacio proyectivo y transformaciones proyectivas

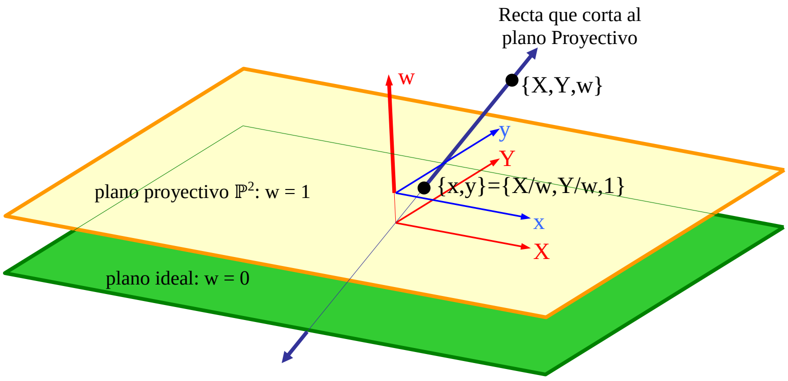
En $\mathbb{R}^3$, cualquier vector, excepto el nulo, define la dirección de una recta por el origen[3]. **Las rectas por el origen en $\mathbb{R}^3$, son los puntos (elementos) de un nuevo espacio: el espacio proyectivo $\mathbb{P}^2$.**

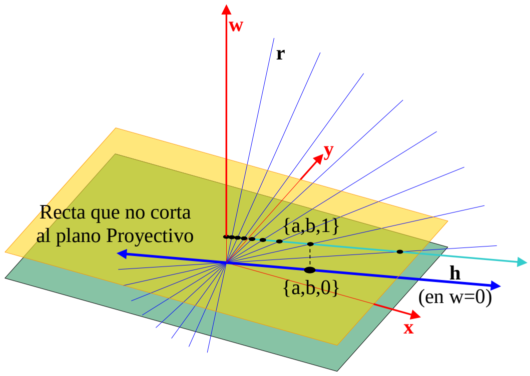
Usaremos {X,Y,w} como coordenadas en $\mathbb{R}^3$, luego quedará claro por qué w en lugar de Z.

En $\mathbb{R}^3$, {X,Y,w} y {αX,αY,αw}, con $\alpha \in \mathbb{R} \neq 0$, representan la misma recta por el origen. Para identificarla (asignarle coordenadas) mediante un único vector, la cortamos con el plano *w*=1. Cada punto de ese plano representa una recta diferente, la que pasa por ese punto y el origen.

Pero a ese plano le faltan puntos para representar las rectas horizontales: las rectas que están en el plano *w*=0 no cortan al plano *w*=1 en ningún punto. Esas rectas quedan determinadas mediante algún vector con *w*=0. Las rectas del plano *w*=0 se suelen llamar "puntos ideales" o "puntos en el infinito" de $\mathbb{P}^2$, digamos que es porque la recta "corta al plano *w*=1 en el infinito".

La figura siguiente quizás aclare un poco más. En la figura izquierda se representan los tres ejes coordenados de $\mathbb{R}^3$, el **plano ideal con *w*=0 y el plano proyectivo, identificado con *w*=1**. Además hay una recta por el origen en dirección a {X,Y,w} y que corta al plano proyectivo en el punto de coordenadas {X/w,Y/w,1} según $\mathbb{R}^3$ o {x,y} en un sistema de referencia del plano proyectivo $\mathbb{P}^2$, cuyo origen está en {0,0,1} de $\mathbb{R}^3$.





La otra recta, más complicada, está en el plano *w*=0, en dirección {a,b,0}. Para entender cuál punto le asigna coordenadas en el plano proyectivo, analizamos, en la figura de la derecha, el plano vertical que forma la recta horizontal **h** con el eje vertical **w**. Cualquier recta **r** ≠ **h**, de ese plano, corta al plano *w*=1 en alguno de los puntos marcados. Se puede ver que a medida que la recta **r** se va horizontalizando, es decir: alejándose del eje **w** y acercándose a **h**, el punto del plano proyectivo se va alejando de su origen. A la recta **h** le corresponde el punto límite (ideal o en el infinito) de la secuencia de puntos en el plano. Cada punto {a,b,1} esta justo por encima del {a,b,0}; la sucesión de puntos marcados en el plano proyectivo puede definirse mediante {γa,γb,1} ≡ {a,b,1/γ}. Cuando γ crece, la recta que representa se va horizontalizando hasta que "en el infinito" coincide con la recta **h**. Cualquier vector {αa,αb,0}, con α≠0, sirve para representar la misma recta horizontal; lo importante es la dirección. Incluso podría hacerse α negativo y decrecer hasta -∞, de donde se ve la equivalencia de los dos puntos ideales opuestos[4].

El vector {X,Y,w} ≠ **0** de $\mathbb{R}^3$, define las coordenadas homogéneas de un punto de $\mathbb{P}^2$. Si w ≠ 0 son las del punto finito {x,y}={X/w,Y/w}; si *w* = 0, son las de un punto en el infinito, en dirección a {X,Y}. Podemos decir que con *w*≠0 tenemos puntos (finitos) y con *w*=0 direcciones hacia puntos en el infinito.

Resumiendo:

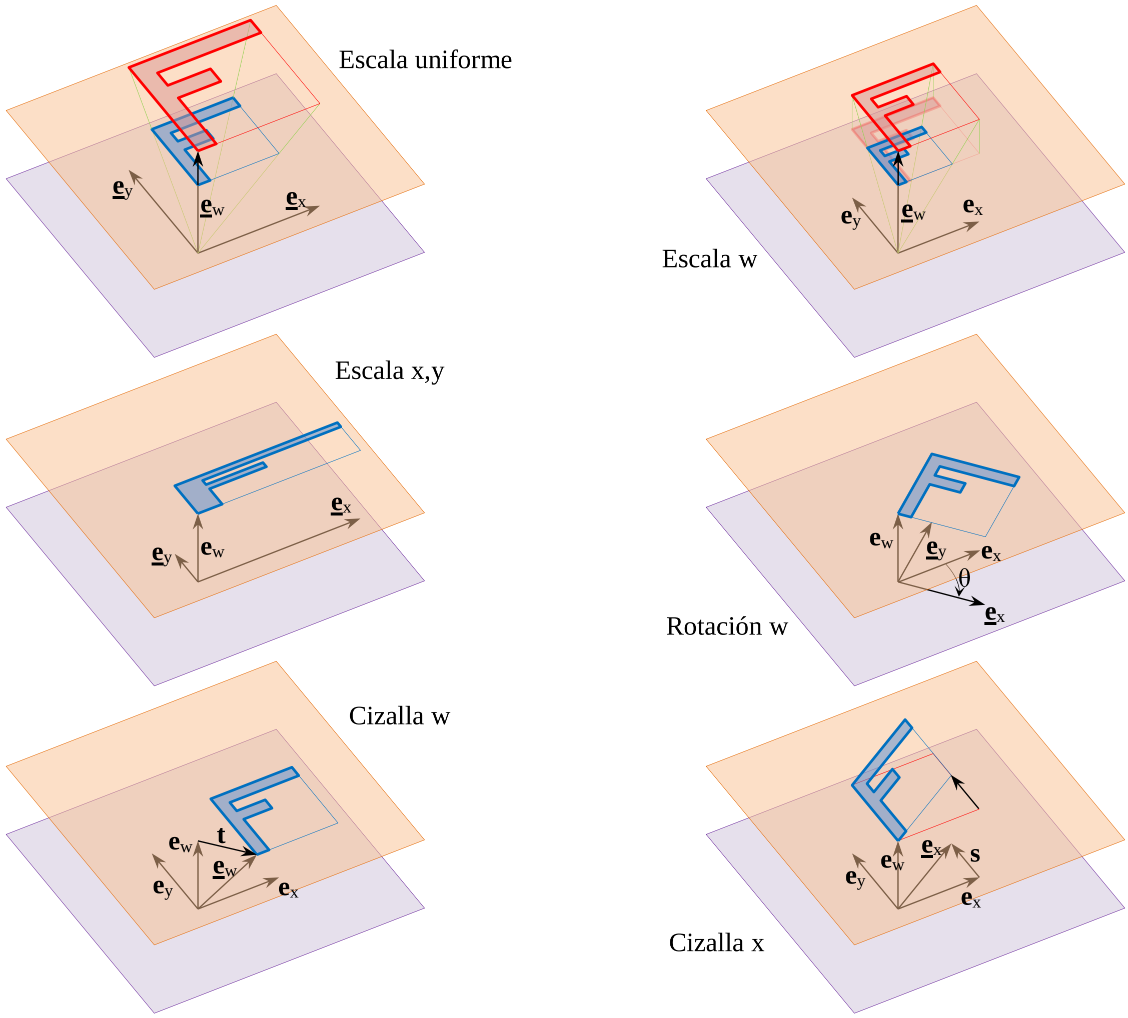
1) Las rectas por el origen de $\mathbb{R}^3$ son, a su vez, los "puntos" de un nuevo espacio denominado "plano proyectivo" $\mathbb{P}^2$,
2) que identificamos con el plano *w*=1 de $\mathbb{R}^3$, donde ubicamos "nuestros puntos", los vértices de coordenadas finitas.
3) Las rectas en el plano ideal (*w*=0) también son puntos de $\mathbb{P}^2$; representan direcciones hacia puntos en el infinito.
4) El vector {X,Y,w} de $\mathbb{R}^3$, define las coordenadas homogéneas de cualquier punto de $\mathbb{P}^2$. Si w≠0 son las del punto estándar 2D {x,y}={X/w,Y/w}, si *w*=0, son las de un punto en el infinito, en dirección a {X,Y}.
   a) Si *w*≠0, el vector define una recta que corta al plano *w*=1 en el punto de coordenadas {x,y} en el plano proyectivo.
   b) Si *w*=0, representa una recta horizontal y se corresponde en $\mathbb{P}^2$ con un punto ideal, en el infinito, definido por el vector {X,Y} que indica la dirección desde el origen del plano proyectivo.

