



Departamento de Física, Facultad de Ciencias, Universidad de Chile.
Las Palmeras 3425, Ñuñoa. Casilla 653, Correo 1, Santiago
FONO: 562 978 7276 FAX: 562 271 2973
E-MAIL: secretaria@fisica.ciencias.uchile.cl

Apuntes de un curso de
FÍSICA MATEMÁTICA

Tercera edición, revisión 080424-10

José Rogan C.
Víctor Muñoz G.

Índice

I	Análisis Tensorial	3
1.	Una breve revisión de álgebra lineal.	5
1.1.	Notación.	5
1.2.	Operaciones vectoriales.	8
1.2.1.	Rotación de vectores.	9
1.2.2.	Productos vectoriales.	11
1.2.3.	Cálculos usando notación de Einstein.	15
2.	Operadores en campos escalares y vectoriales.	19
2.1.	Dibujando campos escalares y vectoriales.	19
2.1.1.	Dibujando campos escalares.	19
2.1.2.	Dibujando campos vectoriales.	20
2.2.	Operadores vectoriales.	21
2.2.1.	Notación del operador integral.	21
2.2.2.	Integrales de línea.	21
2.2.3.	Integrales de superficie.	22
2.2.4.	Integrales de volumen.	23
2.3.	Operadores diferenciales.	24
2.3.1.	Vista física del gradiente.	25
2.3.2.	Vista física de la divergencia.	27
2.3.3.	Vista física del rotor.	30
2.3.4.	Identidades con operadores diferenciales.	33
2.4.	Definiciones integrales de los operadores diferenciales.	34
2.5.	Los teoremas.	35
2.5.1.	Teorema de Gauss.	35
2.5.2.	Teorema de Green.	36
2.5.3.	Teorema de Stokes.	37
2.5.4.	Teorema de Helmholtz.	39
3.	Sistemas de Coordenadas Curvilíneos.	41
3.1.	El vector posición	41
3.2.	El sistema cilíndrico	42
3.3.	Sistema esférico	45
3.4.	Sistemas curvilíneos generales	47
3.4.1.	Coordenadas, vectores base y factores de escala	47

3.4.2.	Geometría diferencial.	49
3.4.3.	El vector desplazamiento	51
3.4.4.	Producto de vectores	51
3.4.5.	La integral de línea	52
3.4.6.	Integral de superficie	52
3.4.7.	La integral de volumen	52
3.4.8.	El gradiente	52
3.4.9.	La divergencia	53
3.4.10.	El rotor	54
3.5.	Gradiente, divergencia y rotor en sistemas cilíndricos y esféricos	56
3.5.1.	Operaciones cilíndricas	56
3.5.2.	Operaciones esféricas	57
4.	Introducción a tensores.	59
4.1.	El tensor de conductividad y la ley de Ohm.	59
4.2.	Notación tensorial general y terminología.	62
4.3.	Transformaciones entre sistemas de coordenadas.	63
4.3.1.	Transformaciones vectoriales entre sistemas cartesianos.	63
4.3.2.	La matriz de transformación.	64
4.3.3.	Resumen de transformaciones de coordenadas.	67
4.3.4.	Transformaciones tensoriales.	68
4.4.	Diagonalización de tensores.	69
4.4.1.	Diagonalización y problema de valores propios.	70
4.5.	Transformaciones tensoriales en sistemas de coordenadas curvilíneos.	76
4.6.	Pseudo-objetos.	77
4.6.1.	Pseudo-vectores.	77
4.6.2.	Pseudo-escalares.	82
4.6.3.	Pseudo-tensores.	83
5.	Sistema de coordenadas no ortogonales.	85
5.1.	Breve recuerdo de transformaciones tensoriales.	85
5.2.	Sistemas de coordenadas no ortogonales.	87
5.2.1.	Un sistema de coordenadas inclinado.	88
5.2.2.	Covarianza, contravarianza y métrica.	90
5.2.3.	Transformaciones de componentes vectoriales contravariantes.	92
5.2.4.	Notación de subíndices y superíndices.	95
5.2.5.	Transformaciones de componentes vectoriales covariantes.	98
5.2.6.	Covarianza y contravarianza en tensores.	101
5.2.7.	Contravarianza y covarianza de derivadas parciales.	103
6.	Determinantes y matrices.	107
6.1.	Determinantes.	107
6.2.	Matrices.	114
6.3.	Matrices ortogonales.	121
6.4.	Matrices Hermíticas, matrices unitarias.	129

6.5.	Diagonalización de matrices.	132
6.6.	Matrices normales.	139
7.	Teoría de grupo.	145
7.1.	Introducción.	145
7.2.	Generadores de grupos continuos.	149
7.3.	Momento angular orbital.	162
7.4.	Grupo homogéneo de Lorentz.	166
7.5.	Covarianza de Lorentz de las ecuaciones de Maxwell.	168
8.	Series infinitas.	175
8.1.	Conceptos fundamentales	175
8.2.	Pruebas de Convergencia	178
8.2.1.	Pruebas de comparación.	178
8.2.2.	Prueba de la raíz de Cauchy.	179
8.2.3.	Prueba de la razón de D' Alembert o Cauchy.	180
8.2.4.	Prueba integral de Cauchy o Maclaurin.	181
8.2.5.	Prueba de Kummer.	183
8.2.6.	Prueba de Raabe.	183
8.2.7.	Prueba de Gauss.	184
8.2.8.	Mejoramiento de convergencia.	185
8.3.	Series alternadas.	186
8.3.1.	Criterio de Leibniz.	186
8.3.2.	Convergencia absoluta.	187
8.4.	Álgebra de series.	188
8.4.1.	Mejoramiento de la convergencia, aproximaciones racionales.	189
8.4.2.	Reordenamiento de series dobles.	190
8.5.	Series de funciones.	192
8.5.1.	Convergencia uniforme.	192
8.5.2.	Prueba M de Weierstrass.	193
8.5.3.	Prueba de Abel.	195
8.6.	Expansión de Taylor.	196
8.6.1.	Teorema de Maclaurin.	197
8.6.2.	Teorema Binomial.	199
8.6.3.	Expansión de Taylor de más de una variable.	201
8.7.	Series de potencias.	201
8.7.1.	Convergencia.	202
8.8.	Convergencia uniforme y absoluta.	202
8.8.1.	Continuidad.	202
8.8.2.	Diferenciación e integración.	202
8.8.3.	Teorema de unicidad.	203
8.8.4.	Inversión de series de potencia.	204
8.9.	Integrales elípticas.	205
8.9.1.	Definiciones.	206
8.9.2.	Expansión de series.	207

8.9.3. Valores límites.	208
8.10. Números de Bernoulli.	209
8.10.1. Funciones de Bernoulli.	211
8.10.2. Fórmula de integración de Euler-Maclaurin.	212
8.11. Función zeta de Riemann.	213
8.11.1. Mejoramiento de la convergencia.	216
8.12. Series asintóticas o semi-convergentes.	216
8.12.1. Función gama incompleta.	217
8.12.2. Integrales coseno y seno.	219
8.12.3. Definición de series asintóticas.	220
8.12.4. Aplicaciones a cálculo numérico.	221
8.13. Productos infinitos.	221
8.13.1. Convergencia de un producto infinito.	222
8.13.2. Funciones seno, coseno y gama.	223
9. Ecuaciones diferenciales.	225
9.1. Ecuaciones diferenciales parciales	225
9.1.1. Ejemplos de PDE.	226
9.1.2. Clases de PDE y característica.	228
9.1.3. Las PDE no lineales.	230
9.1.4. Condiciones de borde.	231
9.2. Ecuaciones diferenciales de primer orden.	231
9.2.1. Variables separables.	232
9.2.2. Ecuaciones diferenciales exactas.	233
9.2.3. Ecuaciones diferenciales ordinarias de primer orden lineales.	234
9.2.4. Conversión a una ecuación integral.	236
9.3. Separación de variables.	237
9.3.1. Coordenadas cartesianas.	237
9.3.2. Coordenadas cilíndricas circulares.	238
9.3.3. Coordenadas polares esféricas.	240

Índice de figuras

1.1. El sistema Cartesiano estandar	6
1.2. Geometría para la rotación vectorial	9
1.3. El producto punto.	12
1.4. El producto cruz.	14
1.5. El arreglo de $3 \times 3 \times 3$ de Levi-Civita	15
2.1. Equipotenciales y líneas de campo eléctrico de dos líneas paralelas de carga.	20
2.2. La integral de línea.	22
2.3. Integrales de superficie.	23
2.4. Superficies de $\Phi = -xy$ constante.	26
2.5. Líneas de campo para $\Phi = -xy$	27
2.6. Volumen diferencial.	28
2.7. Flujo a través de las caras superior e inferior.	28
2.8. Campos vectoriales circulantes y no circulantes.	30
2.9. Camino cerrado para la integral del rotor.	31
2.10. Campos con rotor cero, figura (a) y distinto de cero, figura (b).	32
2.11. La suma de dos volúmenes diferenciales.	36
2.12. La suma de dos volúmenes diferenciales.	36
2.13. La suma de dos superficies diferenciales.	37
2.14. El Teorema de Stokes implica un potencial escalar.	38
3.1. El vector posición	42
3.2. El sistema cilíndrico	44
3.3. El vector posición en el sistema cilíndrico	44
3.4. El sistema polar	45
3.5. Componentes polares de un vector	45
3.6. El sistema esférico	46
3.7. El vector posición en coordenadas esféricas	47
3.8. Coordenadas curvilíneas y vectores bases	48
3.9. Volumen diferencial de un sistema de coordenadas curvilíneas	50
3.10. Orientación de la superficie para la integración curvilínea del rotor	54
3.11. Geometría diferencial para integración curvilínea del rotor	55
4.1. Sistemas rotados.	64
4.2. Componentes del vector.	64
4.3. Vectores base en el sistema primado.	74

4.4. Sistema de la mano derecha.	78
4.5. Vectores en el sistema de la mano derecha.	78
4.6. Sistema de la mano izquierda.	79
4.7. Vectores en el sistema de la mano izquierda.	80
4.8. El paralelogramo.	82
5.1. Los sistemas de coordenadas de la Relatividad Especial.	88
5.2. Un sistema de coordenadas de la Relatividad General.	88
5.3. Un sistema de coordenadas ortonormal y otro inclinado.	89
5.4. Dos sistemas de coordenadas inclinados.	93
5.5. Determinación de la base de vectores contravariante.	99
5.6. Componentes covariantes y contravariantes proyectadas de un vector.	100
6.1. Sistemas de coordenadas cartesianos.	121
6.2. Sistemas de coordenadas rotados en dos dimensiones.	124
6.3. (a) Rotación respecto al eje x_3 en un ángulo α ; (b) Rotación respecto a un eje x'_2 en un ángulo β ; (c) Rotación respecto a un eje x''_3 en un ángulo γ	127
6.4. Vector fijo con coordenadas rotadas.	128
6.5. Elipsoide del momento de inercia.	134
6.6. Vector fijo con coordenadas rotada.	142
7.1. Ilustración de la ecuación (7.13).	150
7.2. Ilustración de $M' = U M U^\dagger$ ecuación (7.42).	155
7.3. Octeto bariónico diagrama de peso para SU(3).	159
7.4. Separación de masa bariónica.	161
7.5. Separación de masa bariónica.	161
8.1. Prueba de comparación.	179
8.2. Comparación de integral con suma de bloques	181
8.3. Rearreglo de serie armónica	189
8.4. Series dobles.	190
8.5. Series dobles.	191
8.6. Series dobles.	191
8.7. Convergencia uniforme.	193
8.8. Péndulo simple	205
8.9. Integrales elípticas.	207
8.10. Función zeta de Riemann.	214
8.11. Sumas parciales.	218

1

Parte I
Análisis Tensorial

Capítulo 1

Una breve revisión de álgebra lineal.

versión final 1.0-080415¹

Este capítulo presenta una rápida revisión del álgebra de vectores y de matrices. No intentamos cubrir por completo estos tópicos sino más bien usarlos como introducción a la notación con subíndices y la convención de suma de Einstein. Estas herramientas nos simplificarán la a menudo complicada manipulación del álgebra lineal.

1.1. Notación.

Una notación estandar y consistente es un hábito muy importante a formar en matemáticas. Una buena notación no sólo facilita los cálculos sino que permite análisis dimensional y ayuda a encontrar y corregir errores. Así comenzamos por explicitar la notación que usaremos a través de los apuntes.

Símbolo	Cantidad
v_i	Una componente de un vector
$M_{i\dots j}$	Un elemento de una matriz o tensor
$[M]$	la matriz completa
\vec{v}	Un vector
\hat{e}_i	Un vector base
\vec{T}	Tensor
\mathcal{L}	Un operador

Cuadro 1.1: Notación

Un vector tridimensional \vec{v} puede ser expresado como

$$\vec{v} = v_x \hat{e}_x + v_y \hat{e}_y + v_z \hat{e}_z , \quad (1.1)$$

donde las componentes (v_x, v_y, v_z) son llamadas las componentes Cartesianas de \vec{v} y $(\hat{e}_x, \hat{e}_y, \hat{e}_z)$ son los vectores bases del sistema de coordenadas. La notación puede ser más eficiente aún si

¹Este capítulo está basado en el primer capítulo del libro: *Mathematical Physics* de Brusse Kusse & Erik Westwig, editorial JOHN WILEY & SONS, INC..

reemplazamos los subíndices con letras (x,y,z) , en las componentes, por subíndices numéricos $(1,2,3)$. Con esto, definimos:

$$\begin{aligned}\hat{e}_1 &= \hat{e}_x & v_1 &= v_x \\ \hat{e}_2 &= \hat{e}_y & v_2 &= v_y \\ \hat{e}_3 &= \hat{e}_z & v_3 &= v_z\end{aligned}\tag{1.2}$$

La ecuación (1.1) se transforma en

$$\vec{v} = v_1\hat{e}_1 + v_2\hat{e}_2 + v_3\hat{e}_3 ,\tag{1.3}$$

o más sucintamente

$$\vec{v} = \sum_{i=1}^3 v_i\hat{e}_i .\tag{1.4}$$

La figura (1.1) muestra esta modificación notacional sobre un típico sistema de coordenadas Cartesiano.

Aunque la notación de subíndices puede ser usada en diferentes tipos de sistemas de coordenadas, en este capítulo limitaremos nuestra discusión al sistema Cartesiano. Los vectores bases Cartesianos son ortonormales y posición independientes. Ortonormal significa que la magnitud de cada vector es unitaria y que ellos son perpendiculares entre ellos. Independiente de la posición significa que los vectores bases no cambian su orientación cuando los movemos a través del espacio. Sistema de coordenadas no-Cartesiano son cubiertos en detalle en el capítulo 3.

La ecuación (1.4) puede ser compactada aún más introduciendo la *convención de suma de Einstein* la cual supone que se suma cada vez que se repiten los subíndices en el mismo término. Por lo tanto

$$\vec{v} = \sum_{i=1}^3 v_i\hat{e}_i = v_i\hat{e}_i .\tag{1.5}$$

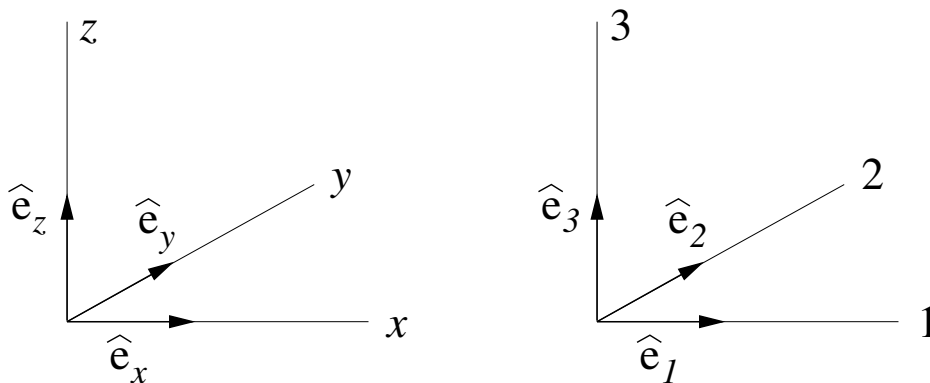


Figura 1.1: El sistema Cartesiano estandar

Nos referimos a la combinación de los subíndices y la convención de suma como la *notación de Einstein*.

Imaginemos ahora que queremos escribir una simple relación vectorial

$$\vec{c} = \vec{a} + \vec{b}. \quad (1.6)$$

Esta ecuación está escrita en lo que se conoce como *notación vectorial*. Notemos que no depende de la elección de un sistema de coordenadas. En un particular sistema de coordenadas, nosotros podemos escribir la relación entre estos vectores en términos de sus componentes:

$$\begin{aligned} c_1 &= a_1 + b_1 \\ c_2 &= a_2 + b_2 \\ c_3 &= a_3 + b_3 \end{aligned} \quad (1.7)$$

Con la notación de subíndices estas tres ecuaciones pueden ser escritas en una sola línea,

$$c_i = a_i + b_i, \quad (1.8)$$

donde el subíndice i se puede reemplazar por cualquiera de los tres valores(1,2,3). Como veremos más adelante el uso de la notación de Einstein puede simplificar drásticamente la derivación de muchas relaciones matemáticas y físicas. Sin embargo, los resultados escritos en esta notación están amarrados a un particular sistema de coordenadas, lo que a menudo dificulta la interpretación. Por esta razón convertiremos nuestros resultados finales de vuelta a una notación vectorial cuando sea posible.

Una matriz es un arreglo dos dimensional de cantidades que puede o no estar asociada con un particular sistema de coordenadas. Las matrices pueden ser expresadas usando diferentes tipos de notación. Si deseamos hablar sobre una matriz como un todo, sin especificar explícitamente todos sus elementos, la escribimos en *notación matricial* como $[M]$. Si, por el contrario necesitamos listar todos los elementos de $[M]$, podemos escribirla como un arreglo rectangular entre un par de paréntesis:

$$[M] = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1c} \\ M_{21} & M_{22} & \cdots & M_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ M_{r1} & M_{r2} & \cdots & M_{rc} \end{bmatrix}. \quad (1.9)$$

Llamaremos a esta *notación de arreglos matriciales* El elemento individual de la tercera fila segunda columna de $[M]$ es escrito como M_{23} . Notemos que la fila de un elemento corresponde al primer índice y la columna al segundo. No todos los arreglos son cuadrados, esto significa que en la ecuación (1.9) r no es necesariamente igual a c .

La multiplicación entre dos matrices es sólo posible si el número de columnas en el pre-multiplicador es igual al número de filas del postmultiplicador. El resultado de tal forma de multiplicación es otra matriz con el mismo número de columnas que el pre-multiplicador y el mismo número de filas que el postmultiplicador. Por ejemplo, el producto entre una matriz 3×2 $[M]$ y una matriz 2×3 $[N]$ forma una matriz de 3×3 $[P]$, con los elementos dados por:

$$\underbrace{\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \\ M_{31} & M_{32} \end{bmatrix}}_{[M]} \underbrace{\begin{bmatrix} N_{11} & N_{12} & N_{13} \\ N_{21} & N_{22} & N_{23} \end{bmatrix}}_{[N]} = \underbrace{\begin{bmatrix} M_{11}N_{11} + M_{12}N_{21} & M_{11}N_{12} + M_{12}N_{22} & M_{11}N_{13} + M_{12}N_{23} \\ M_{21}N_{11} + M_{22}N_{21} & M_{21}N_{12} + M_{22}N_{22} & M_{21}N_{13} + M_{22}N_{23} \\ M_{31}N_{11} + M_{32}N_{21} & M_{31}N_{12} + M_{32}N_{22} & M_{31}N_{13} + M_{32}N_{23} \end{bmatrix}}_{[P]}. \quad (1.10)$$

La multiplicación de la ecuación (1.10) puede ser escrita, en la notación matricial abreviada, como

$$[M][N] = [P] . \quad (1.11)$$

También podemos usar la notación de Einstein para escribir el mismo producto como

$$M_{ij}N_{jk} = P_{ik} , \quad (1.12)$$

con una suma implícita sobre el índice j . notemos que j está en la segunda posición de el término M_{ij} y en la primera posición de el término N_{jk} , tal que la sumatoria es sobre las columnas de $[M]$ y sobre las filas de $[N]$, tal como era en la ecuación (1.10). La ecuación (1.12) es una expresión para el elemento ik -ésimo de la matriz $[P]$.