[3] Por simplicidad decimos "recta por el origen", pero son subespacios lineales 1D, en $\mathbb{R}^3$ no hay origen (se puede usar $\mathbb{A}^3$).

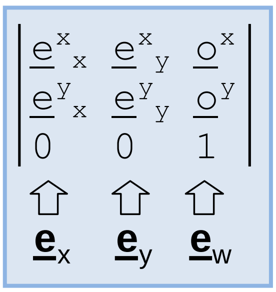
[4] Aunque, en OpenGL y en general en CG, se diferencian los puntos ideales opuestos (dirección de la luz).

El efecto de una transformación lineal en $\mathbb{R}^3$ sobre los puntos de $\mathbb{P}^2$ se denomina transformación proyectiva. Algunos casos particulares se reducen a las transformaciones conocidas:

Lineal en $\mathbb{R}^3$	**Proyectiva en $\mathbb{P}^2$**
Escala uniforme X, Y, *w*:	Identidad (los puntos se mueven en las mismas rectas por el origen)
Escala en *w*:	Escala uniforme (cuanto más se aleja en *w* más chico “se ve” en P^2)
Escalas en X e Y:	Escalas en *x* e *y*
Cizalla al eje *w*:	Traslación (se mueve el punto (0,0,1), que es el origen de “nuestro” espacio)
Cizalla a los ejes X e Y:	Cizalla
Rotación con eje *w*:	Rotación
Otras:	Difícil de ver… sólo usaremos la proyección perspectiva

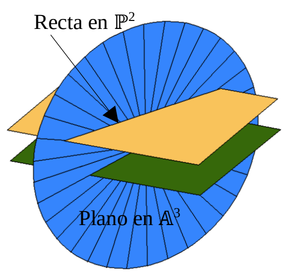


Todas las transformaciones afines pueden representarse como transformaciones lineales (por lo tanto, mediante matrices) en un espacio lineal con una dimensión más. Cualquier transformación afín 2D se puede definir por medio de una matriz en coordenadas homogéneas 3D, como se muestra a la derecha. EN la transformación afín, **siempre se mantienen horizontales los ejes *x* e *y*** (sus bases): $\underline{e}_x^{\,w} = \underline{e}_y^{\,w} = 0$; las traslaciones inclinan el eje *w*, pero solo deslizan los planos *w* = cte. sobre sí mismos, **el deslizamiento del plano *w* = 1 lleva el origen a su nueva posición trasladada:** $\underline{e}_w = \underline{O} \equiv \{\underline{o}^x, \underline{o}^y, 1\}$ es la medida de la traslación.



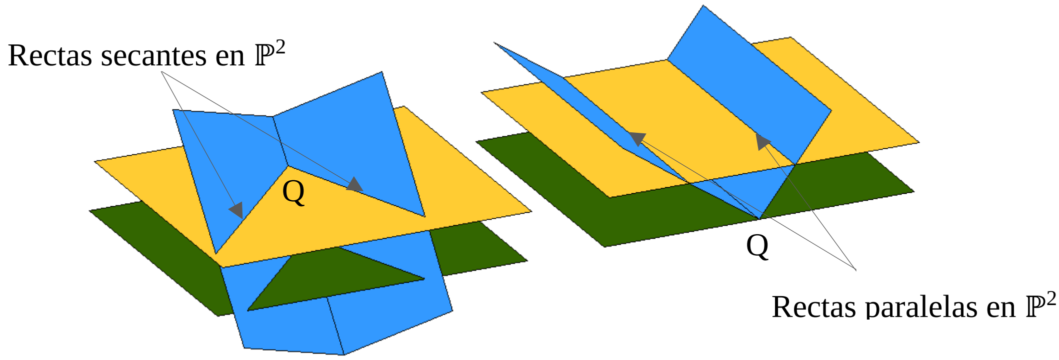
$$\begin{array}{ccc} \underline{e}^x{}_x & \underline{e}^x{}_y & \underline{o}^x \\ \underline{e}^y{}_x & \underline{e}^y{}_y & \underline{o}^y \\ 0 & 0 & 1 \\ \Uparrow & \Uparrow & \Uparrow \\ \underline{\mathbf{e}}_x & \underline{\mathbf{e}}_y & \underline{\mathbf{e}}_w \end{array}$$

En una transformación proyectiva general, las rectas y planos de $\mathbb{R}^3$ siguen siendo rectas y planos y además se conserva el vector nulo, esto es por ser lineal en $\mathbb{R}^3$. Un plano por el origen de $\mathbb{R}^3$ define una recta en $\mathbb{P}^2$. Por lo tanto, la transformación proyectiva transforma rectas en rectas. El plano *w*=0 es la “recta en el infinito” para $\mathbb{P}^2$, contiene la sucesión de puntos ideales o en el infinito, esta se transforma en una recta, pero puede que no se quede en el infinito.

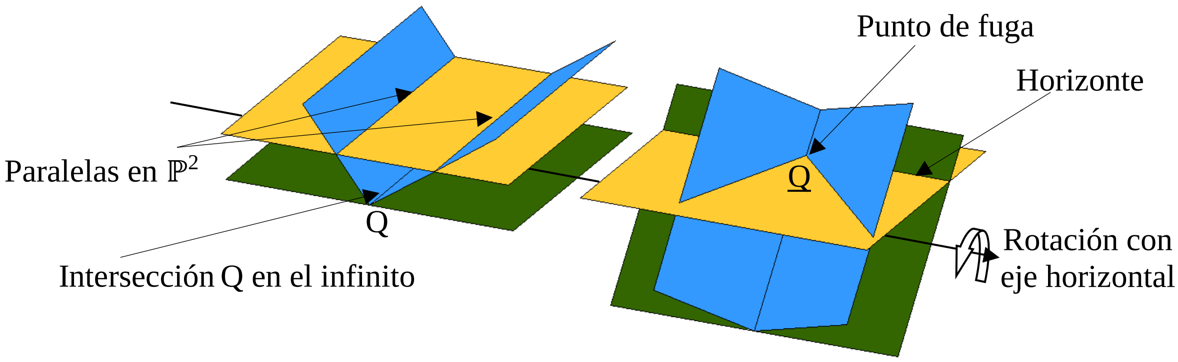


En las afinidades (transformaciones afines) no se inclinan ni se cruzan los planos horizontales (w cte.), en particular el plano ideal y el proyectivo. Pero, en general, en una transformación lineal cualquiera de $\mathbb{R}^3$, el plano proyectivo se puede inclinar y eso hace que puntos que estaban en el infinito pasen a ser finitos y viceversa (ej.: fotografía del horizonte).

Dos rectas que se cortan en $\mathbb{P}^2$ en el punto Q (figura de abajo a la izquierda) corresponden a dos planos que se cortan en la recta que pasa por el origen de $\mathbb{R}^3$ y por el punto Q. Dos rectas paralelas de $\mathbb{P}^2$ (a la derecha) corresponden a dos planos por el origen de $\mathbb{R}^3$, que se interceptan en una recta del plano w=0, paralela a ambas rectas de $\mathbb{P}^2$. **En ambos casos se puede decir que la intersección existe; en el caso sencillo el punto Q tiene coordenadas finitas y en el otro es un punto ideal, en el infinito.**



En las transformaciones afines, el plano ideal, rota sobre sí mismo y/o se estira, pero se queda en su sitio; las rectas paralelas siguen siendo paralelas (el infinito se mapea en el infinito) y las que se interceptan siguen interceptándose en el punto transformado. En una transformación proyectiva en general, si se modifica la posición del plano ideal, las cosas se complican y mucho. El plano ideal original ahora corta al nuevo plano proyectivo. A consecuencia de ello, algunos puntos ideales, que estaban en el infinito, ahora pasan a tener coordenadas finitas y forman una recta en $\mathbb{P}^2$, la recta en el infinito se transforma en la intersección del plano girado y el proyectivo. No casualmente, llamaremos **horizonte** a esa recta.



En la figura de arriba a la izquierda, dos (o más) planos por el origen se cortan en una recta con w=0, en $\mathbb{P}^2$ son rectas paralelas que se cortan en un punto Q del infinito, en la dirección dada por algún punto de la recta en w=0. Luego de una transformación lineal en $\mathbb{R}^3$ (proyectiva en $\mathbb{P}^2$) las rectas se cortan en el punto $\underline{\text{Q}}$ transformado. Los puntos del infinito han pasado a tener coordenadas finitas en $\mathbb{P}^2$ sobre la línea del horizonte. (En una fotografía, "aparece", en el horizonte, el final de un camino largo)

Del mismo modo, rectas en $\mathbb{P}^2$ que antes se interceptaban puede que ahora resulten paralelas.

Para poder realizar los dibujos y las interpretaciones hemos identificado "nuestro espacio" con el plano $\mathbb{P}^2$ inmerso en $\mathbb{R}^3$. Si nos pasamos al espacio de modelos tridimensional, tendremos que identificarlo con el hiperplano $\mathbb{P}^3$ al que le corresponde w=1 en $\mathbb{R}^4$: {X,Y,Z,w}. Las cosas se complican a la hora de visualizar cuatro dimensiones, solo analizaremos un caso sencillo: la proyección central.

La interpretación de una transformación proyectiva general se basa en sus propiedades centrales: se conserva la incidencia, es decir las intersecciones (aun las ideales) y se conserva la razón doble (*cross ratio*) que es la proporción de proporciones.

Todo este lio es principalmente para poder realizar todas las transformaciones y proyecciones que usamos en computación gráfica como productos de matrices y vectores de cuatro dimensiones; pues así se hace, muy rápido y en paralelo en la GPU. Pero también lo utilizaremos para comprender las curvas y superficies paramétricas racionales que se utilizan en computación gráfica y en CAD en general.

Proyecciones:

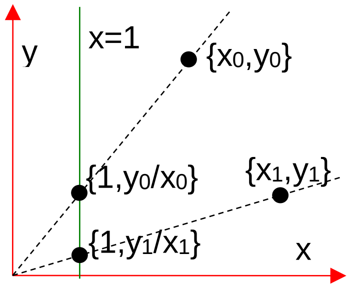
Las proyecciones transforman todos los puntos del espacio tridimensional en puntos de un mismo plano, el plano de la imagen. En el proceso se pierde una dimensión, por lo tanto, son transformaciones de rango 2, no invertibles y no se puede recuperar la posición 3D de un punto de la imagen 2D.

Perspectiva:

La proyección perspectiva central consiste en interceptar, en el plano de la imagen, todos los rayos que van desde cada punto de la escena al ojo del observador. Queda definida por medio de la posición del "ojo", punto de vista o centro y un vector desde el ojo, en dirección a la visual y perpendicular al plano de la imagen, su longitud se denomina distancia focal y define la ubicación del plano.

El punto de vista es un punto finito del espacio, supongamos el ojo en el origen y el plano de dibujo en x=1. Una transformación afín permitirá modificar la posición del ojo y el vector visual, para llevar el caso particular planteado a cualquier otro; <u>incluso para la proyección de sombras sobre un plano</u>.

Dado que proyectamos en x=1, las coordenadas finales de un punto cualquiera {x,y,z} pasan a ser: {1,y/x,z/x}, al igual que las de cualquier otro punto de coordenadas {αx,αy,αz} que esté en la misma recta de proyección.



$$\begin{bmatrix} 1/x & 0 & 0 \\ 0 & 1/x & 0 \\ 0 & 0 & 1/x \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ y/x \\ z/x \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} wx \\ wy \\ wz \\ w \end{bmatrix} = \begin{bmatrix} wx \\ wy \\ wz \\ wx \end{bmatrix}$$

Esa transformación, cuya matriz de elementos variables (funciones de x) está a la izquierda, no es lineal en tres dimensiones. En coordenadas homogéneas, la misma transformación, el mismo resultado, pero como se muestra a la derecha, convierte {x,y,z,1} en {x,y,z,x} ≡ {1,y/x,z/x} y ahora sí resulta lineal.

Proyección ortogonal:

La proyección ortogonal aplasta el espacio como un acordeón. Es como si lleváramos al infinito el ojo de la proyección perspectiva. Los rayos que unen cada punto original con su imagen, son paralelos entre sí y perpendiculares al plano de dibujo. Proyectaremos sobre el plano x=0, pero esa proyección particular es fácilmente generalizable mediante transformaciones afines.

La proyección ortogonal sobre x=0 consiste simplemente en asignar a cualquier punto de coordenadas {x,y,z} el punto {0,y,z}. Es una transformación lineal aún en $\mathbb{R}^3$. En coordenadas homogéneas de $\mathbb{R}^4$ será: {x,y,z,1} → {0,y,z,1}.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} wx \\ wy \\ wz \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ wy \\ wz \\ w \end{bmatrix}$$

La matriz de transformación homogénea es la que se muestra a la derecha. Las columnas de la matriz, como hemos visto antes, representan los ejes transformados por $L(\mathbb{R}^4)$, se puede ver que el nuevo versor $\underline{\mathbf{e}}_x$ se aplastó hasta anularse, ahora tiene coordenadas nulas $\underline{\mathbf{e}}_x = \mathbf{0}$ = {0,0,0,0}.

En ambos tipos de transformación hay puntos del infinito que aparecen en el plano de la imagen. En proyección ortogonal, solo se ve un punto ideal {x,0,0,0} en dirección de la proyección, nada interesante. En perspectiva, en cambio, todos los planos que contienen al ojo se proyectan como rectas por el origen, por lo tanto, se ven sus puntos en el infinito. Con un poco más de esfuerzo intelectual se podrían entender el horizonte y los puntos de fuga y hasta plantear las ecuaciones que describen el dibujo en perspectiva.