La notación de arreglos matriciales es conveniente para hacer cálculos numéricos, especialmente cuando se usan computadores. Cuando derivamos las relaciones entre diversas cantidades en física es a menudo inadecuada porque carece de un mecanismo para mantener la pista de la geometría del sistema de coordenadas. Por ejemplo, en un particular sistema de coordenadas, el vector \vec{v} , puede ser escrito como

$$\vec{v} = 1\hat{e}_1 + 3\hat{e}_2 + 2\hat{e}_3 . \quad (1.13)$$

Cuando realizamos los cálculos es a veces conveniente usar una representación matricial del vector escribiendo

$$\vec{v} \rightarrow [v] = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} . \quad (1.14)$$

El problema con esta notación es que no hay una manera conveniente para incorporar los vectores bases en la matriz. Esta es la razón de que fuimos cuidadosos y usamos una flecha (\rightarrow) en la ecuación (1.14) en vez del signo igual ($=$). En estos apuntes un signo igual entre dos cantidades significa que ellas son perfectamente equivalente en todas sus formas. Una cantidad puede ser substituidas por la otra en cualquier expresión. Por ejemplo, la ecuación (1.13) implica que la cantidad $1\hat{e}_1 + 3\hat{e}_2 + 2\hat{e}_3$ puede reemplazar a \vec{v} en cualquier expresión matemática y vice-versa. En contraste la flecha en (1.14) implica que $[v]$ puede representar a \vec{v} y que los cálculos pueden ser realizados usándolo, pero debemos ser cuidadoso no son directamente substituíbles uno por otro sin especificar los vectores bases asociados con las componentes de $[v]$.

1.2. Operaciones vectoriales.

En esta sección veremos varias de las operaciones vectoriales. Usaremos todas las diferentes formas de notación discutidas en la sección previa para ilustrar sus diferencias. Inicialmente, nos concentraremos en la notación matricial y de arreglo matricial. Cuando progreseemos usaremos la notación de Einstein más frecuentemente.

Como discutimos anteriormente un vector tridimensional \vec{v} puede ser representada usando una matriz. Hay realmente dos maneras de escribir esta matriz. Una puede escribirla como

una matriz columna (3×1) o una matriz fila (1×3), cuyos elementos son las componentes de el vector en alguna base Cartesiana:

$$\vec{v} \rightarrow [v] = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \quad \text{o} \quad \vec{v} \rightarrow [v]^\dagger = [v_1 \ v_2 \ v_3] . \quad (1.15)$$

la notación estandar $[v]^\dagger$ es usada para indicar la traspuesta de $[v]$, indicando un intercambio de filas por columnas. Recordemos que el vector \vec{v} puede tener un número infinito de diferentes representaciones de arreglos matriciales, cada una escrita con respecto a una diferente base coordenada.

1.2.1. Rotación de vectores.

Consideremos la rotación simple de un vector en un sistema de coordenadas Cartesiano. Este ejemplo será trabajado, sin pérdida de generalidad, en dos dimensiones.

Partimos con el vector \vec{a} , el cual está orientado en un ángulo θ respecto al eje-1, como muestra la figura 1.2. Este vector puede ser escrito en términos de sus componentes Cartesianas como

$$\vec{a} = a_1 \hat{e}_1 + a_2 \hat{e}_2 . \quad (1.16)$$

donde

$$a_1 = a \cos \theta \quad a_2 = a \text{ sen } \theta . \quad (1.17)$$

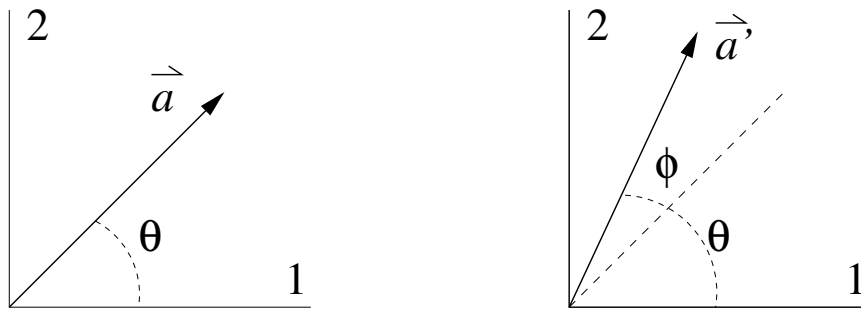


Figura 1.2: Geometría para la rotación vectorial

En esta expresión $a = |\vec{a}| = \sqrt{a_1^2 + a_2^2}$ es la magnitud del vector \vec{a} . El vector \vec{a}' es generado por rotar el vector \vec{a} en el sentido contrario a los punteros del reloj en un ángulo ϕ . Esto cambia la orientación del vector pero no su magnitud. Por lo tanto, podemos escribir

$$\vec{a}' = \underbrace{a \cos(\theta + \phi)}_{a'_1} \hat{e}_1 + \underbrace{a \text{ sen}(\theta + \phi)}_{a'_2} \hat{e}_2 . \quad (1.18)$$

Las componentes a'_1 y a'_2 pueden ser reescritas usando las identidades trigonométricas para seno y el coseno de la suma de ángulos

$$\begin{aligned} a'_1 &= a \cos(\theta + \phi) = \underbrace{a \cos \theta}_{a_1} \cos \phi - \underbrace{a \text{ sen } \theta}_{a_2} \text{ sen } \phi \\ a'_2 &= a \text{ sen}(\theta + \phi) = \underbrace{a \cos \theta}_{a_1} \text{ sen } \phi + \underbrace{a \text{ sen } \theta}_{a_2} \cos \phi \end{aligned} \quad (1.19)$$

Si nosotros representamos a \vec{a} y \vec{a}' como matrices columna.

$$\vec{a} \rightarrow [a] = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad \vec{a}' \rightarrow [a'] = \begin{bmatrix} a'_1 \\ a'_2 \end{bmatrix} . \quad (1.20)$$

La ecuación (1.19) puede ser puesta en forma de arreglo matricial

$$\begin{bmatrix} a'_1 \\ a'_2 \end{bmatrix} = \begin{bmatrix} \cos \phi & -\text{sen } \phi \\ \text{sen } \phi & \cos \phi \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} . \quad (1.21)$$

En notación matricial abreviada, la podemos escribir como

$$[a'] = [R(\phi)] [a] . \quad (1.22)$$

En esta última expresión $[R(\phi)]$ es llamada la matriz de rotación y está claramente definida como

$$[R(\phi)] = \begin{bmatrix} \cos \phi & -\text{sen } \phi \\ \text{sen } \phi & \cos \phi \end{bmatrix} . \quad (1.23)$$

Notemos que para que la ecuación (1.22) sea la misma que la ecuación (1.19), y para que la multiplicación de matrices tenga sentido, las matrices $[a]$ y $[a']$ deben ser matrices columnas y $[R(\phi)]$ debe premultiplicar a $[a]$. El resultado de la ecuación (1.19) también puede escribirse usando una representación fila para $[a]$ y $[a']$. En este caso, las transpuestas de $[R(\phi)]$, $[a]$ y $[a']$ deben ser usadas, y $[R(\phi)]^\dagger$ debe postmultiplicar a $[a]^\dagger$:

$$[a']^\dagger = [a]^\dagger [R(\phi)]^\dagger . \quad (1.24)$$

Escritos usando arreglos de matrices, estas expresiones llegan a ser

$$\begin{bmatrix} a'_1 & a'_2 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} \cos \phi & \text{sen } \phi \\ -\text{sen } \phi & \cos \phi \end{bmatrix} . \quad (1.25)$$

Es fácil ver que la ecuación (1.25) es enteramente equivalente a la ecuación (1.21).

Estas mismas manipulaciones pueden ser logradas usando la notación de Einstein. Por ejemplo, la ecuación (1.19) puede ser expresada como

$$a'_i = R_{ij} a_j . \quad (1.26)$$

La multiplicación de matrices en la ecuación (1.22) suma es sobre las columnas de los elementos de $[R(\phi)]$. Esto se logra en la ecuación (1.26) por la suma implícita sobre j . A diferencia de la notación matricial en la notación de Einstein el orden de a_j y R_{ij} no es ya importante, porque

$$R_{ij} a_j = a_j R_{ij} . \quad (1.27)$$

El vector \vec{a}' puede ser escrito usando la notación de Einstein combinada con la ecuación (1.26) con los vectores bases

$$\vec{a}' = R_{ij} a_j \hat{e}_i . \quad (1.28)$$

Esta expresión demuestra una propiedad de “contabilidad notacional” de la notación de Einstein. La suma sobre un subíndice remueve la dependencia en expresión, de la misma

manera que cuando uno integra sobre una variable. Por esta razón, el proceso de sumar índices es a menudo llamado *contracción* sobre un índice. Hay dos sumas en el lado derecho (LD) de la ecuación (1.28), una sobre i y la otra sobre j . Después de la contracción sobre ambos subíndices, no permanecen subíndices en LD. Esto es consistente con el hecho de que no hay subíndices en el lado izquierdo (LI) de la ecuación. La única notación sobre el LD es una flecha sobre \vec{a} indicando que es un vector, lo cual también existe al LI con el vector unitario \hat{e}_i . Esta suerte de análisis notacional puede ser aplicado a todas las ecuaciones. La notación sobre el LI de un signo igual debe estar siempre de acuerdo con la notación en el LD. Este hecho puede ser usado para chequear las ecuaciones. Por ejemplo,

$$\vec{a}' \neq R_{ij}a_j, \quad (1.29)$$

porque el subíndice i permanece sobre el LD después de contraer sobre j , mientras en el LI no hay subíndices. Adicionalmente, la notación indican que el LI es un cantidad vectorial, mientras el LD no le es.

1.2.2. Productos vectoriales.

Ahora consideraremos los productos punto y cruz de dos vectores usando la notación de Einstein. Este tipo de producto están presente en la física a todo nivel. El producto punto es usualmente encontrado primero cuando calculamos el trabajo W hecho por una fuerza \vec{F} en la integral de línea

$$W = \int d\vec{r} \cdot \vec{F}. \quad (1.30)$$

En esta ecuación, $d\vec{r}$ es un vector desplazamiento diferencial. El producto cruz puede ser usado para encontrar la fuerza sobre una partícula de carga q moviéndose con velocidad \vec{v} en un campo magnético externo \vec{B}

$$\vec{F} = \frac{q}{c}(\vec{v} \times \vec{B}), \quad (1.31)$$

donde c es la velocidad de la luz en el vacío.

El producto punto

El producto punto o interno entre dos vectores \vec{A} y \vec{B} es un escalar definido por

$$\vec{A} \cdot \vec{B} = |\vec{A}||\vec{B}| \cos \theta, \quad (1.32)$$

donde θ es el ángulo entre los dos vectores, como muestra la figura (1.3). Si nosotros tomamos el producto punto de un vector con si mismo tendremos la magnitud al cuadrado de dicho vector

$$\vec{A} \cdot \vec{A} = |\vec{A}|^2. \quad (1.33)$$

En notación de Einstein la ecuación (1.32) se escribe como

$$\vec{A} \cdot \vec{B} = A_i \hat{e}_i \cdot B_j \hat{e}_j. \quad (1.34)$$

Notemos que hemos ocupados dos índices en \vec{A} y \vec{B} , esto es necesario para mantener las sumas independientes de la manipulación que sigue. La contabilidad notacional está trabajando aquí,

porque no hay subíndices en el LD, y ninguno en el LI después de las contracciones sobre ambos i y j . Sólo los vectores bases están involucrados en el producto punto, tal que la ecuación (1.34) puede ser reescrita como

$$\vec{A} \cdot \vec{B} = A_i B_j (\hat{e}_i \cdot \hat{e}_j) . \quad (1.35)$$

Como hemos restringido nuestra atención a sistemas cartesianos donde los vectores bases son ortogonales, tenemos

$$\hat{e}_i \cdot \hat{e}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} . \quad (1.36)$$

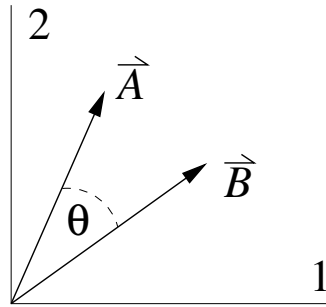


Figura 1.3: El producto punto.

La delta de Kronecker

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} , \quad (1.37)$$

facilita los cálculos que involucran productos puntos. Usándola, podemos escribir $\hat{e}_i \cdot \hat{e}_j = \delta_{ij}$, en la ecuación (1.35) se transforma en

$$\vec{A} \cdot \vec{B} = A_i B_j \delta_{ij} . \quad (1.38)$$

La ecuación (1.38) puede ser expandida haciendo explícitas las sumas sobre ambos índices

$$\vec{A} \cdot \vec{B} = A_1 B_1 \delta_{11} + A_1 B_2 \delta_{12} + A_1 B_3 \delta_{13} + A_2 B_1 \delta_{21} + \dots . \quad (1.39)$$

Ya que la delta de Kronecker es cero a menos que los subíndices sean iguales. La ecuación (1.39) se reduce a sólo tres términos.

$$\vec{A} \cdot \vec{B} = A_1 B_1 + A_2 B_2 + A_3 B_3 = A_i B_i . \quad (1.40)$$

Cuando nos familiaricemos con la notación de Einstein y la delta de Kronecker, estos últimos pasos serán hechos en forma automática. En cualquier momento que aparezca en un término una delta de Kronecker, con uno de sus subíndices repetidos en cualquier otra parte del mismo término, la delta de Kronecker puede ser removida, y cada instancia del subíndice repetido cambiado por el otro subíndice de la delta de Kronecker. Por ejemplo

$$A_i \delta_{ij} = A_j . \quad (1.41)$$

En la ecuación (1.38) la delta de Kronecker puede ser agrupada con el factor B_j , y contraída sobre j para dar

$$A_i(B_j\delta_{ij}) = A_iB_i . \quad (1.42)$$

De la misma manera podemos agruparla con el factor A_i , y sumar sobre i para dar un resultado equivalente

$$B_j(A_i\delta_{ij}) = B_jA_j . \quad (1.43)$$

Esto es cierto para expresiones más complicadas. Por ejemplo,

$$\begin{aligned} M_{ij}(A_k\delta_{ik}) &= M_{ij}A_i \\ &\text{o} \\ B_iT_{jk}(\hat{e}_m\delta_{jm}) &= B_iT_{jk}\hat{e}_j . \end{aligned} \quad (1.44)$$

Esta flexibilidad es una de las cosas que hace los cálculos realizados con notación de Einstein más fácil que trabajar con notación de matrices.

Deberíamos precisar que la delta de Kronecker también puede ser vista como una matriz o arreglo matricial. En tres dimensiones esta representación llega a ser

$$\delta_{ij} \rightarrow [1] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} . \quad (1.45)$$

Esta matriz puede ser usada para escribir la ecuación (1.38) en notación matricial. Notemos que la contracción sobre el índice i suma sobre las filas de la matriz $[1]$, mientras que la contracción sobre j suma sobre las columnas. Así, la ecuación (1.38) en notación matricial es

$$\begin{aligned} \vec{A} \cdot \vec{B} \rightarrow [A]^\dagger [1] [B] &= [A_1 \quad A_2 \quad A_3] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \\ &= [A]^\dagger [B] . \end{aligned} \quad (1.46)$$

El producto cruz

El producto cruz o producto vectorial entre dos vectores \vec{A} y \vec{B} forma un tercer vector \vec{C} , el cual puede ser escrito como

$$\vec{C} = \vec{A} \times \vec{B} . \quad (1.47)$$

La magnitud del vector \vec{C} es

$$|\vec{C}| = |\vec{A}||\vec{B}| \sin \theta , \quad (1.48)$$

donde θ es el ángulo entre los dos vectores, como muestra la figura (1.4). la dirección de \vec{C} depende de que el sistema de coordenadas sea derecho. Por convención, los sistemas de coordenadas tridimensionales en física son usualmente derechos. Extendiendo los dedos de la mano derecha tal que ellos queden alineados con el vector base \hat{e}_1 . Ahora, enrollemoslos hacia el vector base \hat{e}_2 . Si el pulgar apunta a lo largo del vector base \hat{e}_3 el sistema de coordenadas es derecho. Cuando un sistema de coordenadas está dispuesto de esta manera la dirección del producto cruz sigue una regla similar. Para determinar de \vec{C} en la ecuación (1.47), apunte

los dedos a lo largo de \vec{A} , y enrollélos apuntando hacia \vec{B} , el pulgar apuntará la dirección de \vec{C} . Esta definición es a menudo llamada *regla de la mano derecha*. Notemos que la dirección de \vec{C} es siempre perpendicular al plano formado por \vec{A} y \vec{B} . Si por alguna razón, usaremos un sistema zurdo, la definición del producto cruz cambia y deberíamos usar la *regla de la mano izquierda*. Por que la definición del producto cruz cambia levemente cuando movemos la *mano* del sistema de coordenadas, el producto cruz no es exactamente un vector sino más bien un pseudovector. Discutiremos esta distinción más adelante. Por ahora, limitaremos nuestra discusión a sistema de coordenadas derecho, y trataremos el producto cruz como un vector ordinario.

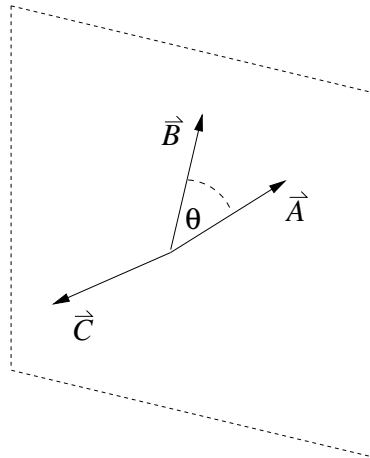


Figura 1.4: El producto cruz.