La matriz de proyección que se utiliza en CG requiere preservar la distancia de cada punto al ojo, para definir la oclusión visual. Entonces "no proyecta", sino que acomoda las cosas de modo que se pueda omitir la profundidad para generar la imagen. OpenGL utiliza la matriz de proyección central que se muestra en el Apéndice IV. Esta matriz hace que los objetos disminuyan su tamaño en proporción con la distancia al ojo; pero, al no aplastar, aún resulta invertible. Finalmente, OpenGL simplemente ignora la coordenada z para armar la imagen, es decir que la única proyección efectiva (aplastamiento) es ortogonal. Es como si la proyección perspectiva se ejecutara en dos pasos matriciales, el primero deforma como una perspectiva y el segundo proyecta como una ortogonal.

Nestor Calvo
Computación Gráfica
FICH – UNL

Apéndice I: Cambio de Base

Dijimos que un vector en $\mathbb{R}^n$ es una n-upla de números reales, pero eso no es del todo exacto en las aplicaciones. Las magnitudes escalares requieren números, pero además unidades; por ejemplo: 0.5in (*inches* o pulgadas) es lo mismo que 12.7mm; las magnitudes vectoriales necesitan además una base o, dado el caso, un sistema de referencia. De tal modo, una “misma magnitud vectorial” o “mismo vector”, para simplificar, puede ser definido en distintas bases, mediante distintas n-uplas de números reales.

Un cambio de base mantiene el mismo vector, pero lo representa según otra base. Vayamos directamente al espacio afín $\mathbb{A}^3$, con un cambio de sistema de referencia ($SR \rightarrow \underline{SR}$), es decir, un cambio de base y de origen. Un punto P se representa mediante las componentes del vector P–O, la nueva base será el conjunto de vectores $\underline{\mathbf{e}}_i$ y el nuevo origen será $\underline{O}$, todo con componentes dadas en el SR de partida.

$$P = \Sigma_k\, p^k \mathbf{e}_k + O = \Sigma_j\, \overline{p}^j \underline{\mathbf{e}}_j + \overline{O} \qquad \Rightarrow \qquad \Sigma_k\, p^k \mathbf{e}_k = \Sigma_j\, \overline{p}^j \underline{\mathbf{e}}_j + (\underline{O} - O)$$

(Notar que “el mismo” P tiene “nuevas componentes” $\overline{p}^k$; mientras que, en una transformación, $\underline{P}$ cambiaba y se conservaban las componentes p^k.)

Esas ecuaciones vectoriales definen las nuevas coordenadas $\{\overline{p}^j\}$ de P en función de los datos: las coordenadas originales, la nueva base y el desplazamiento del origen. Es un sistema de n ecuaciones de componentes en la base original. Escribamos la componente (o ecuación) i-ésima.

$$(P-O)^i = (P-\underline{O})^i + (\underline{O}-O))^i \qquad \Rightarrow \qquad p^i = \Sigma_j\, \overline{p}^j \underline{e}_j^{\,i} + \underline{o}^i$$

Ya vimos cómo poner eso en forma matricial homogénea, con una matriz **m** de 4x4 formada por los vectores $\underline{e}_j^{\,i}$ y el nuevo origen en la cuarta columna:

$$\begin{vmatrix} p^x \\ p^y \\ p^z \\ 1 \end{vmatrix} = \begin{vmatrix} \underline{e}_x{}^x & \underline{e}_y{}^x & \underline{e}_z{}^x & \underline{o}^x \\ \underline{e}_x{}^y & \underline{e}_y{}^y & \underline{e}_z{}^y & \underline{o}^y \\ \underline{e}_x{}^z & \underline{e}_y{}^z & \underline{e}_z{}^z & \underline{o}^z \\ 0 & 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} \overline{p}^x \\ \overline{p}^y \\ \overline{p}^z \\ 1 \end{vmatrix} \qquad \Rightarrow \qquad \begin{vmatrix} \overline{p}^x \\ \overline{p}^y \\ \overline{p}^z \\ 1 \end{vmatrix} = \begin{vmatrix} \underline{e}_x{}^x & \underline{e}_y{}^x & \underline{e}_z{}^x & \underline{o}^x \\ \underline{e}_x{}^y & \underline{e}_y{}^y & \underline{e}_z{}^y & \underline{o}^y \\ \underline{e}_x{}^z & \underline{e}_y{}^z & \underline{e}_z{}^z & \underline{o}^z \\ 0 & 0 & 0 & 1 \end{vmatrix}^{-1} \begin{vmatrix} p^x \\ p^y \\ p^z \\ 1 \end{vmatrix}$$

Puede verse que la matriz del cambio de base $\{O,\mathbf{e}\}$ a $\{\underline{O},\underline{\mathbf{e}}\}$ es inversa a la de transformación.

Una transformación del sistema A al B es lo mismo que un cambio de base de B a A. Pero no debe confundirse con la segunda interpretación de las transformaciones, pues no es lo mismo: En la transformación de A a B, los puntos cambian de lugar, pero las coordenadas en la nueva base B son las mismas que tenían en A; por el contrario, en el cambio de base de B hacia A, los puntos no cambian, pero sí sus coordenadas. Los “números” se calculan como si una fuese inversa de la otra.

Ahora bien, si evitamos la confusión y dado que ya aprendimos a hacer una transformación, podemos plantear el cambio de base de SR a $\underline{SR}$, como una transformación en sentido contrario, es decir con los versores y origen de SR expresados en coordenadas de $\underline{SR}$. En esencia es un modo gráfico o geométrico de calcular la inversa de una matriz de 3x3 o 4x4.

En algunos libros de texto, sobre todo en mecánica, se representan las bases en fila, pero esto sólo sirve para una transformación rígida, donde la submatriz de 3x3 de la base es ortogonal y por lo tanto la inversa es igual a la transpuesta.

Si miramos bien cómo funciona el pipeline gráfico, veremos que hay transformaciones (modelo) y cambios de base (sistema modelo → sistema visual); pero <u>**en Computación Gráfica, para evitar las confusiones, todas las operaciones son analizadas y ejecutadas como transformaciones**</u>.

Apéndice II: Métrica y orientación

Las transformaciones deforman los objetos; para analizar cómo cambian las medidas, debemos analizar primero que es una medida en un espacio afín $\Rightarrow$ euclídeo.

<u>Métrica:</u>

Para medir longitudes de segmentos rectos se requiere uno al que llamar “unidad”, la medida es la cantidad de veces que cabe la unidad en el segmento. La unidad debe poder fraccionarse de forma que la medida sea un número real. Para medir las áreas planas, la unidad es un cuadrado de lado unitario: el interior de una poligonal de cuatro segmentos unitarios con dos diagonales iguales. El cuadrado también define los ángulos rectos y permite medir ángulos. Con un cubo unitario se podrán medir volúmenes.

La recta, el plano y el espacio que habitamos, son espacios euclídeos de una dos y tres dimensiones. Euclides describió sus propiedades geométricas y Descartes, 1900 años después, los dotó de un sistema de referencia analítico. En el plano cartesiano, se define un origen y dos retas numéricas perpendiculares, constituyendo un sistema de referencia ortonormal. Los vectores de la base canónica son unitarios y ortogonales: cada uno tiene componente 1 en su base y 0 en la otra. Pero desde un punto de vista abstracto es esa asignación la que define ortonormalidad.

Algebraicamente, es el producto escalar entre vectores el que define longitudes y ángulos. <u>En la base canónica</u>, se define el producto escalar, interno o punto, entre pares de vectores: $\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\,\mathbf{v} = \Sigma\, u^i\, v^i$. El resultado es un número real. Es conmutativo: $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$ y bilineal: $(\alpha\mathbf{u}+\beta\mathbf{v})\cdot\gamma\mathbf{w} = \alpha\gamma\ \mathbf{u}\cdot\mathbf{w} + \beta\gamma\ \mathbf{v}\cdot\mathbf{w}$.