Otra manera de expresar el producto cruz es usando el determinante de una matriz, donde algunos de sus elementos son los vectores bases:

$$\vec{A} \times \vec{B} = \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix}_{\det} . \quad (1.49)$$

Expandiendo el determinante de la ecuación (1.49) tenemos

$$\vec{A} \times \vec{B} = (A_2 B_3 - A_3 B_2) \hat{e}_1 + (A_3 B_1 - A_1 B_3) \hat{e}_2 + (A_1 B_2 - A_2 B_1) \hat{e}_3 . \quad (1.50)$$

Esta última expresión puede ser escrita usando la notación de Einstein, con la presentación del símbolo de Levi-Civita ϵ_{ijk} :

$$\vec{A} \times \vec{B} = A_i B_j \hat{e}_k \epsilon_{ijk} , \quad (1.51)$$

donde ϵ_{ijk} es definido como

$$\epsilon_{ijk} = \begin{cases} +1 & \text{para } (i, j, k) = \text{a una permutación par de } (1, 2, 3) \\ -1 & \text{para } (i, j, k) = \text{a una permutación impar de } (1, 2, 3) \\ 0 & \text{si dos o más de los subíndices son iguales} \end{cases} . \quad (1.52)$$

Una permutación impar de $(1,2,3)$ es cualquier rearrreglo de estos tres números que pueda ser realizado con un número impar de intercambio de pares. Así, las permutaciones impares de $(1,2,3)$ son $(2,1,3)$, $(1,3,2)$ y $(3,2,1)$. Similarmente las permutaciones pares de $(1,2,3)$ son $(1,2,3)$, $(2,3,1)$ y $(3,1,2)$. Ya que los subíndices i , j y k pueden tomar independientemente los valores $(1,2,3)$, una manera de visualizar el símbolo de Levi-Civita es como un arreglo de $3 \times 3 \times 3$ como lo muestra la figura (1.5)

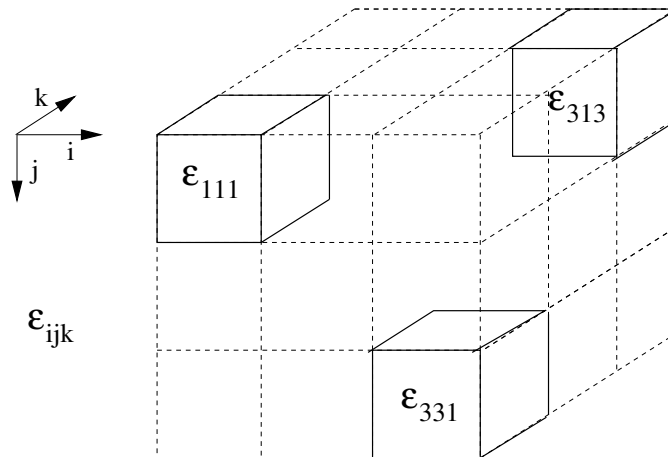


Figura 1.5: El arreglo de $3 \times 3 \times 3$ de Levi-Civita

El producto cruz, escrito usando notación de Einstein en la ecuación (1.51), y el producto punto, escrito en la forma de la ecuación (1.38) son muy útiles para el cálculo manual y lo veremos en los siguientes ejemplos

1.2.3. Cálculos usando notación de Einstein.

Ahora veremos algunos ejemplos para mostrar el uso de la notación de Einstein. El primer ejemplo muestra que la magnitud de un vector no es afectada por rotaciones. El objetivo primario de este ejemplo es mostrar como una derivación que es realizada enteramente con notación matricial también puede ser realizada usando notación de subíndices. El segundo ejemplo deriva una identidad vectorial conocida. Este ejemplo muestra como la notación de subíndices es una poderosa herramienta para derivar complicadas relaciones vectoriales.

Ejemplo 1

Volvamos a la figura de la rotación (1.2), y consideremos el producto $\vec{A} \cdot \vec{A}$ y $\vec{A}' \cdot \vec{A}'$, primero usando notación matricial y luego usando notación de Einstein. Ya que \vec{A}' es generada por una rotación simple de \vec{A} sabemos que estos dos productos puntos, los cuales representan la magnitud al cuadrado de los vectores, debería ser iguales.

Usando matrices:

$$\vec{A} \cdot \vec{A} = [A]^\dagger [A] \quad (1.53)$$

$$\vec{A}' \cdot \vec{A}' = [A']^\dagger [A'] . \quad (1.54)$$

Pero $[A']$ y $[A']^\dagger$ pueden ser expresadas en términos de $[A]$ y $[A]^\dagger$ como

$$[A'] = [R(\phi)][A] \quad [A']^\dagger = [A]^\dagger [R(\phi)]^\dagger, \quad (1.55)$$

donde $R(\phi)$ es la matriz de rotación definida en la ecuación (1.23). Si estas dos ecuaciones son reemplazadas en la ecuación (1.54), tenemos

$$\vec{A}' \cdot \vec{A}' = [A]^\dagger [R(\phi)]^\dagger [R(\phi)] [A]. \quad (1.56)$$

El producto entre las dos matrices de rotación puede realizarse

$$[R(\phi)]^\dagger [R(\phi)] = \begin{bmatrix} \cos \phi & \text{sen } \phi \\ -\text{sen } \phi & \cos \phi \end{bmatrix} \begin{bmatrix} \cos \phi & -\text{sen } \phi \\ \text{sen } \phi & \cos \phi \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (1.57)$$

y la ecuación (1.56) llega a ser

$$\vec{A}' \cdot \vec{A}' = [A]^\dagger [1] [A] = [A']^\dagger [A] \rightarrow \vec{A} \cdot \vec{A}. \quad (1.58)$$

Nuestra conclusión final es que

$$\boxed{\vec{A}' \cdot \vec{A}' = \vec{A} \cdot \vec{A}}. \quad (1.59)$$

Para llegar a este resultado usando matrices, tuvimos cuidado en hacer las operaciones de matrices en el orden correcto.

Ahora repitamos la derivación usando notación de Einstein. La ecuación (1.40) nos permite escribir

$$\vec{A} \cdot \vec{A} = A_i A_i \quad (1.60)$$

$$\vec{A}' \cdot \vec{A}' = A'_j A'_j. \quad (1.61)$$

Notemos que debemos ser cuidadosos en usar diferentes subíndices para las dos sumas en las ecuaciones (1.60) y (1.61). Esto asegura mantenerlas independientes cuando ellas sean manipuladas en los siguientes pasos. Las componentes primas pueden ser expresadas en términos de las componentes sin primas como

$$A'_i = R_{ij} A_j, \quad (1.62)$$

donde R_{ij} es el ij -ésimo elemento de la matriz de rotación $R(\phi)$. Insertando esta expresión en la ecuación (1.61) obtenemos

$$\vec{A}' \cdot \vec{A}' = R_{ru} A_u R_{rv} A_v, \quad (1.63)$$

donde nuevamente hemos sido cuidadosos en usar diferentes subíndices u y v . Esta ecuación tiene tres sumas implícitas, sobre los índices r , u y v .

Un la notación con subíndices, a diferencia de la notación de matrices, el orden de los términos no es importante, así podemos reorganizar la ecuación (1.63) tal que quede

$$\vec{A}' \cdot \vec{A}' = A_u A_v R_{ru} R_{rv}. \quad (1.64)$$

Ahora nos concentramos en la suma sobre r , la cual sólo involucra los elementos de matriz de $[R]$ en el producto $R_{ru}R_{rv}$. ¿Qué significa este producto? Al comparar con las operaciones discutidas previas. En la ecuación (1.12) precisamos la expresión en subíndices $M_{ij}N_{jk}$ representa el producto regular de matrices $[M][N]$ porque el índice sumado j está en la segunda posición de la matriz $[M]$ y en la primera posición en la matriz $[N]$. La expresión $R_{ru}R_{rv}$, sin embargo, tiene una contracción sobre el primer índice en ambas matrices. Para que este producto tenga sentido, escribimos la primera instancia de $[R]$ usando la transpuesta:

$$R_{ru}R_{rv} \rightarrow [R]^\dagger [R] . \quad (1.65)$$

De la ecuación (1.57)

$$R_{ru}R_{rv} = \delta_{uv} . \quad (1.66)$$

Substituyendo este resultado en la ecuación (1.64) nos da

$$\vec{A}' \cdot \vec{A}' = A_u A_v \delta_{uv} = A_u A_v = \vec{A} \cdot \vec{A} . \quad (1.67)$$

Obviamente, este ejemplo es muy fácil. No quedo demostrada ninguna ventaja entre la notación de Einstein y la notación de matrices. Sin embargo, se destaca su equivalencia. En el siguiente ejemplo la notación de Einstein probará ser más indispensable

Ejemplo 2

La notación de Einstein permite la derivación de identidades vectoriales que parecen imposibles usando otra manera. El ejemplo que trabajaremos será la derivación de la identidad del doble producto cruz entre tres vectores $\vec{A} \times (\vec{B} \times \vec{C})$. Este ejemplo muestra la mayoría de las operaciones comunes que ocurren en este tipo de manipulaciones.

La expresión $\vec{A} \times (\vec{B} \times \vec{C})$ está escrita en notación vectorial y es válida en cualquier sistema de coordenadas. Para derivar nuestra identidad, convertiremos esta expresión en notación de Einstein en un sistema de coordenadas Cartesiano. Al final retornaremos a la notación vectorial para obtener un resultado que no dependa de ningún sistema de coordenadas. En este ejemplo, necesitaremos usar la forma de subíndices de un vector

$$\vec{V} = V_i \hat{e}_i , \quad (1.68)$$

Para el producto punto entre dos vectores

$$\vec{A} \cdot \vec{B} = A_i B_i , \quad (1.69)$$

y para el producto cruz

$$\vec{A} \times \vec{B} = A_i B_j \hat{e}_k \epsilon_{ijk} . \quad (1.70)$$

Para comenzar, sea

$$\vec{D} = \vec{B} \times \vec{C} , \quad (1.71)$$

lo cual escribimos usando el símbolo de Levi-Civita como

$$\vec{D} = B_i C_j \hat{e}_k \epsilon_{ijk} . \quad (1.72)$$

Substituyendo la ecuación (1.71) en la expresión $\vec{A} \times (\vec{B} \times \vec{C})$ y usando Levi-Civita nuevamente

$$\vec{A} \times (\vec{B} \times \vec{C}) = A_r D_s \hat{e}_t \epsilon_{rst} . \quad (1.73)$$

La s -ésima componente de \vec{D} es obtenida aplicando el producto punto con \hat{e}_s a ambos lados de la ecuación (1.72) como sigue

$$\begin{aligned} D_s = \hat{e}_s \cdot \vec{D} &= \hat{e}_s \cdot B_i C_j \hat{e}_k \epsilon_{ijk} \\ &= B_i C_j \epsilon_{ijk} (\hat{e}_s \cdot \hat{e}_k) \\ &= B_i C_j \epsilon_{ijk} \delta_{sk} \\ &= B_i C_j \epsilon_{ijs} . \end{aligned} \quad (1.74)$$

Substituyendo el resultado de la ecuación (1.74) en la ecuación (1.73) da

$$\vec{A} \times (\vec{B} \times \vec{C}) = A_r B_i C_j \epsilon_{ijs} \hat{e}_t \epsilon_{rst} , \quad (1.75)$$

lo cual puede ser levemente arreglado para leer

$$\vec{A} \times (\vec{B} \times \vec{C}) = A_r B_i C_j \hat{e}_t \epsilon_{ijs} \epsilon_{rst} . \quad (1.76)$$

Para proceder, necesitamos desarrollar algunas de las propiedades del símbolo de Levi-Civita. Primero, de acuerdo a la definición dada en la ecuación (1.52) es claro que intercambiar cualquier par de índices sólo cambia el signo, *i.e*

$$\epsilon_{ijk} = -\epsilon_{ikj} = \epsilon_{jki} . \quad (1.77)$$

la segunda propiedad involucra el producto de dos símbolos de Levi-Civita que tienen el último índice en común

$$\boxed{\epsilon_{ijk} \epsilon_{mnk} = \delta_{im} \delta_{jn} - \delta_{in} \delta_{jm}} . \quad (1.78)$$

Con una considerable cantidad de esfuerzo se puede mostrar que el LD de la ecuación (1.78) tiene todas las propiedades descrita para el producto de dos símbolos de Levi-Civita en LI.

Con las ecuaciones (1.77) y (1.78) podemos volver a la ecuación (1.76), que ahora puede ser reescrita como

$$\vec{A} \times (\vec{B} \times \vec{C}) = A_r B_i C_j \hat{e}_t (\delta_{rj} \delta_{ti} - \delta_{ri} \delta_{tj}) . \quad (1.79)$$

Después de remover las deltas de Kronecker obtenemos

$$\vec{A} \times (\vec{B} \times \vec{C}) = A_j B_i C_j \hat{e}_i - A_i B_i C_j \hat{e}_j . \quad (1.80)$$

En este punto uno puede realmente ver la utilidad de la notación de Einstein. Los factores en los dos términos del LD de la ecuación (1.80) pueden ser arreglados, agrupados de acuerdo a las sumas, y volver a la notación vectorial ¡en sólo dos líneas! El procedimiento es

$$\vec{A} \times (\vec{B} \times \vec{C}) = (A_j C_j) (B_i \hat{e}_i) - (A_i B_i) (C_j \hat{e}_j) \quad (1.81)$$

$$= (\vec{A} \cdot \vec{C}) \vec{B} - (\vec{A} \cdot \vec{B}) \vec{C} . \quad (1.82)$$

La ecuación (1.81) es válida sólo en un sistema Cartesiano. Como la ecuación (1.82) está en notación vectorial, esta es válida en cualquier sistema de coordenadas.

Capítulo 2

Operadores en campos escalares y vectoriales.

versión final 1.0-080415¹

Un campo es una función que depende del espacio y algunas veces también del tiempo. El potencial eléctrico, la densidad de carga, la temperatura y la presión son sólo una magnitud, y están descritos por campos escalares. En cambio, el campo eléctrico, el campo magnético, la gravedad, la densidad de corriente o la velocidad de un fluido tienen magnitud y dirección y son descritos por campos vectoriales.

Los operadores diferenciales e integrales en campos escalares y vectoriales pueden ser expresados de forma unívoca usando la notación y el formalismo de operadores, los cuales veremos en este capítulo.

2.1. Dibujando campos escalares y vectoriales.

2.1.1. Dibujando campos escalares.

Los dibujos de los campos escalares son mucho más fáciles de construir que los campos vectoriales, ya que los campos escalares están caracterizados por un valor único en cada punto del espacio y del tiempo. Consideremos un ejemplo: el potencial eléctrico Φ producido por dos líneas uniformes con carga $\pm\lambda_0$, las cuales están ubicadas en $(x = \pm 1, y = 0)$. Para este caso, sabemos que

$$\Phi = \lambda_0 \ln \left[\frac{(x+1)^2 + y^2}{(x-1)^2 + y^2} \right]. \quad (2.1)$$

Usualmente queremos construir las superficies donde Φ es constante, usualmente llamadas equipotenciales, contornos o geodésicas, las cuales para este caso son cilindros alrededor de las líneas de carga. Ya que hay simetría en la dirección z , estas superficies pueden ser dibujadas en dos dimensiones como se ve en la figura 2.1. Los centros de estos círculos están ubicados a lo largo del eje x desde $1 < x < \infty$ para los valores positivos de Φ , y desde $-\infty < x < -1$ para los valores negativos de Φ . $\Phi = 0$ se encuentra a lo largo del eje y .

¹Este capítulo está basado en el segundo capítulo del libro: *Mathematical Physics* de Brusse Kusse & Erik Westwig, editorial JOHN WILEY & SONS, INC..

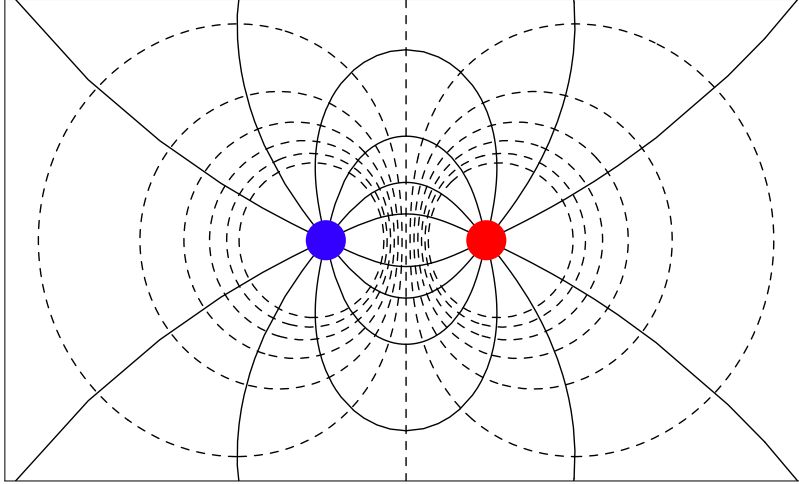


Figura 2.1: Equipotenciales y líneas de campo eléctrico de dos líneas paralelas de carga.

2.1.2. Dibujando campos vectoriales.

Como los vectores poseen magnitud y dirección, los dibujos de los campos que componen son más complicados que los campos vectoriales. Por ejemplo, las componentes cartesianas del campo eléctrico del ejemplo de la sección anterior son

$$E_x = -\frac{\partial\Phi}{\partial x} = 4\lambda_0 \left[\frac{x^2 - y^2 - 1}{[(x-1)^2 + y^2][(x+1)^2 + y^2]} \right] \quad (2.2)$$

$$E_y = -\frac{\partial\Phi}{\partial y} = 4\lambda_0 \left[\frac{2xy}{[(x-1)^2 + y^2][(x+1)^2 + y^2]} \right]. \quad (2.3)$$

Un campo vectorial es dibujado típicamente construyendo líneas tangentes al campo vectorial en cada punto del espacio. Por convención, la densidad de estas líneas de campo indican la magnitud del campo, y flechas muestran su dirección. Si suponemos que las líneas de campo eléctrico que expresan las ecuaciones (2.2) y (2.3) está dada por la ecuación $y = y(x)$, entonces

$$\frac{dy(x)}{dx} = \frac{E_y}{E_x} = \frac{2xy}{x^2 - y^2 - 1}. \quad (2.4)$$

Con un poco de álgebra, la ecuación (2.4) puede ser integrada, obteniendo

$$x^2 + (y - c)^2 = 1 + c^2, \quad (2.5)$$

donde c es una constante de integración. Esta constante puede ser variada desde $-\infty$ a ∞ para generar la familia de líneas de campo. Para este caso, estas líneas son círculos centrados en $y = c$ con un radio dado por $\sqrt{1 + c^2}$. Estas son mostradas como líneas sólidas en la figura 2.1. Las flechas indican como el campo apunta desde la carga positiva a la negativa. Recordemos que donde las líneas están más densamente pobladas (entre las dos cargas) es donde el campo eléctrico es más fuerte.

2.2. Operadores vectoriales.

2.2.1. Notación del operador integral.

El gradiente, la divergencia y el rotor están descritos naturalmente por su forma de operador. Esto es, que ellos son representados por un símbolo que opera sobre otra cantidad. Por ejemplo, el gradiente de Φ es escrito por $\vec{\nabla}\Phi$. Aquí el operador es $\vec{\nabla}$, el cual actúa sobre el operando Φ , lo cual resulta en el gradiente.

En cambio, la integral no es generalmente escrito en su forma de operador. La integral de $f(x)$ sobre x es escrita de la siguiente forma

$$\int f(x) dx , \quad (2.6)$$

la cual no está escrita en su forma de operador ya que la integral y el operando $f(x)$ están mezclados. Sin embargo, podemos poner la ecuación (2.6) en forma de operador reorganizando los términos en la ecuación, como sigue

$$\int dx f(x) . \quad (2.7)$$

Ahora el operador $\int dx$ actúa sobre $f(x)$ para formar la integral, tal como el operador $\vec{\nabla}$ actúa sobre Φ para formar el gradiente. En la práctica, el operador integral es colocado en la derecha, pasando a través de todos los términos del integrando que no dependen de la variable de integración. Por ejemplo,

$$\int dx x^2(x+y)y^2 = y^2 \int dx x^2(x+y) . \quad (2.8)$$

2.2.2. Integrales de línea.

El proceso de tomar una integral a lo largo de un camino es llamado integral de línea y es una operación común en todas las ramas de la Física. Por ejemplo, el trabajo que una fuerza \vec{F} realiza cuando se mueve a través de un camino C es

$$W = \int_C d\vec{r} \cdot \vec{F} . \quad (2.9)$$

Aquí el operador integral de línea $\int_C d\vec{r}$ actúa sobre la fuerza \vec{F} . El vector de desplazamiento diferencial $d\vec{r}$ es tangencial a cada punto a lo largo de C , como es mostrado en la figura 2.2. Si C se cierra sobre sí mismo, escribimos el operador con un círculo sobre el signo de integral,

$$\oint_C d\vec{r} . \quad (2.10)$$

Ya que la ecuación (2.9) está escrita en notación vectorial, es válido en cualquier sistema de coordenadas. En el sistema de coordenadas Cartesiano, $d\vec{r} = dx_i \hat{e}_i$ y la ecuación (2.9) se convierte en

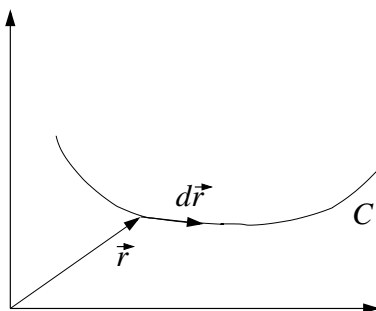


Figura 2.2: La integral de línea.

$$W = \int_C d\vec{r} \cdot \vec{F} = \int_C dx_i F_i . \quad (2.11)$$

Notemos que la cantidad producida por esta integración es un escalar, ya que el subíndice i está sumado implícitamente.

Hay otras operaciones integrales, las cuales son poco comunes. Por ejemplo, el operador

$$\int_C d\vec{r} \Phi = \hat{e}_i \int_C dx_i \Phi \quad (2.12)$$

actúa sobre el escalar Φ para producir un vector. Otro ejemplo,

$$\int_C d\vec{r} \times \vec{v} = \hat{e}_k \int_C dx_i \epsilon_{ijk} v_j , \quad (2.13)$$

genera un vector usando el producto cruz. Notemos que todas las expresiones con subíndices están escritas en el sistema Cartesiano donde la base de vectores es ortonormal e independientes de la posición.

2.2.3. Integrales de superficie.

Las integrales de superficie son representadas por su operador integral

$$\int_S d\vec{\sigma} , \quad (2.14)$$

donde $d\vec{\sigma}$ es un vector que representa un área diferencial. Este vector tiene una magnitud igual a un área diferencial de S , y una dirección perpendicular a la superficie. Si escribimos el diferencial de área como $d\sigma$ y el vector unitario normal \hat{n} , el vector de área diferencial puede ser reescrito como $d\vec{\sigma} = \hat{n} d\sigma$. Como la superficie tiene dos lados, hay un problema para definir \hat{n} . Para una superficie simple y cerrada, como por ejemplo la que se muestra en la figura 2.3(a), definimos \hat{n} para que siempre apunte hacia afuera. Si la superficie no es cerrada, es decir, no encierra un volumen, la dirección de \hat{n} es definida por el camino cerrado C que define los bordes de la superficie, y la regla de la mano derecha, como se muestra en la figura 2.3(b).

Frecuentemente, el operador integral de superficie actúa sobre una cantidad vectorial mediante el producto punto

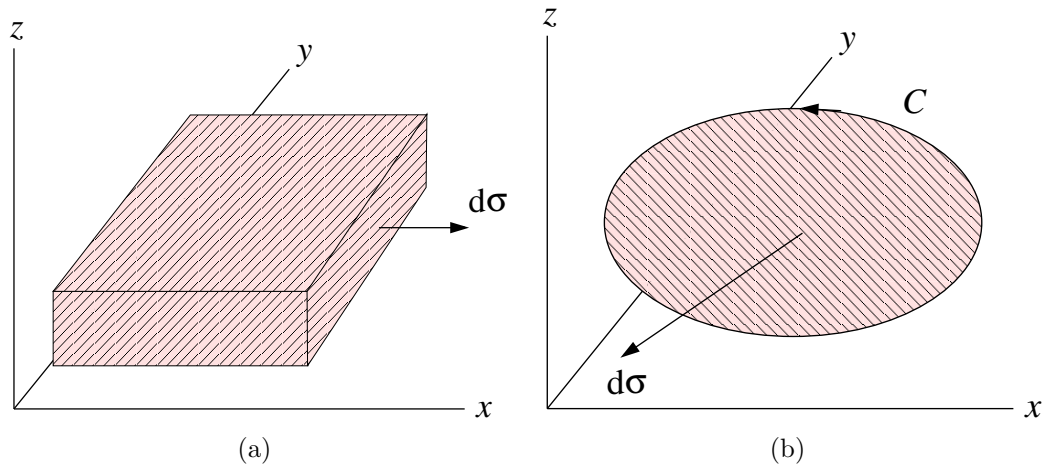


Figura 2.3: Integrales de superficie.

$$\int_S d\vec{\sigma} \cdot \vec{v} . \quad (2.15)$$

En coordenadas cartesianas,

$$d\vec{\sigma} = d\sigma_i \hat{e}_i , \quad (2.16)$$

donde $d\sigma_i$ es positivo ó negativo dependiendo del signo de $\hat{n} \cdot \hat{e}_i$, como se discutió en el párrafo anterior. Esta integral de superficie se transforma en

$$\int_S d\vec{\sigma} \cdot \vec{v} = \int_S d\sigma_i v_i . \quad (2.17)$$

Hay integrales de superficie poco comunes, como

$$\int_S d\vec{\sigma} \Phi = \hat{e}_i \int_S d\sigma_i \Phi , \quad (2.18)$$

la cual es una operación sobre un escalar, la cual produce un vector, y

$$\int_S d\vec{\sigma} \times \vec{v} = \hat{e}_k \int_S d\sigma_i \epsilon_{ijk} v_j \quad (2.19)$$

la cual también produce un vector.

2.2.4. Integrales de volumen.

Las integrales de volumen son los operadores integrales más sencillos, ya que las variables de integración son escalares. Son escritas

$$\int_V d\tau , \quad (2.20)$$

donde $d\tau$ es un volumen diferencial, y V representa el volumen total de integración. La integral de volumen más común actúa sobre una cantidad escalar y, como resultado, produce un escalar

$$\int_V d\tau \Phi . \quad (2.21)$$

En coordenadas cartesianas, esto es escrito como

$$\int_V dx_1 dx_2 dx_3 \Phi . \quad (2.22)$$

Las integrales de volumen de cantidades vectoriales también son posibles,

$$\int_V d\tau \vec{v} = \int_V dx_1 dx_2 dx_3 \vec{v} . \quad (2.23)$$

2.3. Operadores diferenciales.

Por su definición, los campos son funciones de la posición. Análogamente a cómo cambia una función de una variable, lo cual está descrito por su derivada, la dependencia de la posición de un campo escalar puede ser descrito por su gradiente, y la dependencia de la posición de un campo vectorial puede ser descrito por su rotor y su divergencia. El operador nabla $\vec{\nabla}$ es usado para describir estas tres operaciones fundamentales.

El operador $\vec{\nabla}$ está escrito en una notación independiente del sistema de coordenadas. Este puede ser expresado con notación de Einstein en el sistema cartesiano como

$$\vec{\nabla} = \hat{e}_i \frac{\partial}{\partial x_i} . \quad (2.24)$$

Esta expresión será vista en otros sistemas de coordenadas en el capítulo siguiente.

Cuando opera sobre un campo escalar, el operador $\vec{\nabla}$ produce un vector llamado el gradiente

$$\vec{\nabla}\Phi(x_1, x_2, x_3) = \hat{e}_i \frac{\partial\Phi(x_1, x_2, x_3)}{\partial x_i} . \quad (2.25)$$

Por ejemplo, en electrostática el campo eléctrico es igual a menos el gradiente del potencial eléctrico

$$\vec{E} = -\vec{\nabla}\Phi = -\hat{e}_i \frac{\partial\Phi(x_1, x_2, x_3)}{\partial x_i} . \quad (2.26)$$

El operador nabla también actúa sobre campos vectoriales vía el producto punto o el producto cruz. La divergencia de un campo vectorial es una cantidad escalar creada usando el producto punto

$$\vec{\nabla} \cdot \vec{A} = \hat{e}_i \frac{\partial}{\partial x_i} \cdot A_j \hat{e}_j = \frac{\partial A_i}{\partial x_i} . \quad (2.27)$$

La densidad de carga ρ en una región del espacio puede ser calculada usando la divergencia de la relación

$$\vec{\nabla} \cdot \vec{E} = 4\pi\rho . \quad (2.28)$$

En cambio, si utilizamos el producto cruz, generamos una cantidad vectorial llamada el rotor

$$\vec{\nabla} \times \vec{A} = \left(\hat{e}_i \frac{\partial}{\partial x_i} \right) \times A_j \hat{e}_j = \left(\frac{\partial}{\partial x_i} A_j \right) \epsilon_{ijk} \hat{e}_k , \quad (2.29)$$

donde hemos utilizado el símbolo de Levi-Civita para expresar el producto cruz en notación de Einstein. Una de las ecuaciones de Maxwell relaciona el campo eléctrico con la tasa de cambio del campo magnético usando el rotor,

$$\vec{\nabla} \times \vec{E} = -\frac{1}{c} \frac{\partial \vec{B}}{\partial t} . \quad (2.30)$$

2.3.1. Vista física del gradiente.

El gradiente de un campo escalar es un vector que describe, en cada punto, cómo el campo cambia con la posición. Aplicando producto punto a ambos lados de la ecuación (2.25) con $d\vec{r} = dx_i \hat{e}_i$ obtenemos

$$d\vec{r} \cdot \vec{\nabla} \Phi = dx_i \hat{e}_i \cdot \hat{e}_j \frac{\partial \Phi}{\partial x_j} . \quad (2.31)$$

Haciendo un poco de álgebra en el lado derecho de la ecuación, obtenemos

$$d\vec{r} \cdot \vec{\nabla} \Phi = \frac{\partial \Phi}{\partial x_i} dx_i . \quad (2.32)$$

El lado derecho de esta expresión puede ser reorganizado como la diferencia total de carga de Φ debido al cambio diferencial de posición $d\vec{r}$. El resultado puede ser escrito en notación vectorial como sigue

$$d\Phi = \vec{\nabla} \Phi \cdot d\vec{r} . \quad (2.33)$$

De la ecuación (2.33), es claro que el valor máximo de $d\Phi$ ocurre cuando $d\vec{r}$ apunta en la misma dirección que $\vec{\nabla} \Phi$. Por otra parte, un desplazamiento perpendicular a $\vec{\nabla} \Phi$ no produce cambio en Φ , ya que $d\Phi = 0$. Esto significa que el gradiente siempre apuntará perpendicular a las superficies donde Φ es constante.

Al comienzo de este capítulo discutimos la función potencial eléctrico generado por dos líneas de carga. El campo eléctrico fue generado tomando el gradiente de este potencial escalar, y fue dibujado en la figura 2.1. Pudimos haber usado este ejemplo como modelo para desarrollar una vista Física del operador gradiente, pero es un poco complicado. En cambio, observaremos una función de dos dimensiones mucho más simple

$$\Phi = -xy . \quad (2.34)$$

Un dibujo de las líneas equipotenciales en el plano $x-y$ es mostrado en la figura 2.4. Haciendo la operación gradiente obtenemos un vector de campo

$$\vec{\nabla} \Phi = -y \hat{e}_x - x \hat{e}_y . \quad (2.35)$$

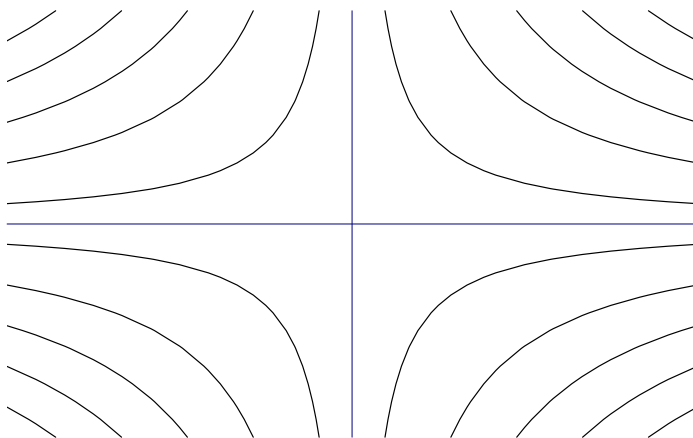


Figura 2.4: Superficies de $\Phi = -xy$ constante.

Ahora imaginemos que estamos en el punto $(1, 2)$ y nos movemos a la derecha una cantidad infinitesimal dr a lo largo del eje x positivo. El cambio correspondiente en Φ puede ser determinado calculando

$$\begin{aligned} d\Phi &= \vec{\nabla}\Phi \cdot d\vec{r} \\ &= (-2\hat{e}_x - 1\hat{e}_y) \cdot (dr\hat{e}_x) \\ &= -2dr . \end{aligned} \quad (2.36)$$

Esto dice que Φ disminuye en 2 unidades por cada paso infinitesimal en esa dirección. En cambio, si estamos sentados en el punto $(3, 4)$ y nos movemos una cantidad infinitesimal dr , con un ángulo de 45° con respecto al eje x , Φ cambia de la siguiente forma

$$\begin{aligned} d\Phi &= \vec{\nabla}\Phi \cdot d\vec{r} \\ &= (-4\hat{e}_x - 3\hat{e}_y) \cdot \frac{dr}{\sqrt{2}}(\hat{e}_x + \hat{e}_y) \\ &= -\frac{7}{\sqrt{2}}dr . \end{aligned} \quad (2.37)$$

Notemos que estos cambios son por pasos infinitesimales. Para calcular el cambio de Φ sobre un camino finito, donde el gradiente cambia mientras nos vamos moviendo punto a punto, necesitamos usar la integral de línea

$$\Delta\Phi = \int_C d\vec{r} \cdot \vec{\nabla}\Phi . \quad (2.38)$$

Cuando utilizamos el gradiente para generar un campo vectorial, usualmente se añade un signo negativo en la definición. Por ejemplo, el campo eléctrico es generado desde el potencial electrostático por

$$\vec{E} = -\vec{\nabla}\Phi . \quad (2.39)$$

Usando esta convención, si nos movemos en contra de las líneas de campo Φ aumenta. Para el potencial de la ecuación (2.34), el gradiente negativo es

$$-\vec{\nabla}\Phi = y\hat{e}_x + x\hat{e}_y . \quad (2.40)$$

Las líneas de campo para esta función pueden ser determinadas como sigue

$$\begin{aligned} \frac{dy}{dx} &= \frac{x}{y} \\ dy y &= dx x \\ y^2 &= x^2 + c \\ x^2 - y^2 &= c . \end{aligned} \quad (2.41)$$

Estas líneas son perpendiculares a las líneas donde Φ es constante, como es mostrado en la figura 2.5. Notemos cómo la densidad de las líneas del campo vectorial muestran que la magnitud del campo aumenta a medida que nos movemos al origen.

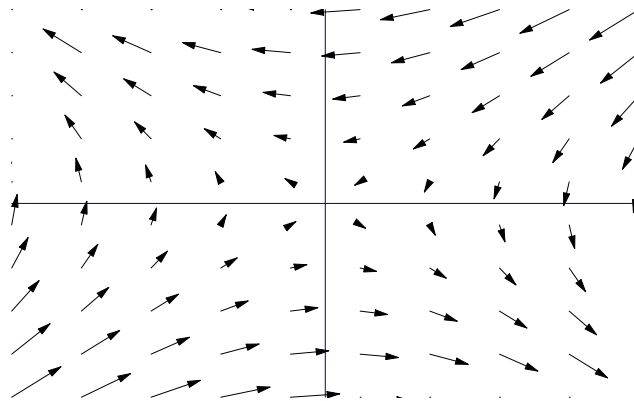


Figura 2.5: Líneas de campo para $\Phi = -xy$.

En resumen, el gradiente de una función escalar Φ genera un campo vectorial el cual, en cada punto indica la dirección de crecimiento de Φ y yacen perpendiculares a las líneas o superficies donde Φ es constante. El ejemplo discutido anteriormente ha sido puesto en práctica en dos dimensiones, pero el proceso también puede ser visualizado en tres dimensiones, donde $\Phi = \text{constante}$ generan superficies, y el gradiente es siempre normal a estas superficies. Salvo esta visualización, no hay límites en el número de dimensiones para el operador gradiente.

2.3.2. Vista física de la divergencia.

El operador divergencia será descrito físicamente desarrollando la ecuación de continuidad, la cual describe el campo en la densidad local de una partícula en un fluido como función del tiempo. Sea $\rho(x, y, z, t)$ el número de partículas por unidad de volumen y $\vec{v}(x, y, z, t)$ la velocidad de estas partículas ubicadas en el punto (x, y, z) y en el tiempo t . Consideremos un volumen diferencial $d\tau = dx dy dz$ localizado en (x_0, y_0, z_0) como se muestra en la figura 2.6. La ecuación de continuidad es obtenida haciendo la suposición que las partículas pueden

entrar ó salir de este volumen, y después equiparar el flujo neto de partículas con cuántas partículas salieron o entraron con el consecuente cambio en ρ . Si llamamos $N \equiv \rho d\tau$ al número total de partículas en el volumen infinitesimal, tenemos

$$\frac{\partial N}{\partial t} = \frac{\partial \rho(x_0, y_0, z_0, t)}{\partial t} dx dy dz . \quad (2.42)$$

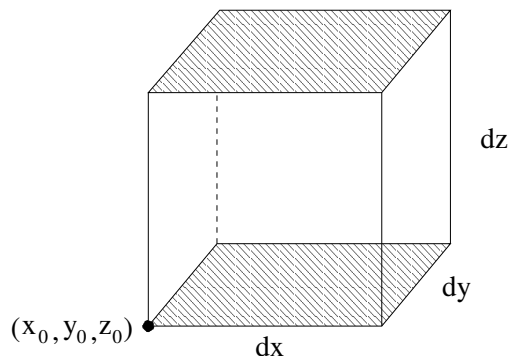


Figura 2.6: Volumen diferencial.

La tasa de cambio de N en la ecuación (2.42) la tomaremos midiendo el flujo de partículas que pasa a través de las seis caras del volumen diferencial $d\tau$. Consideremos la superficie achurada inferior de la figura 2.6. El flujo a través de esta superficie puede ser determinado con la ayuda de la figura 2.7(a). El número total de partículas que entran al volumen en un tiempo dt a través de esta superficie es igual al número de partículas en la región sombreada $dx dy v_z dt$. Notemos que una velocidad positiva v_z agrega partículas a este volumen, mientras que si es negativa sucederá lo contrario. Luego, la contribución inferior a $\partial N/\partial t$ es

$$\frac{\partial N_{\text{inferior}}}{\partial t} = \rho(x_0, y_0, z_0, t) v_z(x_0, y_0, z_0, t) dx dy . \quad (2.43)$$

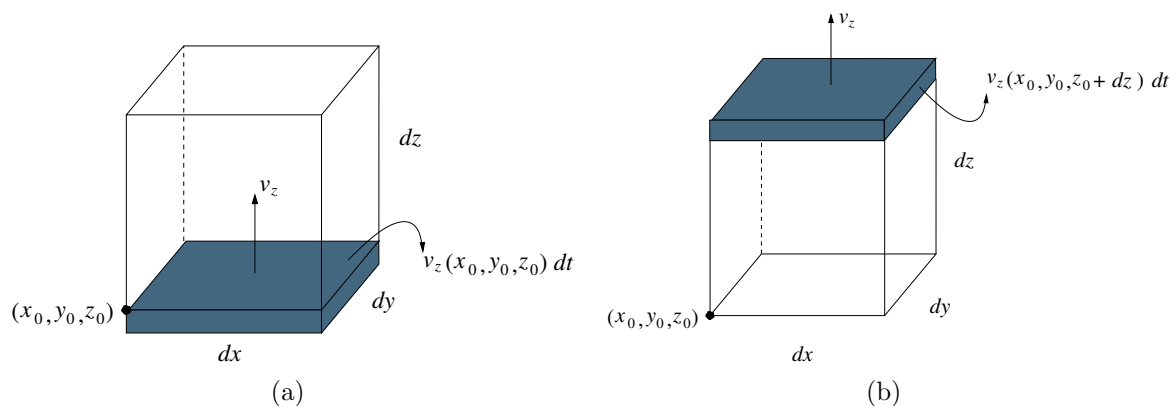


Figura 2.7: Flujo a través de las caras superior e inferior.