La norma, módulo o longitud de un vector se define como $|\mathbf{v}| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{\mathbf{v}^2}$, siempre positivo. El único vector de módulo 0 es el vector $\mathbf{0}$, todas sus componentes son nulas. Para dos puntos, se define la distancia como el módulo del vector diferencia.

El producto escalar también define el ángulo entre dos vectores; por definición, el coseno de dicho ángulo es el producto escalar dividido por el producto de los módulos.

Para las mediciones lineales y angulares, aun entre puntos, solamente importa el producto escalar entre vectores. En general: $\mathbf{u}\cdot\mathbf{v} = u^i\,\mathbf{e}_i \cdot v^j\,\mathbf{e}_j = u^i\,v^j\,\mathbf{e}_i\cdot\mathbf{e}_j$; pero sólo para una base ortonormal: $\mathbf{e}_i\cdot\mathbf{e}_j := \delta_{ij}$ (módulos unitarios, coseno nulo); estos productos escalares de la base definen la **métrica** euclídea y ésta define al espacio afín como euclídeo. Por ejemplo, poniendo un vector como combinación de la base, al calcular el módulo se puede ver el teorema de Pitágoras.

Aún para la distancia entre puntos, la ubicación del origen no importa; pues, tanto para medir longitudes como ángulos, se utilizan vectores diferencia entre pares de puntos. Por lo tanto, una traslación en el espacio afín no cambia la métrica ni las medidas. Pero ante una transformación lineal cualquiera, puede que la nueva base ya no sea ortonormal y, en general, cambian los módulos y los ángulos de los vectores: $\underline{\mathbf{u}}\cdot\underline{\mathbf{v}} = \underline{\mathbf{u}}^T\,\underline{\mathbf{v}} = (L\,\mathbf{u})^T\,L\,\mathbf{v} = \mathbf{u}^T L^T\,L\,\mathbf{v}$.

La matriz **l** de la transformación lineal L, se forma con las componentes de las nuevas bases como columnas: $l_i^{\,j} = \underline{e}_i^{\,j}$; su transpuesta L^T las tendrá en las filas; por lo tanto, si $\mathbf{g} = L^T\,L \Rightarrow g_{ij} = (\underline{\mathbf{e}}_i)^T\,\underline{\mathbf{e}}_j = \underline{\mathbf{e}}_i\cdot\underline{\mathbf{e}}_j$. El conjunto de productos escalares de los vectores de la nueva base define la métrica (longitudes y ángulos) del espacio transformado: $\underline{\mathbf{u}}\cdot\underline{\mathbf{v}} = (u^i\underline{\mathbf{e}}_i)\cdot(v^j\underline{\mathbf{e}}_j) = \underline{\mathbf{e}}_i\cdot\underline{\mathbf{e}}_j\ u^i\,v^j = g_{ij}\,u^i\,v^j$.

Se puede observar que el producto escalar quedará igual que el original solamente cuando $\mathbf{g} = L^T L = I$, o $g_{ij} = \delta_{ij}$; es decir si L es ortogonal: $L^T = L^{-1}$, lo que implica que sus columnas siguen formando una base ortonormal y la métrica sigue siendo euclídea. Estas son las **transformaciones rígidas**, pues preservan las longitudes y los ángulos. Las transformaciones rígidas son o combinan rotaciones y traslaciones, lo que indica que, para medir, no importan ni el origen ni la orientación angular de las bases ortonormales. Pero sí hay que tener un cuidado: para medir ángulos orientados, la segunda base está girada 90º “a la derecha” de la primera.

<u>Orientación</u>:

En la recta numérica, la flecha unitaria puede apuntar para un lado o para el otro, es indiferente. Vista como espacio afín, resultan irrelevantes la ubicación del origen y la dirección de la única base, pero las coordenadas de los puntos en 1D tienen signo, según estén del lado al que apunta la base o el contrario.

En dos dimensiones, la segunda base define otra recta numérica, perpendicular a la primera y rotada hacia la derecha. Pero el álgebra no tiene mano derecha, entonces utiliza el signo de un determinante.

En tres dimensiones aparece otra arbitrariedad: ¿hacia arriba o hacia abajo? ¿de qué? Las respuestas son las mismas: humana, mano derecha y algebraica, signo de un determinante.

Además del producto escalar, para dos vectores en el espacio euclídeo bidimensional se puede definir el producto cruz y, en el espacio tridimensional, se definen el producto vectorial, externo o también, producto cruz y, combinándolo con el escalar, el triple producto o producto mixto:

$$a = \mathbf{u}\times\mathbf{v} = \begin{vmatrix} u^x & v^x \\ u^y & v^y \end{vmatrix} \qquad \mathbf{n} = \mathbf{u}\times\mathbf{v} = \begin{vmatrix} u^x & v^x & \mathbf{e}_x \\ u^y & v^y & \mathbf{e}_y \\ u^z & v^z & \mathbf{e}_z \end{vmatrix} \qquad v = \mathbf{u}\times\mathbf{v}\cdot\mathbf{w} = \begin{vmatrix} u^x & v^x & w^x \\ u^y & v^y & w^y \\ u^z & v^z & w^z \end{vmatrix}$$

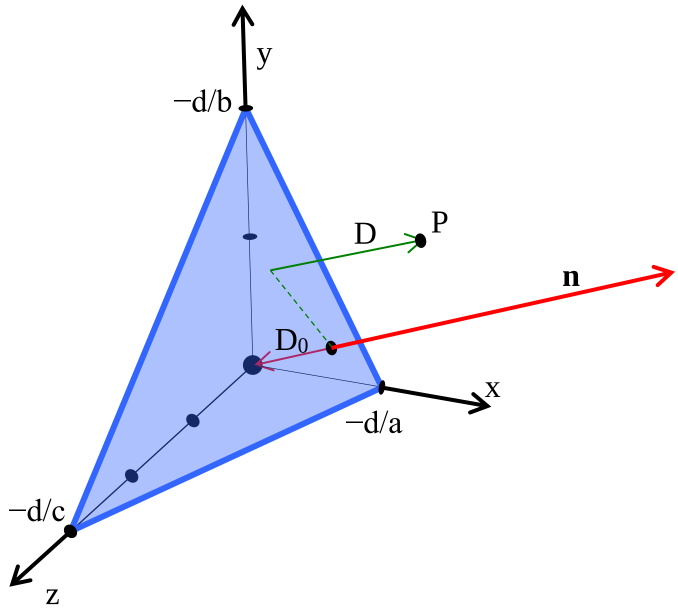
El producto cruz 2D o 3D es anti-conmutativo: intercambiando el orden de **u** y **v**, el determinante cambia de signo. En 2D mide el área del paralelogramo que se usa para sumar los dos vectores, pero con signo: el área orientada $a = |\mathbf{u}||\mathbf{v}|\sin(\mathbf{u},\mathbf{v})$, es el seno del ángulo orientado el que impone el signo. En 3D, el vector **n**, es perpendicular a **u**, a **v** y a sus combinaciones lineales ($\mathbf{n}\cdot(\alpha\mathbf{u}+\beta\mathbf{v}) = 0$); su módulo $|\mathbf{n}|$ mide el área del paralelogramo y su dirección está definida por la regla de la mano derecha: parado como **n** y mirando en dirección a **u**, el vector **v** queda a la izquierda. El producto cruz define el área orientada, los giros y circuitos ordenados igual o al contrario del giro de las agujas del reloj (CW / CCW).