Notemos que tanto como ρ y v_z están evaluados en (x_0, y_0, z_0) en la última ecuación. Definamos el vector densidad de corriente como $\vec{J} = \rho \vec{v}$. La ecuación (2.43) puede ser escrita de forma más compacta,

$$\frac{\partial N_{\text{inferior}}}{\partial t} = J_z(x_0, y_0, z_0, t) dx dy . \quad (2.44)$$

El mismo tipo de cálculo se hace de forma análoga para la cara superior mostrada en la figura 2.6. La figura 2.7(b) muestra que ahora v_z positivo acarrea el número de partículas en la región sombreada del volumen. Este lado contribuye

$$\frac{\partial N_{\text{superior}}}{\partial t} = -J_z(x_0, y_0, z_0 + dz, t) dx dy \quad (2.45)$$

al cambio total $\partial N/\partial t$. Notemos que en este caso, evaluamos J_z en el punto $(x_0, y_0, z_0 + dz)$. Combinando las ecuaciones (2.44) y (2.45) obtenemos

$$\frac{\partial N_{\text{inferior}}}{\partial t} + \frac{\partial N_{\text{superior}}}{\partial t} = [J_z(x_0, y_0, z_0, t) - J_z(x_0, y_0, z_0 + dz, t)] dx dy . \quad (2.46)$$

Esta expresión puede ser escrita en términos de la derivada de J_z , ya que en el límite diferencial tenemos

$$J_z(x_0, y_0, z_0 + dz, t) = J_z(x_0, y_0, z_0, t) + \left. \frac{\partial J_z}{\partial z} \right|_{(x_0, y_0, z_0)} dz . \quad (2.47)$$

Substituyendo la ecuación (2.47) en (2.46) obtenemos

$$\frac{\partial N_{\text{inferior}}}{\partial t} + \frac{\partial N_{\text{superior}}}{\partial t} = - \left. \frac{\partial J_z}{\partial z} \right|_{(x_0, y_0, z_0)} dx dy dz . \quad (2.48)$$

Realizando el proceso análogo para las otras cuatro superficies se obtienen resultados similares. Por tanto, el flujo total en el volumen diferencial es

$$\frac{\partial N}{\partial t} = \left[- \left. \frac{\partial J_x}{\partial x} \right|_{(x_0, y_0, z_0)} - \left. \frac{\partial J_y}{\partial y} \right|_{(x_0, y_0, z_0)} - \left. \frac{\partial J_z}{\partial z} \right|_{(x_0, y_0, z_0)} \right] dx dy dz . \quad (2.49)$$

Lo cual es reconocido como $-\vec{\nabla} \cdot \vec{J}$ por $d\tau$. Combinando este resultado con la ecuación (2.42), obtenemos la ecuación de continuidad

$$\frac{\partial \rho}{\partial t} = -\vec{\nabla} \cdot \vec{J} . \quad (2.50)$$

Para una cantidad positiva de la divergencia de \vec{J} , más partículas están dejando la región que entrando en ella, por tanto $\partial \rho/\partial t$ es negativo.

Este ejercicio nos provee la interpretación física de la divergencia. Si la divergencia de un campo vectorial es positiva en una región, la región es una fuente. Las líneas de campo “nacen” en las regiones tipo fuente. Por otra parte, si la divergencia en una región es negativa, la región es considerada un sumidero. Las líneas de campo “mueren” en las regiones tipo sumidero. Si la divergencia de un campo vectorial es cero en una región, todas las líneas de campo que entran deben salir de esa región.

2.3.3. Vista física del rotor.

El rotor de un campo vectorial es un vector, el cual describe en una escala local la circulación del campo. De la palabra rotor parece razonable concluir que si un campo vectorial tiene rotor distinto de cero las líneas de campo deben ser “curvadas”, mientras que si un campo vectorial tiene rotor cero las líneas de campo debiesen ser “rectas”. Esta concepción está errada. Es posible que las líneas de un campo vectorial aparezcan como es mostrado en la figura 2.8(a), describiendo una situación “curvada” y tener rotor igual a cero. También las líneas de campo mostradas en la figura 2.8(b) las cuales son “rectas” pueden tener rotor distinto de cero. Para resolver esta confusión, debemos mirar el rotor en una escala distinta.

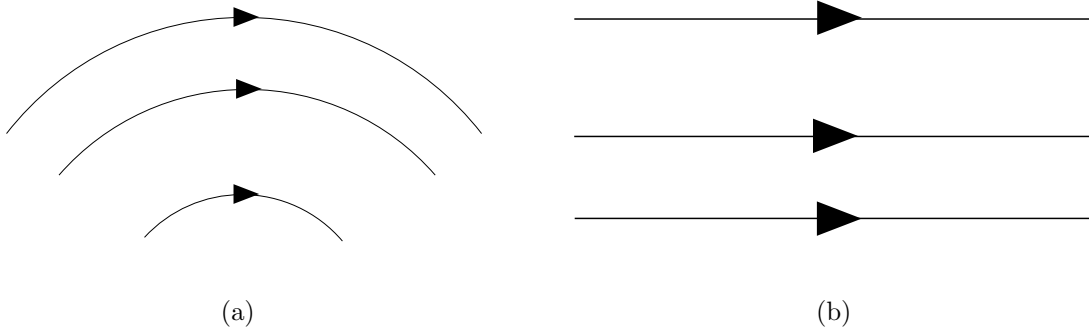


Figura 2.8: Campos vectoriales circulantes y no circulantes.

Consideremos un campo vectorial \vec{v} que sólo es función de x e y . El rotor de este campo apunta en la dirección z , y de acuerdo a la ecuación (2.29) está dado por

$$\vec{\nabla} \times \vec{v} = \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \right) \hat{e}_z \quad (2.51)$$

para un sistema de coordenadas Cartesiano.

Consideremos la integral de línea del campo vectorial \vec{v} alrededor de un camino cerrado, tal como se muestra en la figura 2.9,

$$\oint_C d\vec{r} \cdot \vec{v} . \quad (2.52)$$

El punto de comienzo para la integración es (x_0, y_0) . En esta derivación, necesitamos tener un poco más de cuidado con las cantidades infinitesimales, en contraste con la sección anterior. Por esta razón, imponemos que las dimensiones del camino cerrado sean Δx y Δy , como es mostrado en la figura. Luego comprimiremos estas cantidades a infinitesimales para obtener el resultado final.

La integral a lo largo de C puede ser dividida en cuatro partes. Consideremos la integración a lo largo de C_1 , donde $y = y_0$ y x varía desde x_0 a Δx

$$\int_{C_1} d\vec{r} \cdot \vec{v} = \int_{x_0}^{x_0 + \Delta x} dx v_x . \quad (2.53)$$

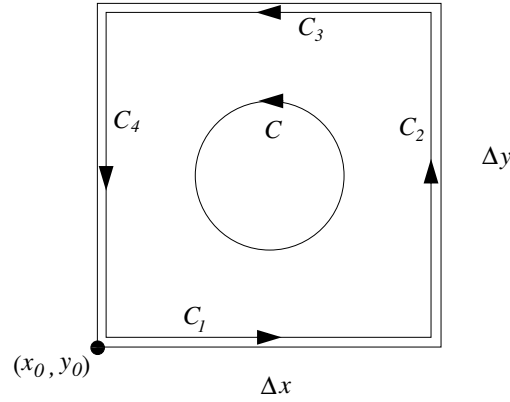


Figura 2.9: Camino cerrado para la integral del rotor.

A lo largo de este segmento podemos expandir $v_x(x, y_0)$ en serie de Taylor, reteniendo el término lineal en x

$$v_x(x, y_0) \approx v_x(x_0, y_0) + \left. \frac{\partial v_x}{\partial x} \right|_{(x_0, y_0)} (x - x_0) . \quad (2.54)$$

No mantendremos los términos de más alto orden, ya que no harán ninguna diferencia significativa en los resultados. Sustituyendo la ecuación (2.54) en (2.53) y realizando la integración, obtenemos

$$\int_{C_1} d\vec{r} \cdot \vec{v} \approx v_x(x_0, y_0) \Delta x + \frac{1}{2} \left. \frac{\partial v_x}{\partial x} \right|_{(x_0, y_0)} (\Delta x)^2 . \quad (2.55)$$

La próxima integración la realizaremos a lo largo de C_3 , la sección superior del camino. A lo largo de este camino, mantendremos fijo $y = y_0 + \Delta y$, mientras que x varía desde x_0 a $x_0 + \Delta x$. Por tanto,

$$\int_{C_3} d\vec{r} \cdot \vec{v} = \int_{x_0 + \Delta x}^{x_0} dx v_x . \quad (2.56)$$

Nuevamente, expandimos en Taylor $v_x(x, y_0 + \Delta y)$ a primer orden

$$v_x(x, y_0 + \Delta y) \approx v_x(x_0, y_0) + \left. \frac{\partial v_x}{\partial x} \right|_{(x_0, y_0)} (x - x_0) + \left. \frac{\partial v_x}{\partial y} \right|_{(x_0, y_0)} \Delta y . \quad (2.57)$$

Reemplazando (2.57) en (2.56) y realizando la integral, obtenemos

$$\int_{C_3} d\vec{r} \cdot \vec{v} \approx -v_x(x_0, y_0) \Delta x - \frac{1}{2} \left. \frac{\partial v_x}{\partial x} \right|_{(x_0, y_0)} (\Delta x)^2 - \left. \frac{\partial v_x}{\partial y} \right|_{(x_0, y_0)} \Delta x \Delta y . \quad (2.58)$$

Combinando las ecuaciones (2.55) y (2.58) obtenemos

$$\int_{C_1} d\vec{r} \cdot \vec{v} + \int_{C_3} d\vec{r} \cdot \vec{v} \approx - \left. \frac{\partial v_x}{\partial y} \right|_{(x_0, y_0)} \Delta x \Delta y . \quad (2.59)$$

Si hacemos el proceso análogo para los caminos C_2 y C_4 , podemos combinar todos los resultados, obteniendo

$$\oint_C d\vec{r} \cdot \vec{v} \approx \left(\frac{\partial v_y}{\partial x} \Big|_{(x_0, y_0)} - \frac{\partial v_x}{\partial y} \Big|_{(x_0, y_0)} \right) \Delta x \Delta y . \quad (2.60)$$

El error de la ecuación (2.60) desaparece cuando las dimensiones del camino disminuyen a dimensiones infinitesimales, es decir cuando $\Delta x \rightarrow 0$ y $\Delta y \rightarrow 0$. Además, utilizando la ecuación (2.51), el término entre paréntesis del lado derecho de la ecuación (2.60) puede ser identificado como la componente z de $\nabla \times \vec{v}$. Por tanto, podemos escribir

$$\lim_{C \rightarrow 0} \oint_C d\vec{r} \cdot \vec{v} = \hat{e}_z \cdot (\vec{\nabla} \times \vec{v}) \lim_{S \rightarrow 0} \int_S d\sigma_z , \quad (2.61)$$

donde C es el contorno que encierra a S y $d\sigma_z = dx dy$ es el área diferencial de esta superficie. ¿Qué nos dice esto acerca del rotor? El resultado en la ecuación (2.61) puede ser reescrito como

$$\hat{e}_z \cdot (\vec{\nabla} \times \vec{v}) = \lim_{C, S \rightarrow 0} \frac{\oint_C d\vec{r} \cdot \vec{v}}{\int_S d\sigma_z} . \quad (2.62)$$

Esto nos dice que la componente z de $\vec{\nabla} \times \vec{v}$ en un punto es la integral de línea de \vec{v} en un camino alrededor de este punto, dividido por el área del camino, en el límite cuando el camino se vuelve muy pequeño. Por tanto, el rotor no nos dice nada acerca de la circulación en una escala macroscópica.

Por tanto, ahora podemos entender las situaciones de la figura 2.8. Si el campo “curvado” mostrado en la figura 2.10(a) tiene una magnitud que decae como $1/r$, exactamente suficiente como para compensar el crecimiento en el camino mientras que r aumenta, luego la integral alrededor del camino diferencial cerrado mostrado en la figura es cero. Por tanto, el rotor en este punto también es cero. Si la magnitud del campo vectorial “recto” mostrado en la figura 2.10(b) varía como indican las líneas de densidad, la integral alrededor del camino cerrado mostrado no puede ser cero y, por tanto, el rotor también tendrá un valor distinto de cero.

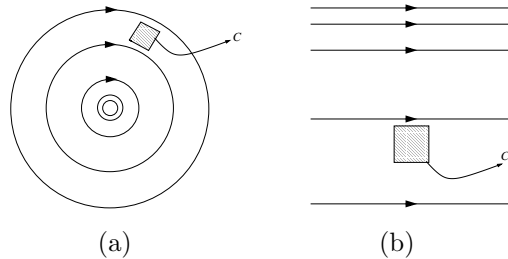


Figura 2.10: Campos con rotor cero, figura (a) y distinto de cero, figura (b).

Hemos derivado la ecuación (2.61) en dos dimensiones y sólo escogimos la componente z del rotor. La generalización de este resultado a tres dimensiones y cualquier orientación del camino diferencial viene dada por

$$\lim_{C \rightarrow 0} \oint_C d\vec{r} \cdot \vec{v} = (\vec{\nabla} \times \vec{v}) \cdot \lim_{S \rightarrow 0} \int_S d\vec{\sigma} . \quad (2.63)$$

2.3.4. Identidades con operadores diferenciales.

La notación de Einstein facilita mucho el trabajo al tener que demostrar igualdades con los operadores diferenciales. Las relaciones presentadas en esta sección son similares a las identidades vectoriales discutidas en el capítulo anterior, excepto que ahora debemos considerar las reglas del cálculo diferencial. Como las identidades vectoriales, utilizaremos el sistema de coordenadas cartesiano, pero los resultados finales están expresados en notación vectorial independiente del sistema de coordenadas.

Ejemplo 1: Consideremos la expresión de operadores $\vec{\nabla} \cdot (\vec{\nabla} \Phi)$. Escribamos esta expresión en notación de Einstein, hagamos la sustitución

$$\vec{\nabla} = \hat{e}_i \frac{\partial}{\partial x_i} . \quad (2.64)$$

Los dos operadores $\vec{\nabla}$ en la expresión original debe ser escrita usando índices independientes

$$\vec{\nabla} \cdot (\vec{\nabla} \Phi) = \hat{e}_i \frac{\partial}{\partial x_i} \cdot \left(\hat{e}_j \frac{\partial}{\partial x_j} \Phi \right) . \quad (2.65)$$

Como los vectores base en el sistema cartesiano son independientes de la posición, $\partial \hat{e}_j / \partial x_i = 0$, y la ecuación (2.65) queda

$$\begin{aligned} \vec{\nabla} \cdot (\vec{\nabla} \Phi) &= (\hat{e}_i \cdot \hat{e}_j) \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_j} \Phi \right) \\ &= \delta_{ij} \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_j} \Phi \right) \\ &= \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_i} \Phi \right) \\ &= \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right) \Phi . \end{aligned} \quad (2.66)$$

En la última línea hemos escrito la suma explícitamente para hacer notar cómo se trabaja con la notación para este caso. La ecuación (2.66) puede ser escrita en notación vectorial definiendo el operador laplaciano ∇^2 como

$$\nabla^2 = \vec{\nabla} \cdot \vec{\nabla} = \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_i} \Phi \right) = \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right) \Phi , \quad (2.67)$$

por tanto

$$\vec{\nabla} \cdot (\vec{\nabla} \Phi) = \nabla^2 \Phi \quad (2.68)$$

Ejemplo 2: Consideremos la expresión $\vec{\nabla} \times \vec{\nabla} \times \vec{v}$, la cual es el rotor del rotor de \vec{v} . Esta identidad será útil cuando desarrollemos la ecuación de las ondas electromagnéticas desde las

ecuaciones de Maxwell. Para escribir esto en notación de Einstein, usaremos los símbolos de Levi-Civita,

$$\vec{\nabla} \times \vec{\nabla} \times \vec{v} = \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_r} v_s \epsilon_{rsj} \right) \epsilon_{ijk} \hat{e}_k . \quad (2.69)$$

El álgebra para encontrar la relación es como sigue

$$\begin{aligned} \vec{\nabla} \times \vec{\nabla} \times \vec{v} &= \frac{\partial}{\partial x_i} \left(\frac{\partial v_s}{\partial x_r} \right) \epsilon_{rsj} \epsilon_{ijk} \hat{e}_k \\ &= -\frac{\partial}{\partial x_i} \left(\frac{\partial v_s}{\partial x_r} \right) \epsilon_{rsj} \epsilon_{ikj} \hat{e}_k \\ &= -\frac{\partial}{\partial x_i} \left(\frac{\partial v_s}{\partial x_r} \right) (\delta_{ri} \delta_{sk} - \delta_{rk} \delta_{si}) \hat{e}_k \\ &= \frac{\partial}{\partial x_i} \left(\frac{\partial v_i}{\partial x_k} \right) \hat{e}_k - \frac{\partial}{\partial x_i} \left(\frac{\partial v_k}{\partial x_i} \right) \hat{e}_k \\ &= \frac{\partial}{\partial x_k} \left(\frac{\partial v_i}{\partial x_i} \right) \hat{e}_k - \frac{\partial}{\partial x_i} \left(\frac{\partial (v_k \hat{e}_k)}{\partial x_i} \right) . \end{aligned} \quad (2.70)$$

Así, el lado derecho de la ecuación (2.70) es convertida a notación vectorial para obtener la igualdad

$$\vec{\nabla} \times \vec{\nabla} \times \vec{v} = \vec{\nabla} \left(\vec{\nabla} \cdot \vec{v} \right) - \nabla^2 \vec{v} . \quad (2.71)$$

Notemos que el operador Laplaciano puede actuar tanto en campos escalares como vectoriales. En la ecuación (2.68) el Laplaciano opera en un campo escalar, obteniendo un escalar. En cambio, en la ecuación (2.71) opera sobre un campo vectorial, obteniendo un vector.

2.4. Definiciones integrales de los operadores diferenciales.

En las ecuaciones (2.25), (2.27) y (2.29) se muestran relaciones para hacer cálculos con la divergencia, el gradiente y el rotor. Cada una de estas relaciones son válidas sólo en un sistema de coordenadas cartesianas y están en términos de las derivadas espaciales de los campos. Las definiciones integrales de cada operador también existen. Ya derivamos la expresión para el rotor en la ecuación (2.63). En esta sección, presentamos definiciones similares para el gradiente y la divergencia. Sus derivaciones, las cuales son similares a la ecuación (2.63) están en los textos de cálculo. Sólo presentaremos los resultados.

El gradiente de un campo escalar en un punto particular puede ser generado por

$$\vec{\nabla} \Phi = \lim_{S, V \rightarrow 0} \frac{\oint_S \Phi d\vec{s}}{\int_V d\tau} , \quad (2.72)$$

donde V es el volumen que incluye el punto de interés y S es la superficie cerrada que encierra a V . Tanto V como S deben ser reducidas a tamaño infinitesimal para que esta relación se cumpla.

Para obtener la divergencia de un campo vectorial en un punto, debemos integrar el campo vectorial sobre una superficie infinitesimal S que encierre al punto, y dividimos por el volumen infinitesimal,

$$\vec{\nabla} \cdot \vec{A} = \lim_{S, V \rightarrow 0} \frac{\oint_S d\vec{\sigma} \cdot \vec{A}}{\int_V d\tau}. \quad (2.73)$$

Ya habíamos obtenido la definición integral para el rotor,

$$\lim_{C \rightarrow 0} \oint_C d\vec{r} \cdot \vec{v} = \left(\vec{\nabla} \times \vec{v} \right) \cdot \lim_{S \rightarrow 0} \int_S d\vec{s}. \quad (2.74)$$

Esta definición es un poco torpe, ya que requiere el cálculo de tres integrales diferentes, cada una con diferentes orientaciones de S , para obtener las tres componentes del rotor. La definición integral que daremos a continuación no tiene este problema, pero usa una forma poco común de integral de superficie

$$\vec{\nabla} \times \vec{A} = \lim_{S, V \rightarrow 0} \frac{\oint_S d\vec{\sigma} \times \vec{A}}{\int_V d\tau}. \quad (2.75)$$

2.5. Los teoremas.

Los operadores diferenciales nos proveen información acerca de la variación de campos escalares y vectoriales en una escala infinitesimal. Para aplicarlos en escala macroscópica necesitamos introducir cuatro teoremas importantes. Estos son el Teorema de Gauss, el Teorema de Green, el Teorema de Stokes y el Teorema de Helmholtz, los cuales pueden ser directamente derivados de las definiciones integrales de los operadores. Damos especial atención en la demostración y discusión del Teorema de Helmholtz ya que no es cubierto adecuadamente en muchos textos.

2.5.1. Teorema de Gauss.

El teorema de Gauss lo podemos deducir de la ecuación (2.73), escribiéndola de una manera ligeramente distinta

$$\vec{\nabla} \cdot \vec{A} d\tau = \lim_{S \rightarrow 0} \oint_S d\vec{\sigma} \cdot \vec{A}. \quad (2.76)$$

En esta ecuación, la superficie cerrada S rodea completamente el volumen $d\tau$, el cual ha sido escrito infinitesimalmente.

La ecuación (2.76) puede ser aplicada en dos volúmenes adyacentes $d\tau_1$ y $d\tau_2$ que tienen una superficie en común, como se muestra en la figura 2.11

$$\vec{\nabla} \cdot \vec{A} d\tau_1 + \vec{\nabla} \cdot \vec{A} d\tau_2 = \oint_{S_1} d\vec{\sigma} \cdot \vec{A} + \oint_{S_2} d\vec{\sigma} \cdot \vec{A}. \quad (2.77)$$

Las contribuciones a la integral de superficie de las superficies comunes se cancelan como se ve en la figura, por lo que la ecuación (2.77) puede ser escrita como

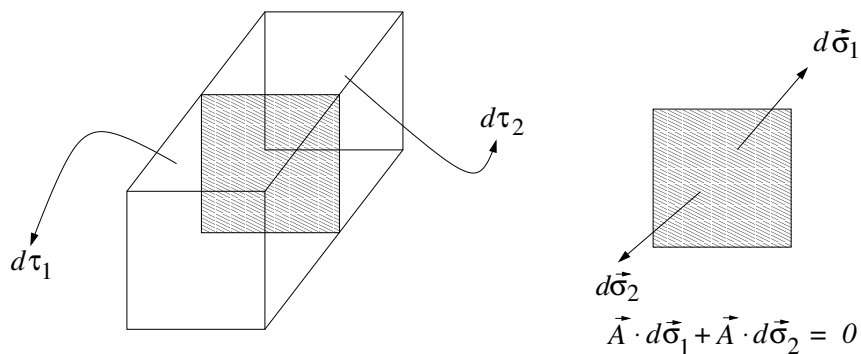


Figura 2.11: La suma de dos volúmenes diferenciales.

$$\vec{\nabla} \cdot \vec{A} d\tau_1 + \vec{\nabla} \cdot \vec{A} d\tau_2 = \oint_{S_{1+2}} d\vec{\sigma} \cdot \vec{A}, \quad (2.78)$$

donde S_{1+2} es la superficie exterior que encierra tanto a $d\tau_1$ como a $d\tau_2$, como es mostrado en la figura 2.12. Podemos continuar este proceso sumando volúmenes diferenciales contiguos para formar un volumen arbitrario V encerrado por una superficie cerrada S . El resultado es llamado el Teorema de Gauss

$$\int_V d\tau \vec{\nabla} \cdot \vec{A} = \oint_S d\vec{\sigma} \cdot \vec{A}. \quad (2.79)$$

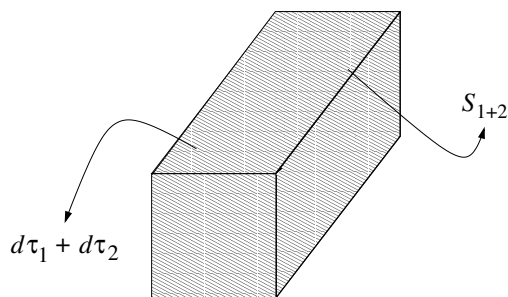


Figura 2.12: La suma de dos volúmenes diferenciales.

2.5.2. Teorema de Green.

El teorema de Green puede ser escrito de dos formas y se puede derivar directamente usando el Teorema de Gauss y algunas manipulaciones algebraicas. Comencemos considerando la expresión $\vec{\nabla} \cdot (u\vec{\nabla}v)$, donde u y v son campos escalares. Usando una identidad de operadores, la cual puede ser demostrada fácilmente, podemos escribir

$$\vec{\nabla} \cdot (u\vec{\nabla}v) = \vec{\nabla}u \cdot \vec{\nabla}v + u\nabla^2v. \quad (2.80)$$

Cambiando u con v , tenemos

$$\vec{\nabla} \cdot (v\vec{\nabla}u) = \vec{\nabla}v \cdot \vec{\nabla}u + v\nabla^2u . \quad (2.81)$$

Restando la ecuación (2.80) con (2.81), tenemos

$$\vec{\nabla} \cdot (u\vec{\nabla}v) - \vec{\nabla} \cdot (v\vec{\nabla}u) = u\nabla^2v - v\nabla^2u . \quad (2.82)$$

Finalmente, integramos ambos lados en la ecuación (2.82) sobre un volumen V , y aplicando el Teorema de Gauss, obtenemos una forma del Teorema de Green

$$\oint_S d\vec{\sigma} \cdot (u\vec{\nabla}v - v\vec{\nabla}u) = \int_V d\tau [u\nabla^2v - v\nabla^2u] . \quad (2.83)$$

En esta expresión, la superficie cerrada S rodea el volumen V . El mismo proceso es aplicado directamente en la ecuación (2.80), con lo cual obtenemos una segunda forma del Teorema de Green

$$\oint_S d\vec{\sigma} \cdot (u\vec{\nabla}v) = \int_V d\tau [\vec{\nabla}u \cdot \vec{\nabla}v + u\nabla^2v] . \quad (2.84)$$

2.5.3. Teorema de Stokes.

El teorema de Stokes se deriva de la ecuación (2.74)

$$(\vec{\nabla} \times \vec{A}) \cdot d\vec{\sigma} = \lim_{C \rightarrow 0} \oint_C d\vec{r} \cdot \vec{A} , \quad (2.85)$$

donde C es el camino que encierra la superficie diferencial $d\vec{\sigma}$.

La deducción del Teorema de Stokes se sigue de forma similar que para el Teorema de Gauss. La ecuación (2.85) es aplicada a dos superficies diferenciales adyacentes que tienen un borde en común, como es mostrado en la figura 2.13. El resultado es

$$(\vec{\nabla} \times \vec{A}) \cdot d\vec{\sigma}_1 + (\vec{\nabla} \times \vec{A}) \cdot d\vec{\sigma}_2 = \oint_{C_{1+2}} d\vec{r} \cdot \vec{A} \quad (2.86)$$

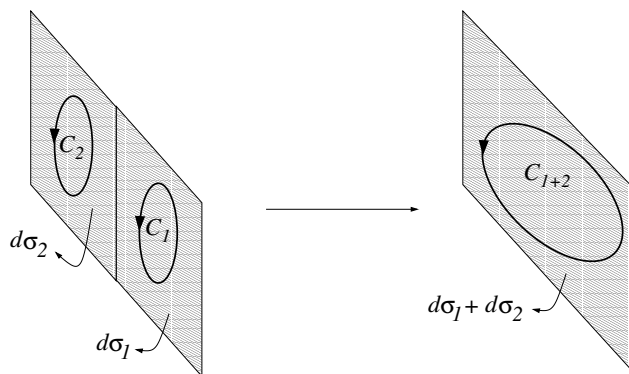


Figura 2.13: La suma de dos superficies diferenciales.

donde el camino C_{1+2} es el camino cerrado por $d\sigma_1$ y $d\sigma_2$. Las integrales de línea a lo largo de los bordes C_1 y C_2 se cancelan. Cualquier número de éstas áreas diferenciales pueden ser

sumadas para formar una superficie S arbitraria y el contorno cerrado C el cual rodea a S . El resultado es el Teorema de Stokes

$$\int_S d\vec{\sigma} \cdot (\vec{\nabla} \times \vec{A}) = \oint_C d\vec{r} \cdot \vec{A}. \quad (2.87)$$

Hay una consecuencia importante del Teorema de Stokes para los campos vectoriales que tienen rotor cero. Tales campos pueden ser siempre derivados desde un potencial escalar. Es decir, si $\vec{\nabla} \times \vec{A} = 0$ en todo el espacio, existe una función escalar $\Phi(\vec{r})$ tal que $\vec{A} = -\vec{\nabla}\Phi$. Para ver esto, consideremos los puntos 1 y 2 y dos caminos A y B arbitrarios entre ellos, como se muestra en la figura 2.14. Una integral de línea cerrada puede ser formada combinando los caminos A y el contrario del camino B . Si $\vec{\nabla} \times \vec{A} = 0$ en todos lados, la ecuación (2.87) nos permite escribir

$$\int_A d\vec{r} \cdot \vec{A} - \int_B d\vec{r} \cdot \vec{A} = \oint d\vec{r} \cdot \vec{A} = 0, \quad (2.88)$$

ó

$$\int_A d\vec{r} \cdot \vec{A} = \int_B d\vec{r} \cdot \vec{A}. \quad (2.89)$$

La ecuación (2.89) nos dice que la integral de línea de \vec{A} entre los dos puntos es independiente del camino que se elija. Esto significa que es posible definir una función escalar de la posición $\Phi(\vec{r})$ tal que su diferencial total esté dado por

$$d\Phi = -d\vec{r} \cdot \vec{A}. \quad (2.90)$$

Es convencional poner el signo negativo tal que Φ aumente cuando se mueve en contra de las líneas de campo de \vec{A} . Reemplazando la ecuación (2.90) en las integrales de línea (2.89) muestra que estas integrales de línea son iguales a

$$\int_1^2 -d\Phi = \Phi(1) - \Phi(2). \quad (2.91)$$

Recordando la ecuación (2.33), es claro que la condición de la ecuación (2.90) puede ser reescrito como

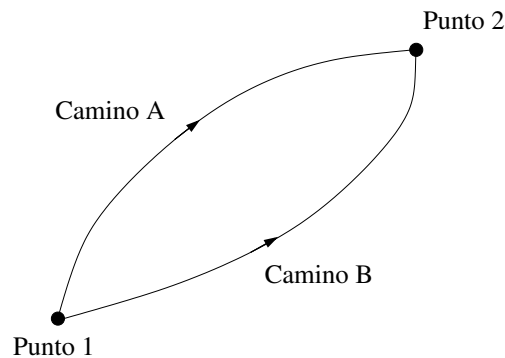


Figura 2.14: El Teorema de Stokes implica un potencial escalar.

$$\vec{A} = -\vec{\nabla}\Phi . \quad (2.92)$$

En resumen, si el rotor de un campo vectorial es cero, el campo es derivable desde un campo escalar. Las integrales de línea de este tipo de campos vectoriales es siempre independiente del camino tomado. Estos tipos de campo son llamados campos conservativos.

2.5.4. Teorema de Helmholtz.

El Teorema de Helmholtz se enuncia de la siguiente manera:

Un campo vectorial, si existe, es determinado en forma única especificando su divergencia y rotor en cualquier punto dentro de una región y su componente normal en la superficie cerrada que rodea esta región.

Hay dos partes muy importantes en este enunciado. Por una parte, dice que si tenemos un campo \vec{v} que estamos tratando de determinar, y conocemos los valores de $\vec{\nabla} \cdot \vec{v}$ y $\vec{\nabla} \times \vec{v}$ en todos los puntos en algún volumen más la componente normal de \vec{v} en la superficie de este volumen, hay un sólo \vec{v} que hará todo el trabajo. Por otra parte, hemos hecho la aclaración “si es que existe”. Esta calificación es necesaria ya que es enteramente posible especificar valores para la divergencia, el gradiente, el rotor y la componente normal de un campo vectorial que no pueden ser satisfechas por cualquier campo.

Para probar el Teorema de Helmholtz supondremos dos campos vectoriales \vec{v}_1 y \vec{v}_2 que poseen los mismos valores de la divergencia, el gradiente, el rotor y la componente normal. Luego, mostraremos que si se da este caso, las dos soluciones deben ser iguales. Además, sea $\vec{w} = \vec{v}_1 - \vec{v}_2$. Ya que la divergencia, el rotor y el producto punto son operadores lineales, \vec{w} debe cumplir las siguientes propiedades

$$\begin{aligned} \vec{\nabla} \cdot \vec{w} &= 0 && \text{en la región} \\ \vec{\nabla} \times \vec{w} &= 0 && \text{en la región} \\ \hat{n} \cdot \vec{w} &= 0 && \text{en la superficie.} \end{aligned} \quad (2.93)$$

Ya que $\vec{\nabla} \times \vec{w} = 0$, \vec{w} puede ser derivado de un potencial escalar

$$\vec{w} = -\vec{\nabla}\Phi . \quad (2.94)$$

Ahora aplicamos el Teorema de Green, en la forma de la ecuación (2.84), con $u = v = \Phi$, obteniendo

$$\oint_S d\vec{\sigma} \cdot \Phi(\vec{\nabla}\Phi) = \int_V d\tau \left[\Phi\vec{\nabla} \cdot (\vec{\nabla}\Phi) + \vec{\nabla}\Phi \cdot \vec{\nabla}\Phi \right] . \quad (2.95)$$

Reemplazando en la ecuación (2.95), obtenemos

$$\oint_S d\vec{\sigma} \cdot \Phi\vec{w} = \int_V d\tau (\Phi\vec{\nabla} \cdot \vec{w} - \vec{w} \cdot \vec{w}) . \quad (2.96)$$

Usando la ecuación (2.93), que la integral de superficie en el lado izquierdo de la ecuación y la integral de volumen de $\Phi \vec{\nabla} \cdot \vec{w}$ son ambas cero y que se cumple

$$\int_V d\tau \vec{w} \cdot \vec{w} = \int_V d\tau |\vec{w}|^2 = 0 . \quad (2.97)$$

Ya que $|\vec{w}|^2$ es siempre una cantidad positiva, la única manera de que se satisfaga la ecuación (2.97) es que se cumpla $\vec{w} = 0$ en todo el espacio. Por tanto, $\vec{v}_1 = \vec{v}_2$ y hemos probado el Teorema de Helmholtz.

El Teorema de Helmholtz es útil para separar los campos vectoriales en dos partes, una con rotor cero y otra con divergencia cero. Esta discusión se apoya en dos identidades

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = 0 \quad (2.98)$$

$$\vec{\nabla} \times \vec{\nabla} \Phi = 0 , \quad (2.99)$$

lo cual puede ser probado fácilmente. Escribimos \vec{v} como

$$\vec{v} = \vec{\nabla} \times \vec{A} - \vec{\nabla} \Phi . \quad (2.100)$$

Luego, podemos escribir

$$\begin{aligned} \vec{\nabla} \cdot \vec{v} &= -\nabla^2 \Phi \\ \vec{\nabla} \times \vec{v} &= \vec{\nabla} \times \vec{\nabla} \times \vec{A} \\ \hat{n} \cdot \vec{v} &= \hat{n} \cdot (\vec{\nabla} \times \vec{A} - \vec{\nabla} \Phi) , \end{aligned} \quad (2.101)$$

ya que la divergencia, el rotor y la componente normal están todas fijas si \vec{A} y Φ están fijos, el Teorema de Helmholtz dice que \vec{v} es único. Notemos que la contribución a \vec{v} que viene de \vec{A} no tiene divergencia, ya que $\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = 0$. Esto es llamado el rotacional o la parte solenoidal del campo y \vec{A} es llamado el potencial vector. La porción de \vec{v} que viene de Φ no tiene rotor, ya que $\vec{\nabla} \times \vec{\nabla} \Phi = 0$. Esto es llamado la parte irrotacional del campo y Φ es llamado el potencial escalar.

Capítulo 3

Sistemas de Coordenadas Curvilíneas.

versión final 1.0-080415¹

Hasta este punto, nuestra discusión de operadores vectoriales, diferenciales e integrales ha estado limitada a sistemas de coordenadas cartesianas. Aunque conceptualmente son simples, estos sistemas a menudo no utilizan la simetría natural de ciertos problemas. Considere el vector campo eléctrico creado por una carga puntual q ubicada en el origen de un sistema cartesiano. Usando una base cartesiana de vectores, este campo es

$$\vec{E} = q \frac{x\hat{x} + y\hat{y} + z\hat{z}}{(x^2 + y^2 + z^2)^{3/2}} . \quad (3.1)$$

En contraste, un sistema esférico, descrito por las coordenadas (r, θ, ϕ) , explota completamente la simetría de éste campo y simplifica la ecuación (3.1) a

$$\vec{E} = \frac{q}{r^2} \hat{r} , \quad (3.2)$$

El sistema esférico pertenece a la clase de sistema de coordenadas *curvilíneas*. Los vectores base de un sistema curvilíneo son ortonormales, tal como los de un sistema cartesiano, pero sus direcciones pueden ser funciones de la posición.

Este capítulo generaliza los conceptos de los capítulos previos para incluir sistemas de coordenadas curvilíneas. Los dos sistemas más comunes, esféricos y cilíndricos son descritos primero con el fin de proporcionar un marco para una discusión más abstracta de coordenadas curvilíneas generalizadas que sigue.

3.1. El vector posición

El vector posición $\vec{r}(P)$ asociado con un punto P describe el desplazamiento desde el origen del sistema de coordenadas. Este tiene una magnitud igual a la distancia desde el origen hasta P y una dirección que apunta desde el origen a este punto.

¹Este capítulo está basado en el tercer capítulo del libro: *Mathematical Physics* de Brusse Kusse & Erik Westwig, editorial JOHN WILEY & SONS, INC..