El producto mixto $v = \mathbf{u}\times\mathbf{v}\cdot\mathbf{w}$, define el volumen del paralelepípedo que se usa para sumar los vectores y su signo determina la orientación (derecha/izquierda) de las ternas ordenadas de vectores. El signo cambia siguiendo el orden cíclico de los factores: $\mathbf{u}\times\mathbf{v}\cdot\mathbf{w} = \mathbf{v}\times\mathbf{w}\cdot\mathbf{u} = \mathbf{w}\times\mathbf{u}\cdot\mathbf{v} = -\,\mathbf{v}\times\mathbf{u}\cdot\mathbf{w} = -\,\mathbf{w}\times\mathbf{v}\cdot\mathbf{u} = -\,\mathbf{u}\times\mathbf{w}\cdot\mathbf{v}$.

Ante una transformación lineal, las áreas y volúmenes se multiplican por el determinante de la matriz de transformación; es muy sencillo verlo, analizando como cambia el cuadradito o el cubito que usamos para medir área o volumen.

El producto vectorial no cambia de forma sencilla, pero en CG nos interesa de sobremanera porque es el que define normales a planos y, con ello, determina la iluminación. De modo que el mecanismo de cambio es parte de teoría principal, fuera de este apéndice.

Apéndice III: Planos y normales



Los puntos (x,y,z) de un plano son los que cumplen una función de 1er grado en las coordenadas:

$$ax + by + cz + d = 0$$

Los parámetros a, b, c y d definen al plano. Dado que está igualada a cero, la ecuación es homogénea: un factor escalar α no la afecta: Π = {a,b,c,d} = {αa,αb,αc,αd}.

Haciendo cero cada par de coordenadas, se ve que las intersecciones con los ejes son las indicadas en la figura. Por ejemplo: cuando y = z = 0, se puede despejar x = −d/a.

Un vector normal al plano se obtiene mediante el producto vectorial de dos vectores no paralelos del plano; por ejemplo, el producto de dos aristas del triángulo de la figura de arriba:

$$\left(\begin{Bmatrix}0\\-d/b\\0\end{Bmatrix}-\begin{Bmatrix}-d/a\\0\\0\end{Bmatrix}\right)\times\left(\begin{Bmatrix}0\\0\\-d/c\end{Bmatrix}-\begin{Bmatrix}-d/a\\0\\0\end{Bmatrix}\right)=\begin{Bmatrix}d/a\\-d/b\\0\end{Bmatrix}\times\begin{Bmatrix}d/a\\0\\-d/c\end{Bmatrix}=\begin{Bmatrix}d^2/bc\\d^2/ca\\d^2/ab\end{Bmatrix}=\frac{d^2}{abc}\begin{Bmatrix}a\\b\\c\end{Bmatrix}.$$

El factor escalar d^2/abc modifica la longitud del vector, pero no interesa; el vector {a,b,c} ya es normal al plano. Para obtener un versor normal, normalizamos el vector {a,b,c} dividiéndolo por su módulo:

$$\mathbf{n}=\frac{1}{\sqrt{a^2+b^2+c^2}}\begin{Bmatrix}a\\b\\c\end{Bmatrix}.$$

Para calcular la distancia del plano a un punto P{x,y,z}, formamos un vector entre un punto cualquiera del plano y el punto P; para ese vector, se calcula la componente perpendicular al plano:

$$D=(P-P_{plano})\cdot\mathbf{n}=\frac{1}{\sqrt{a^2+b^2+c^2}}\left(\begin{Bmatrix}x\\y\\z\end{Bmatrix}-\begin{Bmatrix}-d/a\\0\\0\end{Bmatrix}\right)\cdot\begin{Bmatrix}a\\b\\c\end{Bmatrix}=\frac{ax+by+cz+d}{\sqrt{a^2+b^2+c^2}}.$$

Esa distancia tiene signo que indica si el punto está del lado de la normal o en el opuesto.

Si el punto es el origen O{0,0,0}, puede verse que d indica la distancia del plano al origen: D(O)=d/||n||; y entonces, un plano por el origen tiene d=0. El plano del infinito tiene 1/d = 0: se puede ver que, cuando $d \to \infty \Rightarrow \Pi = \{a/d,b/d,c/d,1\}$, por la homogeneidad.

En la fórmula de la distancia, el numerador puede verse como un producto escalar de dos vectores de cuatro dimensiones: el punto P = {x,y,z,1} y el plano Π = {a,b,c,d}. En CG se utiliza esa ecuación del plano porque se opera con vectores 4D; se considera al punto P{wx,wy,wz,w} con sus cuatro componentes: y al plano como un vector $\Pi\{n_x,n_y,n_z,d\}$. Multiplicando escalarmente y luego dividiendo por w, la ecuación queda: D = Π·P/w|**n**|.

Para algunas consultas no hace falta dividir; por ejemplo, si solo se requiere saber si el punto está en el plano o de qué lado está (clipping) o si se quiere ordenar puntos de acuerdo a la distancia, o saber cuál es más lejano. OpenGL, por ejemplo, **nunca** hace esa división; si se quiere que el producto escalar sea la distancia real, hay que normalizar la ecuación del plano haciendo $a^2 + b^2 + c^2 = 1$ y usar w=1.

Transformación proyectiva de planos y normales:

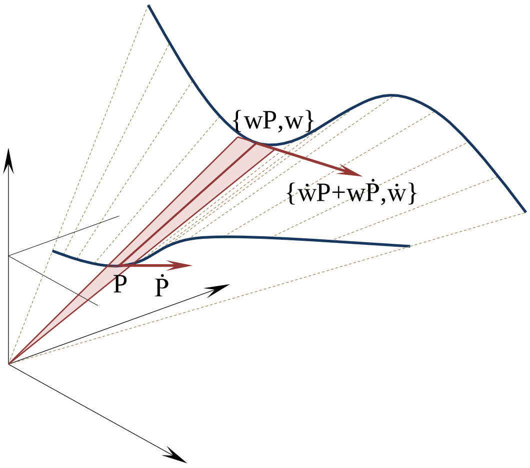
Anteriormente vimos como afectaba la transformación afín a un vector normal, aquí veremos que esa misma consideración vale para los planos en las transformaciones proyectivas en general. Dijimos que el plano se puede considerar como un vector de cuatro dimensiones $\Pi = \{a,b,c,d\}$; esto no es del todo exacto puesto que no se trasforma del mismo modo que un vector columna estándar 4D, sino como un vector fila. Al igual que los vectores normales, los que definen los planos tienen una esencia diferente, se transforman de distinto modo que un punto o vector común 3D.

La propiedad del plano que debe conservarse invariante es que un punto P del plano Π debe cumplir con la ecuación $\Pi{\cdot}P = 0 \Rightarrow \underline{\Pi}{\cdot}\underline{P} = 0$. Aquí utilizamos la notación matricial, el producto escalar entre dos vectores columna se obtiene como el producto matricial de un vector fila por un vector columna, en ese orden. Una transformación proyectiva *M* con una matriz **m** de 4x4 invertible, transforma puntos y vectores (columna) haciendo $\underline{P} = \mathbf{m}P$ y tiene inversa, de modo que $P = \mathbf{m}^{-1}\underline{P}$. Hagamos la deducción paso a paso para que sea trivial:

$$0 = \Pi^T P = \Pi^T (\mathbf{m}^{-1}\,\underline{P}) = (\Pi^T\,\mathbf{m}^{-1})\,\underline{P} = [(\mathbf{m}^{-1})^T\,\Pi]^T\,\underline{P} = \underline{\Pi}^T\,\underline{P}$$

Es decir que el vector columna del plano transformado $\underline{\Pi} = (\mathbf{m}^{-1})^T\,\Pi$ se obtiene con la inversa transpuesta de la matriz de transformación. Dado que las tres primeras componentes de Π definen la normal del plano, ésta también se transforma del mismo modo[5].