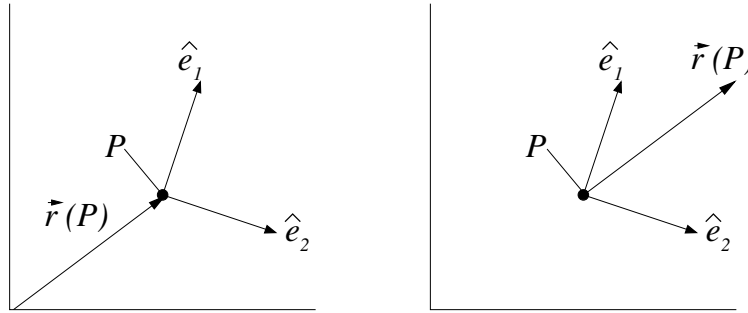


Figura 3.1: El vector posición

Parece natural dibujar el vector posición entre el origen y P como muestra la figura 3.1a. Aunque esto está bien para sistemas de coordenadas cartesianas, esto puede acarrear dificultades en sistemas curvilíneos. Los problemas surgen debido a la dependencia de la posición de los vectores base del sistema curvilíneo. Cuando dibujamos un vector, debemos ser cuidadosos de dónde está ubicado. Si no lo hacemos, podría no ser claro como descomponer el vector en términos de su base. A pesar de esta dificultad, el vector y su base podrían ser dibujados partiendo desde un mismo punto. Las componentes del vector curvilíneo son entonces fácilmente obtenidas proyectando el vector en sus bases. Consecuentemente, para determinar las componentes del vector posición, es mejor dibujarlo, así como sus vectores base, emanando desde P . Esto es mostrado en la figura 3.1b. Hay situaciones, sin embargo, en que es mejor dibujar el vector posición desde el origen. Por ejemplo, integrales de línea, como la mostrada en la figura 2.2 son mejor descritas de esta forma, porque la punta del vector posición sigue el camino de integración. Nosotros ubicaremos el vector posición como se muestra en la figura 3.1, dependiendo de cuál es la más apropiada para la situación dada.

En coordenadas cartesianas, la expresión para el vector posición es intuitiva y simple:

$$\vec{r} = r_i \hat{e}_i = x_i \hat{e}_i \quad (3.3)$$

Las componentes (r_1, r_2, r_3) son fácilmente identificadas como las coordenadas cartesianas (x_1, x_2, x_3) . Formalmente, r_1 es obtenida haciendo el producto punto entre el vector base \hat{e}_1 y el vector posición \vec{r} :

$$r_1 = \hat{e}_1 \cdot \vec{r} = x_1 \quad (3.4)$$

Si bien esto puede parecer exagerada, esta técnica puede ser usada para encontrar las componentes de un vector en cualquier sistema de coordenadas ortogonales.

3.2. El sistema cilíndrico

Las coordenadas de un punto P descrito en un sistema cilíndrico son (ρ, ϕ, z) . Las ecuaciones

$$\begin{aligned}
 x &= \rho \cos \phi \\
 y &= \rho \operatorname{sen} \phi \\
 z &= z
 \end{aligned}
 \tag{3.5}$$

y las correspondientes ecuaciones inversas

$$\begin{aligned}
 \rho &= \sqrt{x^2 + y^2} \\
 \phi &= \tan^{-1}(y/x) \\
 z &= z
 \end{aligned}
 \tag{3.6}$$

gobiernan la relación entre coordenadas cilíndricas y las coordenadas de un superpuesto sistema cartesiano, como muestra la figura 3.2a.

Los vectores base unitarios para el sistema cilíndrico son mostrados en la figura 3.2b. Cada vector base apunta en la dirección en que P se mueve cuando el correspondiente valor de la coordenada es incrementado. Por ejemplo, la dirección de \hat{e}_ρ se encuentra observando como P se mueve al incrementar ρ . Este método puede ser utilizado para determinar la dirección de los vectores base de cualquier conjunto de coordenadas. A diferencia del sistema cartesiano, los vectores base cilíndricos no están fijos. Como el punto P se mueve, las direcciones de \hat{e}_ρ y \hat{e}_ϕ cambian. Notemos también que si P se encuentra exactamente en el eje z , es decir, $\rho = 0$, las direcciones de \hat{e}_ρ y \hat{e}_ϕ se indefinen.

Las coordenadas cilíndricas, tomadas en el orden (ρ, ϕ, z) , forman un sistema de la mano derecha. Si usted alinea su mano derecha a través de \hat{e}_ρ , y entonces rotar sus dedos apuntando en la dirección de \hat{e}_ϕ , su pulgar apuntará en la dirección de \hat{e}_z . Los vectores base son también así

$$\begin{aligned}
 \hat{e}_\rho \cdot \hat{e}_\phi &= \hat{e}_\rho \cdot \hat{e}_z = \hat{e}_z \cdot \hat{e}_\phi = 0 \\
 \hat{e}_\rho \cdot \hat{e}_\rho &= \hat{e}_\phi \cdot \hat{e}_\phi = \hat{e}_z \cdot \hat{e}_z = 1
 \end{aligned}
 \tag{3.7}$$

El vector posición expresado en coordenadas cilíndricas es

$$\vec{r} = (\vec{r} \cdot \hat{e}_\rho) \hat{e}_\rho + (\vec{r} \cdot \hat{e}_\phi) \hat{e}_\phi + (\vec{r} \cdot \hat{e}_z) \hat{e}_z
 \tag{3.8}$$

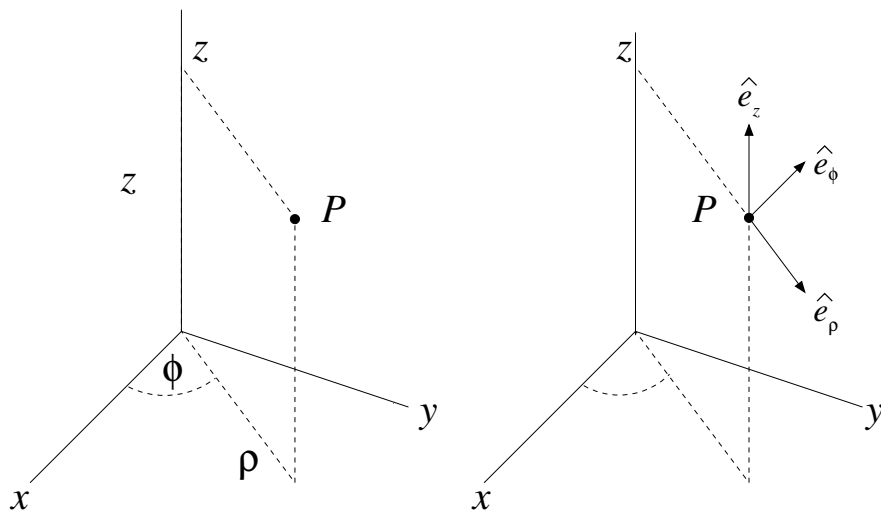


Figura 3.2: El sistema cilíndrico

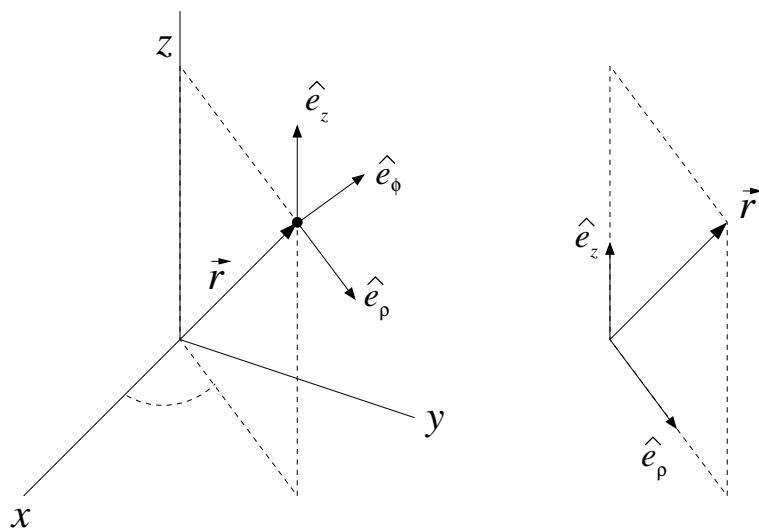


Figura 3.3: El vector posición en el sistema cilíndrico

Notemos que \hat{e}_ϕ está siempre perpendicular a \vec{r} , como se muestra en la figura 3.3, por lo tanto la ecuación (3.8) se reduce a

$$\vec{r} = r_\rho \hat{e}_\rho + r_z \hat{e}_z \quad (3.9)$$

La versión bidimensional del sistema cilíndrico, con sólo las coordenadas (ρ, ϕ) , es llamado un sistema polar plano. Este sistema, mostrado en la figura 3.4a, tiene vectores base \hat{e}_ρ y \hat{e}_ϕ . El vector posición, mostrado en la figura 3.4b, tiene sólo una componente ρ y es expresado como

$$\vec{r} = \rho \hat{e}_\rho \quad (3.10)$$

Recuerde que un vector arbitrario \vec{v} , a diferencia del vector posición, puede tener ambas componentes (ρ y ϕ), como se muestra en la figura 3.5.

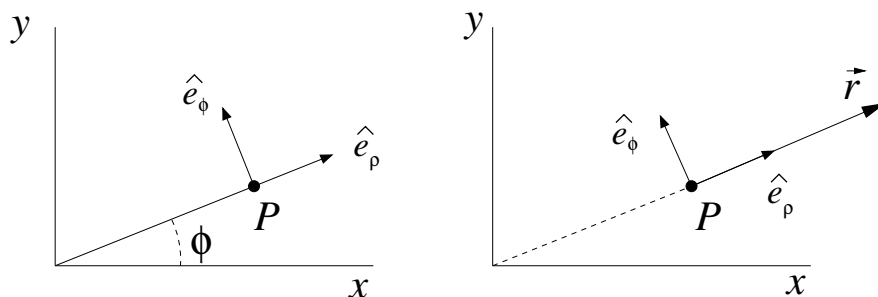


Figura 3.4: El sistema polar

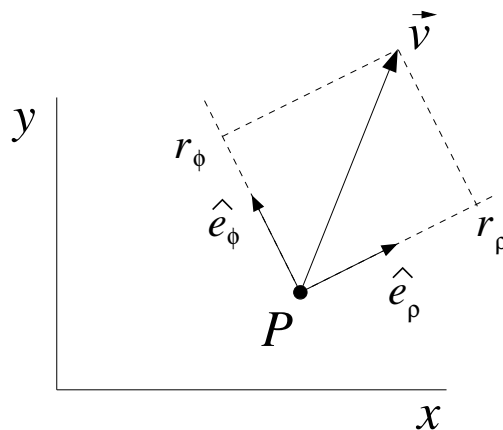


Figura 3.5: Componentes polares de un vector

3.3. Sistema esférico

Las tres coordenadas (r, θ, ϕ) describen un punto en un sistema de coordenadas polares esféricas. Su relación con un conjunto de coordenadas cartesianas se muestra en la figura 3.6. Las ecuaciones

$$\begin{aligned} x &= r \operatorname{sen} \theta \cos \phi \\ y &= r \operatorname{sen} \theta \operatorname{sen} \phi \\ z &= r \cos \theta \end{aligned} \tag{3.11}$$

y las inversas

$$\begin{aligned}
 r &= \sqrt{x^2 + y^2 + z^2} \\
 \phi &= \cos^{-1} \left(\frac{x}{\sqrt{x^2 + y^2}} \right) \\
 \theta &= \cos^{-1} \left(\frac{z}{\sqrt{x^2 + y^2 + z^2}} \right)
 \end{aligned}
 \tag{3.12}$$

permiten una conversión entre los dos sistemas de coordenadas.

La base de vectores unitarios para el sistema esférico se muestra en la figura 3.6b. Como con las coordenadas cilíndricas, obtenemos la dirección de cada vector base incrementando la coordenada asociada y observando como se mueve P . Note como los vectores base cambian con la posición de el punto P . Si P se encuentra en el eje z las direcciones para \hat{e}_θ y \hat{e}_ϕ no están definidas. Si P se encuentra en el origen \hat{e}_r tampoco está definido.

El sistema esférico, con las coordenadas en el orden (r, θ, ϕ) , es un sistema de la mano derecha, tal como en el sistema Cartesiano y en el sistema cilíndrico. También es un sistema ortonormal porque

$$\begin{aligned}
 \hat{e}_r \cdot \hat{e}_\theta &= \hat{e}_r \cdot \hat{e}_\phi = \hat{e}_\theta \cdot \hat{e}_\phi = 0 \\
 \hat{e}_r \cdot \hat{e}_r &= \hat{e}_\theta \cdot \hat{e}_\theta = \hat{e}_\phi \cdot \hat{e}_\phi = 1
 \end{aligned}
 \tag{3.13}$$

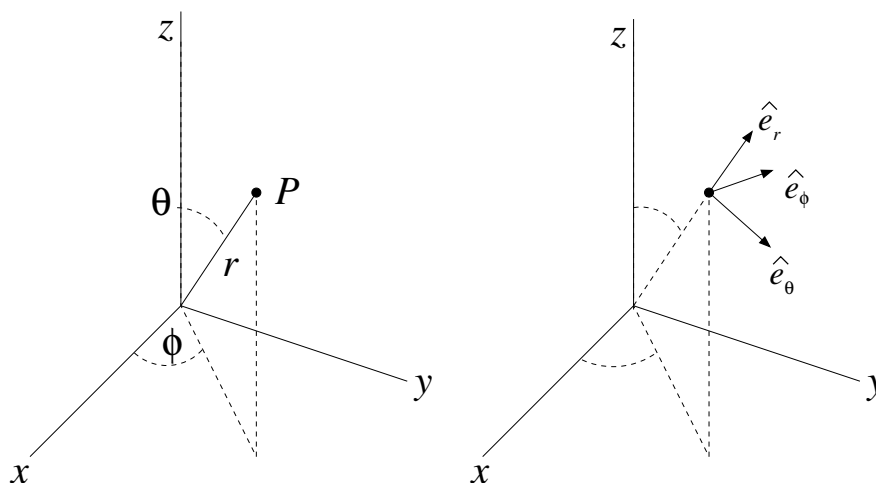


Figura 3.6: El sistema esférico

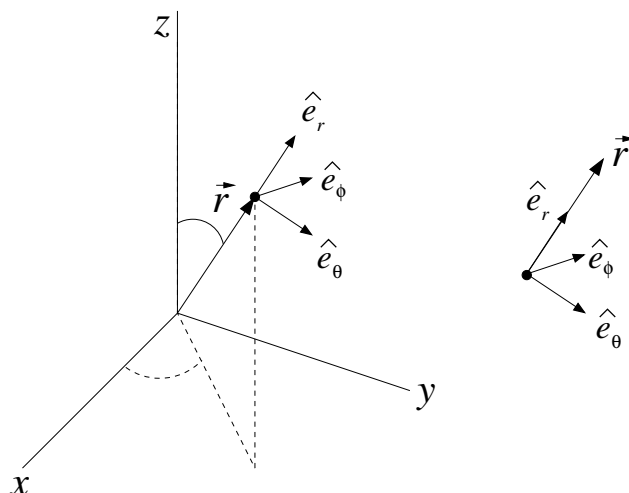


Figura 3.7: El vector posición en coordenadas esféricas

El vector posición, mostrado en la figura 3.7, está expresado en el sistema esférico como

$$\vec{r} = (\vec{r} \cdot \hat{e}_\rho) \hat{e}_\rho + (\vec{r} \cdot \hat{e}_\theta) \hat{e}_\theta + (\vec{r} \cdot \hat{e}_\phi) \hat{e}_\phi \quad (3.14)$$

Como \vec{r} es siempre perpendicular a \hat{e}_θ y a \hat{e}_ϕ , la ecuación (3.14) se simplifica a

$$\vec{r} = r \hat{e}_r \quad (3.15)$$

3.4. Sistemas curvilíneos generales

Aunque los más comunes, el sistema de coordenadas cilíndricas y el sistema de coordenadas polares esféricas son sólo dos ejemplos de una gran familia de sistemas curvilíneos. Un sistema es clasificado como curvilíneo si este tiene vectores base ortonormales, pero no necesariamente constantes. Otros sistemas curvilíneos menos comunes son el toroidal, el hiperbólico y el elíptico. En lugar de trabajar individualmente con las operaciones vectoriales del capítulo anterior para cada uno de estos sistemas, se presenta un enfoque general que pueda abordar cualquier geometría curvilínea.

3.4.1. Coordenadas, vectores base y factores de escala

Las coordenadas (q_1, q_2, q_3) y los correspondientes vectores base \hat{q}_1 , \hat{q}_2 y \hat{q}_3 serán usados para representar cualquier sistema curvilíneo genérico como se ve en la figura 3.8. Debido a que estos vectores base son funciones de posición, deberíamos siempre tener el cuidado de dibujarlos saliendo desde un punto particular, como se mencionó anteriormente en este capítulo.

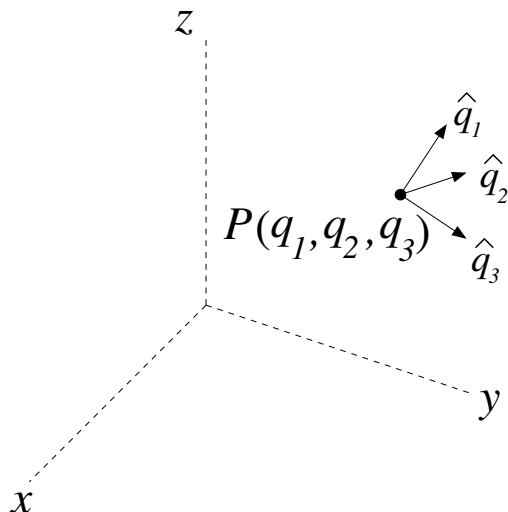


Figura 3.8: Coordenadas curvilíneas y vectores bases

En el sistema de coordenadas cilíndrico y esférico existía un conjunto de ecuaciones que relacionaban sus coordenadas con un conjunto “standard” de coordenadas cartesianas. Para el caso general, escribimos estas ecuaciones como

$$x_i = x_i(q_1, q_2, q_3) \quad (3.16)$$

$$q_i = q_i(x_1, x_2, x_3) , \quad (3.17)$$

Donde el subíndice de la notación se ha arrastrado para mantener las cosas concisas. En estas dos ecuaciones, el subíndice i toma los valores $(1, 2, 3)$. Las variables x_i siempre representan coordenadas Cartesianas, mientras que los q_i son coordenadas curvilíneas generales.

Una expresión para \hat{q}_i , el vector base unitario asociado con la coordenada q_i , puede ser construida incrementando q_i , observando como el vector posición cambia y entonces normalizando:

$$\hat{q}_i = \frac{\partial \vec{r} / \partial q_i}{h_i} \quad (3.18)$$

donde $h_i = |\partial \vec{r} / \partial q_i|$. Esta ecuación es un poco confusa porque no hay una suma sobre el índice i en el lado derecho, aunque el índice aparece dos veces. Esto está sutilmente implícito en la notación, porque hay un subíndice i al lado izquierdo. Los h_i , los cuales a veces son llamados factores de escala, obligan a los vectores base a tener largo unitario. Ellos pueden ser escritos en términos de las coordenadas curvilíneas. Para ver esto, escriba el vector posición en términos de sus componentes Cartesianas, que a su vez se escriben como función de las coordenadas curvilíneas:

$$\vec{r} = x_j(q_1, q_2, q_3) \hat{e}_j \quad (3.19)$$

Para ello,

$$\frac{\partial \vec{r}}{\partial q_i} = \frac{\partial x_j(q_1, q_2, q_3)}{\partial q_i} \hat{e}_j \quad (3.20)$$

y

$$h_i = \left| \frac{\partial \vec{r}}{\partial q_i} \right| = \sqrt{\left(\frac{\partial x_1}{\partial q_i} \right)^2 + \left(\frac{\partial x_2}{\partial q_i} \right)^2 + \left(\frac{\partial x_3}{\partial q_i} \right)^2} \quad (3.21)$$

La interpretación física de los factores de escala es simple. Para un cambio dq_1 de la coordenada q_1 , el vector posición cambia una distancia $|dq_1 h_1|$. Para ello, usando la ecuación (3.18), el vector desplazamiento puede ser escrito en el sistema curvilíneo como

$$\begin{aligned} d\vec{r} &= \frac{\partial \vec{r}}{\partial q_i} dq_i \\ &= dq_i h_i(q_1, q_2, q_3) \hat{q}_i \end{aligned} \quad (3.22)$$

donde ahora hay una suma sobre el índice i en el lado derecho ya que no hay subíndice en el lado izquierdo. Ya que los factores h_i pueden cambiar con la posición, un elemento de volumen diferencial en un sistema curvilíneo, no es necesariamente un cubo como en el sistema Cartesiano. Como veremos en la próxima sección, los lados de un volumen diferencial en un sistema curvilíneo varían en largo y pueden tener curvatura.