Para definir la normal a una superficie general, usábamos tangentes a curvas, justificando que las tangentes se transformaban en tangentes por la invariancia afín; pero las transformaciones proyectivas no mantienen la invariancia afín; de modo que no podemos usar ese argumento. Nos preguntamos si al proyectar una curva 4D junto con su tangente en un punto, sobre el plano *w*=1, obtendremos “nuestra” curva y su tangente. La respuesta es: en general sí; a menos que la recta tangente pase por el origen (cuando el vector tangente apunta desde el punto al origen). Pero, en general, el origen 4D y el vector tangente en el punto, definen un triángulo y este un plano, ese plano corta al proyectivo en una línea que es tangente a la curva proyectada. La figura derecha muestra una curva y su tangente proyectadas desde $\mathbb{A}^3$ en $\mathbb{P}^2$.



Consideremos, sobre la superficie una curva P(*t*) en función de un parámetro *t*:

$$\frac{d}{dt}\{wP,w\} = \{\dot{w}P+w\dot{P},\dot{w}\}$$

La derivada proyectada $(P+\dot{P}w/\dot{w})$ no es la derivada 3D $(\dot{P})$, y ni siquiera es un vector, es un punto, a menos que $\dot{w} = 0$; pero veremos que el vector derivada 4D pertenece al plano tangente a la curva 3D.

El plano por el origen y tangente a la curva 3D en *w*=1, se forma mediante las combinaciones lineales del vector posición y el vector derivada. Un punto cualquiera de ese plano es $Q=\alpha\{P,1\}+\beta\{\dot{P},0\}$, tomando $\alpha=w$ y $\beta=0$, vemos que P pertenece al plano (obvio); con $\alpha=\dot{w}$ y $\beta=w$, vemos que el punto que identifica la derivada 4D: $\dot{w}\{P,1\}+w\{\dot{P},0\}$, también está en ese plano.

Como conclusión, la normal transformada por una transformación proyectiva se obtiene mediante la transformación del plano por el origen y tangente a la superficie (sus curvas 4D) en el punto. Si el punto (vector posición) es {wP,w}, el plano tangente es {**n**,d}, de modo que su producto escalar 4D sea nulo, es decir: $d = -P{\cdot}\mathbf{n}$, haciendo este producto punto en 3D.

---

5 En algunos textos e incluso en la especificación de OpenGL se utiliza en forma errónea la inversa transpuesta de la primera parte 3x3 de la matriz de transformación. Eso sólo es correcto cuando los versores base transformados no contienen componente w, es decir: para una transformación puramente afín (que es el caso general) y no para una proyectiva. Para el vector normal solo importan las tres primeras componentes y las traslaciones no lo afectan, es decir que sí se puede usar solo la primera parte 3x3, pero de la inversa transpuesta de la matriz 4x4 completa. El caso es distinto para un plano, pues la cuarta componente d determina la posición del plano.

Apéndice IV: Matriz de proyección perspectiva

Trataremos la deducción de la matriz de perspectiva que utiliza OpenGL para explicarnos porqué se mapea *z* en 1/*z*, lo cual tiene implicaciones en los cálculos de oclusión.

Se parte desde el espacio visual, con la convención de que $\mathbf{e}_z$ "pincha el ojo", $\mathbf{e}_x$ apunta a la derecha y $\mathbf{e}_y$ hacia arriba. La matriz de proyección que se utiliza no proyecta, sino que transforma el frustum en el cubo canónico $[-1,1]^3$. Ese cubo luego será estirado al tamaño del *viewport* y la profundidad *z* será ignorada. La primera transformación es la que interesa aquí, debe llevar los planos *left*, *right*, *bottom*, *top*, *near* y *far* a los límites del cubo. Los planos *near* y *far* vienen dados por sus distancias positivas al ojo: N y F; por lo que los valores de *z* son –N y –F respectivamente. Los laterales están inclinados y entonces, por convención, se definen por sus coordenadas visuales en el plano *near*: L, R, B y T.

Para "proyectar" en el plano *z*=h, con el ojo en el origen, la matriz debería ser como la que se muestra a la derecha. Esa matriz transforma (x,y,z,1) en (hx,hy,hz,z), que reducido a 3D es (hx/z,hy/z,h) como se pretendía. Todos los vértices con el mismo par (x/z,y/z) tienen su proyección central en $\underline{z}$=h y se perdió la información necesaria para oclusión.

$$\begin{vmatrix} h & 0 & 0 & 0 \\ 0 & h & 0 & 0 \\ 0 & 0 & h & 0 \\ 0 & 0 & 1 & 0 \end{vmatrix}$$

Al proyectar se pierde la profundidad (*depth* o *z*) que se requiere para descartar los fragmentos ocultos por otros más cercanos al ojo. Calculados *x* e *y* de la proyección, el valor $\underline{z}$=h ya es inútil, pues la imagen está generada. Podemos "poner" en ese lugar una función monótona de la profundidad que resulte útil para oclusión. Eso se puede lograr con la matriz que de la derecha (∀a,b≠0), que transforma (x,y,z,1) → (hx,hy,az+b,z) → (hx/z,hy/z,a+b/z).

$$\begin{vmatrix} h & 0 & 0 & 0 \\ 0 & h & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & 1 & 0 \end{vmatrix}$$

Si se quiere el plano de proyección en h=–N (por convención) y además, asignar $\underline{z}$=−1 a los puntos del plano *near* y $\underline{z}$=1 a los del *far plane*; se logra haciendo a=(F+N)/(F−N), b=2FN/(F−N).

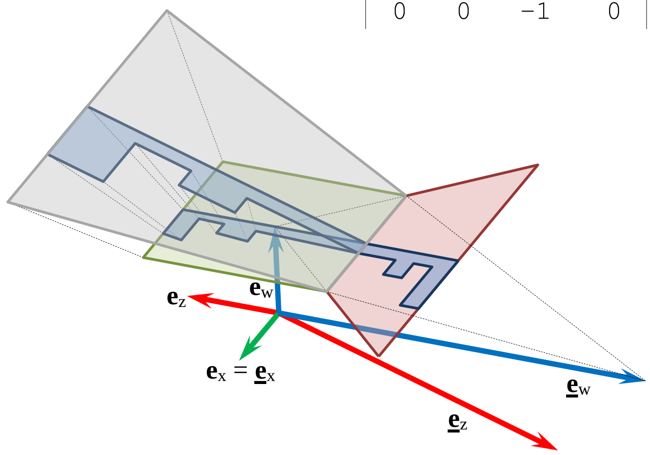
Teniendo *x* e *y* proyectados, sólo resta ubicar los planos laterales entre −1 y 1 mediante traslación y escalado como se muestra a la derecha (notar que la traslación (d,f) va en la tercera columna, la que multiplica al *z* del punto que será el divisor, en lugar del *w*).

$$\begin{vmatrix} ch & 0 & d & 0 \\ 0 & eh & f & 0 \\ 0 & 0 & a & b \\ 0 & 0 & 1 & 0 \end{vmatrix}$$

El nuevo *x* debe ser –1 en todo el plano *left* y 1 en todo el plano *right*, y lo mismo para la coordenada *y* en *bottom* y en *top;* lo que se logra del mismo modo que antes para *z*, resolviendo, en cada caso, dos ecuaciones con dos incógnitas. La matriz que emplea OpenGL es la que se muestra a la derecha. La matriz se ha multiplicado por −1, lo que no cambia el resultado, debido a la división por $\underline{w}$.

$$\begin{vmatrix} \frac{2N}{R-L} & 0 & \frac{R+L}{R-L} & 0 \\ 0 & \frac{2N}{T-B} & \frac{T+B}{T-B} & 0 \\ 0 & 0 & -\frac{F+N}{F-N} & -\frac{2FN}{F-N} \\ 0 & 0 & -1 & 0 \end{vmatrix}$$

En la figura se muestra, en 2D+1, el efecto de esta transformación proyectiva que, como ya se dijo, no proyecta, sino que transforma el frustum en el cubo canónico. Se descarta el eje *y* por la dificultad de visualizar cuatro dimensiones, pero recibe un tratamiento similar al del eje *x*.



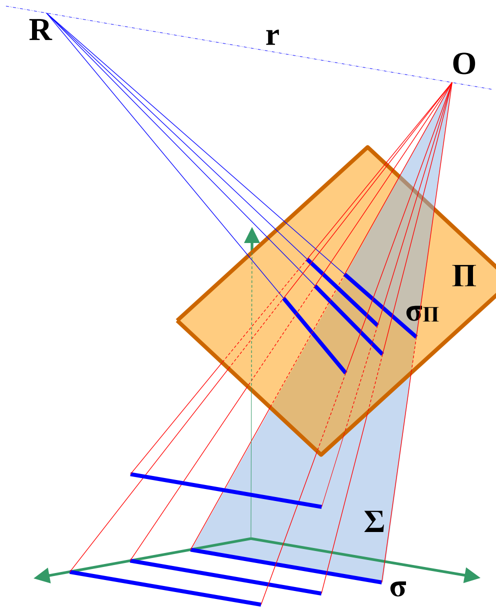
El "frustum" original es el trapecio rojo de la derecha, situado en el plano *w* = 1, comprendido entre *z* = −1 y *z* = −2, con R = 1 y L = −1 en *x*, medidos en el "plano" *near*. Usando la matriz con los valores adecuados, los nuevos versores base, representados en los originales son $\underline{\mathbf{e}}_x$ (1,0,0), que coincide con el original, $\underline{\mathbf{e}}_z$ (0,−3,−1) y $\underline{\mathbf{e}}_w$ (0,−4,0). El frustum transformado es el trapecio gris, inclinado respecto al plano proyectivo y al hacer la (mal llamada) división perspectiva se transforma en el cuadrado canónico verde con *w* = 1 y $(x,z) \in [-1,1]^2$.

Con una observación detenida y algo de análisis, se puede ver que el ancho (Δz) de las líneas horizontales de la F es constante en el plano inclinado pero variable en el cuadrado canónico, siendo más ancha la línea media que la superior. Es por esto que el *z−buffer* tiene más precisión cerca del plano *near*. El segmento recto diagonal del frustum se convierte en un segmento recto diagonal del cuadrado; pero, mediante marcas, se puede ver que se recorre más rápido cerca del plano near del cuadrado.

Otra observación es que el plano ideal transformado ($\underline{\mathbf{e}}_x$-$\underline{\mathbf{e}}_z$ o $\underline{w}$ = 0) queda inclinado y corta al proyectivo (*w* = 1) en una línea que contiene a todos los puntos de fuga (ex puntos ideales, ahora visibles).

Se puede analizar que el momento más adecuado para hacer el clipping de los objetos es sobre el trapecio gris, recortando cualquier coordenada *x*, *y* o *z* que caiga fuera de [−*w*,*w*].

Apéndice V: Puntos y Líneas de Fuga



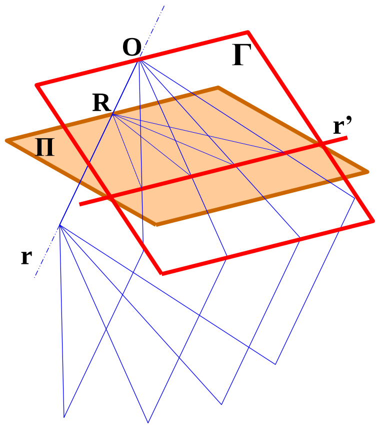
La imagen de un punto cualquiera es la intersección, en el plano **Π** de la imagen, del rayo que une el ojo **O** y el punto.

Un segmento **σ** y el ojo forman un triángulo y por lo tanto un plano **Σ**. En ese plano hay una recta **r** que pasa por el ojo y es paralela al segmento, su imagen es el punto **R**. La recta que sostiene la imagen **σ_Π** del segmento corta a la recta **r** en un punto que pertenece a la vez al plano **Σ** y al de la imagen **Π**, es decir en el punto **R**.

Otros segmentos o rectas, paralelas a **σ**, forman con el ojo un conjunto de planos y todos tienen en común a la recta **r**, paralela a todas y que pasa por el ojo. Esto es: toda recta paralela al segmento está contenida en uno de los planos del haz de planos definidos por **r** y en cada uno de esos planos, el punto **R** es el punto de fuga, común a todo segmento o recta en esa dirección.

Visto de otro modo: Dada una dirección en el espacio, todas las rectas en esa dirección (paralelas) se cortan en el mismo punto ideal en el infinito. De entre todas, hay una que pasa por el ojo o centro de proyección, la llamamos línea de fuga **r**. Toda la recta **r** se ve de punta en la imagen, todos sus puntos se proyectan en el mismo punto **R**, entre ellos el punto ideal que pertenece a todas las paralelas.

Esto sucede siempre y cuando las líneas no sean paralelas al plano de la imagen, en ese caso, la línea que pasa por el centro de proyección no intercepta al plano de la imagen en un punto finito, no tiene imagen visible. La imagen no visible de **R** es, en ese caso, un punto ideal o infinito en esa dirección y en plano de la imagen. Ej.: al sacar una foto, con la cámara perfectamente vertical (la posición estándar de una cámara), las líneas verticales son paralelas al plano de la imagen y se ven verticales y paralelas, no tienen un punto de fuga en la imagen sino en el infinito. En caso contrario, cuando se fotografían edificios mirando hacia arriba, sí hay un punto de fuga vertical.

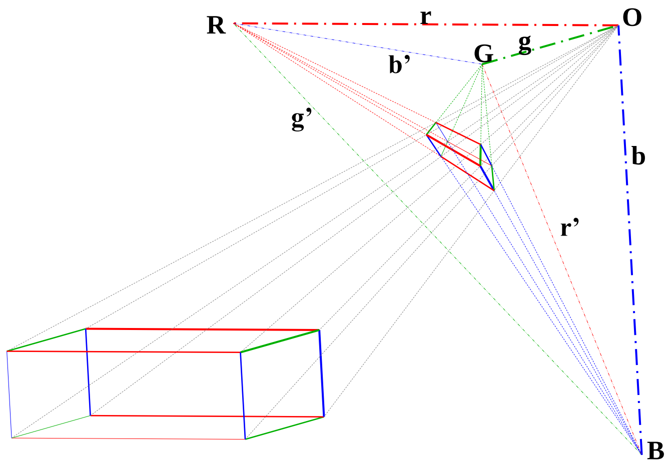


De todos los planos perpendiculares a **r** (y a todos los planos del haz), hay uno: **Γ**, que pasa por el centro de proyección. El plano **Γ** corta al plano **Π** de la imagen en una recta **r'**.

Cualquier línea perpendicular a **r** tendrá una paralela que pasa por el centro **O** y estará contenida en el plano **Γ**. Ésta última cortará al plano de la imagen en algún punto de **r'**: los puntos de fuga de las normales a **r** están sobre la recta **r'**.

Un ladrillo, por ejemplo, tiene tres conjuntos de cuatro aristas paralelas cada uno. Las líneas de conjuntos distintos son perpendiculares. La imagen del ladrillo tiene (en general) tres puntos de fuga.

Cuando alguna dirección es paralela o perpendicular al plano de la imagen aparecen casos particulares: Si la foto se toma paralela a un plano, el punto de fuga de las normales es el centro de la foto. Si se toma paralela a una línea, los puntos de fuga de las normales están en una línea que pasa por el centro (horizonte a la altura de ojo).



En la figura derecha, se muestra la geometría que induce la imagen proyectada de un ladrillo. El plano de la imagen contiene a los tres puntos de fuga **R**, **G** y **B**, que corresponden a los segmentos de colores rojo, verde y azul. El triedro formado por el centro **O** y los puntos de fuga (**O**, **R**, **G**, **B**) sostiene las tres líneas de fuga **r**, **g** y **b**, que son perpendiculares entre sí. Cada par de puntos de fuga define una línea donde están los puntos de fuga de cualquier perpendicular al restante; por ejemplo: cualquier segmento perpendicular a los rojos, tendrá su punto de fuga en algún punto de la línea **GB** = **r'**.