3.4.2. Geometría diferencial.

La figura 3.9 representa una superficie diferencial encerrando en volumen infinitesimal en un sistema curvilíneo. Esta figura será la base para la derivación, en coordenadas curvilíneas generales, de la divergencia y el rotor, así como de integrales de superficie y de volumen.

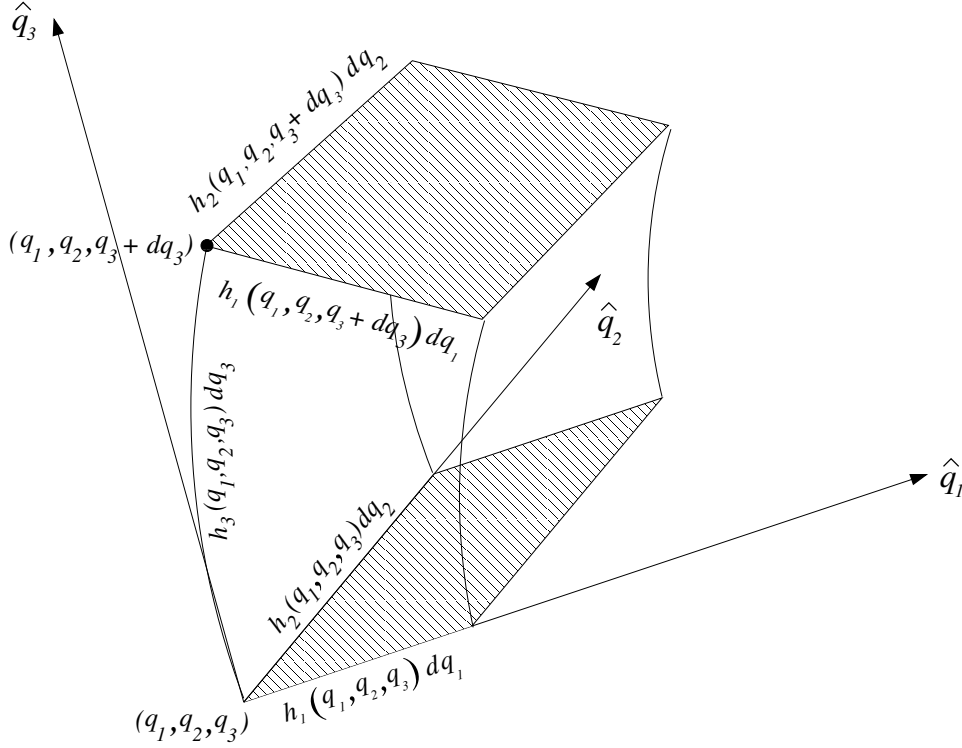


Figura 3.9: Volumen diferencial de un sistema de coordenadas curvilíneas

El volumen está formado escogiendo un punto de partida (q_1, q_2, q_3) y luego construyendo otros siete vértices moviéndose desde este punto con pequeños cambios de coordenadas dq_1, dq_2 y dq_3 . En el límite diferencial, la longitud a lo largo de cada borde del cubo deformado está dado por el correspondiente dq_i veces su factor de escala. El factor de escala se evalúa en un conjunto de coordenadas que corresponde a su valor inicial en el borde. Si la coordenada q_i es igual a $q_i + dq_i$ todos a lo largo de un borde, éste se fija en $q_i + dq_i$. Si la coordenada q_i va desde q_i hasta $q_i + dq_i$ en un borde, éste se fija en q_i . Esta es una manera un poco arrogante de tratar la dependencia de la posición de estos factores, aunque se den todos los resultados correctos para nuestras derivaciones. Un acercamiento más riguroso, en el cual evalúa el valor medio del factor de escala en cada borde puede ser hecha explícitamente.

Siguiendo este acercamiento, un elemento diferencial de volumen es simplemente

$$d\tau = dq_1 dq_2 dq_3 h_1 h_2 h_3 |_{(q_1, q_2, q_3)} , \quad (3.23)$$

donde los h_i están evaluados en el punto (q_1, q_2, q_3) . La superficie diferencial de la cara sombreada inferior es

$$d\vec{\sigma}_{\text{inferior}} = - dq_1 dq_2 h_1 h_2 \hat{q}_3 |_{(q_1, q_2, q_3)} , \quad (3.24)$$

donde el signo menos es debido a que la superficie normal es antiparalela a \hat{q}_3 . Por el contrario, la superficie diferencial de la cara sombreada superior es

$$d\vec{\sigma}_{\text{superior}} = dq_1 dq_2 h_1 h_2 \hat{q}_3 |_{(q_1, q_2, q_3 + dq_3)} , \quad (3.25)$$

El signo menos no aparece porque ahora la superficie es paralela a \hat{q}_3 . En este caso h_1 , h_2 y el vector base \hat{q}_3 están evaluados en el punto $(q_1, q_2, q_3 + dq_3)$.

3.4.3. El vector desplazamiento

El vector desplazamiento $d\vec{r}$ juega un rol central en las matemáticas de sistemas curvilíneos. Una vez que la forma de $d\vec{r}$ es conocida, las ecuaciones para la mayoría de las operaciones vectoriales puede ser fácilmente determinada. Según el cálculo diferencial multi-variable, $d\vec{r}$ puede ser escrito

$$d\vec{r} = \frac{\partial \vec{r}}{\partial q_i} dq_i . \quad (3.26)$$

Como mostramos en la ecuación (3.22), este puede ser escrito usando los factores de escala como

$$d\vec{r} = dq_i h_i \hat{q}_i , \quad (3.27)$$

En un sistema Cartesiano $q_i = x_i$, $\hat{q}_i = \hat{e}_i$ y $h_i = 1$, así la ecuación (3.27) se convierte en la familiar

$$d\vec{r} = dx_i \hat{e}_i . \quad (3.28)$$

En coordenadas cilíndricas, $h_1 = h_\rho = 1$, $h_2 = h_\phi = \rho$ y $h_3 = h_z = 1$ así

$$\begin{aligned} d\vec{r} &= dq_1 \hat{q}_1 + \rho dq_2 \hat{q}_2 + dq_3 \hat{q}_3 \\ &= d\rho \hat{e}_\rho + \rho d\phi \hat{e}_\phi + dz \hat{e}_z . \end{aligned} \quad (3.29)$$

3.4.4. Producto de vectores

Como los sistemas curvilíneos son ortonormales, tenemos que

$$\hat{q}_i \cdot \hat{q}_j = \delta_{ij} . \quad (3.30)$$

Esto significa que el producto punto de dos vectores, \vec{A} y \vec{B} , tienen la misma forma que en un sistema Cartesiano:

$$\vec{A} \cdot \vec{B} = A_i \hat{q}_i \cdot B_j \hat{q}_j = A_i B_j \delta_{ij} = A_i B_i . \quad (3.31)$$

Aquí A_i y B_i son las componentes curvilíneas de los vectores, las cuales pueden ser obtenidas tomando las proyecciones a los ejes paralelos de los vectores en los vectores base:

$$A_i = \vec{A} \cdot \hat{q}_i . \quad (3.32)$$

Con el orden correcto, siempre podemos arreglar nuestras tres coordenadas curvilíneas para ser un sistema de la mano derecha. Entonces, la forma del producto cruz es también

la misma como en un sistema Cartesiano. El producto cruz de \vec{A} y \vec{B} expresado usando los símbolos de Levi-Civita es

$$\vec{A} \times \vec{B} = A_i \hat{q}_i \times B_j \hat{q}_j = A_i B_j \hat{q}_k \epsilon_{ijk} . \quad (3.33)$$

3.4.5. La integral de línea

Usando la expresión para el vector desplazamiento en la ecuación (3.27), la integral de línea en sistemas curvilíneos es sencillamente

$$\int_C d\vec{r} \cdot \vec{v} = \int dq_j h_j \hat{q}_j \cdot v_i \hat{q}_i . \quad (3.34)$$

En el lado derecho de la ecuación hay una suma sobre i y j . Ya que la base de vectores curvilíneos es ortonormal, la integral de línea se transforma en

$$\int_C d\vec{r} \cdot \vec{v} = \int dq_j h_j v_j . \quad (3.35)$$

3.4.6. Integral de superficie

Las integrales de superficies curvilíneas son un poco más complicadas debido a que se debe considerar la orientación de la superficie. Recordando la figura 3.9 y las ecuaciones (3.24) y (3.25), la integral de superficie de un vector \vec{V} es

$$\int_S d\vec{\sigma} \cdot \vec{V} = \int_S \pm dq_1 dq_2 h_1 h_2 V_3 \pm dq_2 dq_3 h_2 h_3 V_1 \pm dq_1 dq_3 h_1 h_3 V_2 , \quad (3.36)$$

donde cada signo más o menos debe ser elegido dependiendo del signo de $d\vec{\sigma} \cdot \hat{q}_i$.

3.4.7. La integral de volumen

La geometría de la figura 3.9 puede ser usada para derivar la forma de integrales de línea en sistemas curvilíneos. El elemento de volumen en el límite infinitesimal, es simplemente $d\tau = dq_1 dq_2 dq_3 h_1 h_2 h_3 \rho(q_1, q_2, q_3)$. Para ello la integral de una función $\rho(\vec{r})$ sobre algún volumen V es expresada como

$$\int_V d\tau \rho(\vec{r}) = \int_V dq_1 dq_2 dq_3 h_1 h_2 h_3 \rho(q_1, q_2, q_3) . \quad (3.37)$$

3.4.8. El gradiente

En el capítulo 2, mostramos como el gradiente de una función escalar Φ se define como

$$d\Phi = \vec{\nabla} \Phi \cdot d\vec{r} . \quad (3.38)$$

Usando la ecuación (3.27) para $d\vec{r}$ tenemos que

$$d\Phi = \vec{\nabla} \Phi \cdot dq_j h_j \hat{q}_j . \quad (3.39)$$

El cálculo diferencial implica que $d\Phi = (\partial\Phi/\partial q_i)dq_i$, así

$$\frac{\partial\Phi}{\partial q_i}dq_i = \vec{\nabla}\Phi \cdot dq_j h_j \hat{q}_j . \quad (3.40)$$

La única forma de que se cumpla la ecuación (3.40) es que

$$\vec{\nabla}\Phi = \frac{1}{h_i} \left(\frac{\partial\Phi}{\partial q_i} \right) \hat{q}_i . \quad (3.41)$$

3.4.9. La divergencia

La operación divergencia en un sistema curvilíneo es más complicada que el gradiente y debe ser obtenida desde la definición de integral

$$\vec{\nabla} \cdot \vec{A} = \lim_{S, V \rightarrow 0} \frac{\oint_S d\vec{\sigma} \cdot \vec{A}}{\int_V d\tau} \quad (3.42)$$

donde S es la superficie cerrada que encierra al volumen V .

Consideremos nuevamente el volumen diferencial de la figura 3.9. El denominador de la ecuación (3.42) para este volumen en el límite infinitesimal es sencillo

$$\int_V d\tau = dq_1 dq_2 dq_3 h_1 h_2 h_3 . \quad (3.43)$$

Para evaluar el numerador, la integración sobre las seis superficies de V debe ser desarrollada. Primero, consideremos las dos caras sombreadas de la figura 3.9, con las normales alineadas paralela o antiparalelamente a \hat{q}_3 . La integral sobre la superficie interior es

$$\int_{\text{inferior}} d\vec{\sigma} \cdot \vec{A} = - dq_1 dq_2 (h_1 h_2 A_3)|_{(q_1, q_2, q_3)} . \quad (3.44)$$

El signo menos surge porque en esta superficie $d\vec{\sigma}$ y \hat{q}_3 están antiparalelas. Note también que A_3 , h_1 y h_2 son todas funciones de las coordenadas curvilíneas y están evaluadas en (q_1, q_2, q_3) , el valor inicial de las coordenadas en esta superficie. La integral sobre la superficie superior es

$$\int_{\text{superior}} d\vec{\sigma} \cdot \vec{A} = dq_1 dq_2 (h_1 h_2 A_3)|_{(q_1, q_2, q_3 + dq_3)} . \quad (3.45)$$

En este caso no hay signo menos porque la superficie normal está orientada paralela a \hat{q}_3 . El valor inicial de la coordenada q_3 ha cambiado en dq_3 comparado con la superficie inferior y por lo tanto A_3 , h_1 y h_2 deben ser evaluados en el punto $(q_1, q_2, q_3 + dq_3)$. En el límite diferencial

$$(h_1, h_2 A_3)|_{(q_1, q_2, q_3 + dq_3)} = (h_1, h_2 A_3)|_{(q_1, q_2, q_3)} + \frac{\partial(h_1, h_2 A_3)}{\partial q_3} \Big|_{(q_1, q_2, q_3)} , \quad (3.46)$$

así la suma de las ecuaciones (3.44) y (3.45) es

$$\int_{\text{ambas}} d\vec{\sigma} \cdot \vec{A} = \frac{\partial(h_1 h_2 A_3)}{\partial q_3} dq_1 dq_2 dq_3 . \quad (3.47)$$

Combinando este resultado con integraciones similares sobre las restantes cuatro superficies

$$\int_S d\vec{\sigma} \cdot \vec{A} = \left[\frac{\partial(h_2 h_3 A_1)}{\partial q_1} + \frac{\partial(h_1 h_3 A_2)}{\partial q_2} + \frac{\partial(h_1 h_2 A_3)}{\partial q_3} \right] dq_1 dq_2 dq_3 . \quad (3.48)$$

Sustituyendo las ecuaciones (3.43) y (3.48) en la ecuación (3.42) obtenemos el resultado

$$\vec{\nabla} \cdot \vec{A} = \frac{1}{h_1 h_2 h_3} \left[\frac{\partial(h_2 h_3 A_1)}{\partial q_1} + \frac{\partial(h_1 h_3 A_2)}{\partial q_2} + \frac{\partial(h_1 h_2 A_3)}{\partial q_3} \right] . \quad (3.49)$$

3.4.10. El rotor

El rotor para un sistema de coordenadas curvilíneas también puede ser derivada desde la definición de integral:

$$\vec{\nabla} \times \vec{A} \cdot \lim_{S \rightarrow 0} \int_S d\vec{\sigma} = \lim_{C \rightarrow 0} \oint_C d\vec{r} \cdot \vec{A} , \quad (3.50)$$

donde C es un camino cerrado que encierra a la superficie S y la dirección de $d\vec{\sigma}$ es definida vía C y por la convención de la mano derecha.

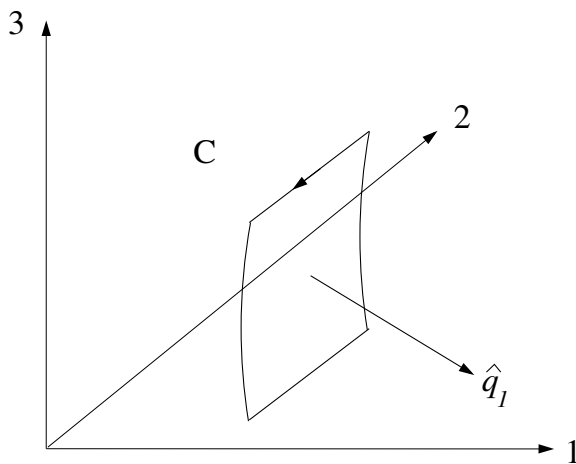


Figura 3.10: Orientación de la superficie para la integración curvilínea del rotor

Una componente del rotor puede ser escogida orientando $d\vec{\sigma}$ en la dirección de un vector base. Consideremos la figura 3.10, donde $d\vec{\sigma}$ está orientado para elegir la componente \hat{q}_1 , en este caso $d\vec{\sigma} = h_2 q_2 h_3 dq_3 \hat{q}_1$, así el lado izquierdo de la ecuación (3.50) en el límite diferencial se convierte en

$$\begin{aligned} \vec{\nabla} \times \vec{A} \cdot \lim_{S \rightarrow 0} \int_S d\vec{\sigma} &= dq_2 dq_3 h_2 h_3 \hat{q}_1 \cdot [\vec{\nabla} \times \vec{A}] , \\ &= dq_2 dq_3 h_2 h_3 [\vec{\nabla} \times \vec{A}]_1 . \end{aligned} \quad (3.51)$$

La integral de línea en el lado derecho de la ecuación (3.50) naturalmente divide en cuatro partes a lo largo de C_a , C_b , C_c y C_d , como se muestra en la figura 3.11. La integral completa es entonces dada por

$$\oint_C d\vec{r} \cdot \vec{A} = \int_{C_a} dq_2 h_2 A_2 + \int_{C_b} dq_3 h_3 A_3 + \int_{C_c} dq_2 h_2 A_2 + \int_{C_d} dq_3 h_3 A_3 . \quad (3.52)$$

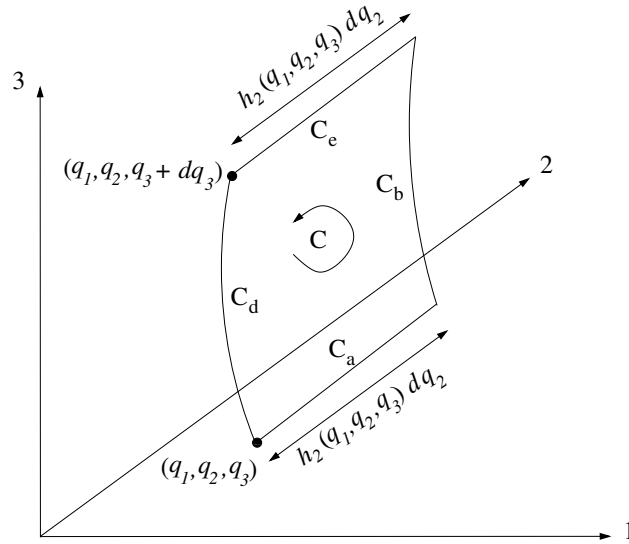


Figura 3.11: Geometría diferencial para integración curvilínea del rotor

En el límite diferencial, la integral a través de C_a vale

$$\int_{C_a} dq_2 h_2 A_2 = (h_2 A_2)|_{(q_1, q_2, q_3)} dq_2 , \quad (3.53)$$

donde evaluamos A_2 y h_2 en el punto (q_1, q_2, q_3) . Igualmente, la contribución a lo largo de C_c es

$$\int_{C_c} dq_2 h_2 A_2 = -(h_2 A_2)|_{(q_1, q_2, q_3 + dq_3)} dq_2 \quad (3.54)$$

donde ahora evaluamos A_2 y h_2 en $(q_1, q_2, q_3 + dq_3)$. En el límite diferencial,

$$(h_2 A_2)|_{(q_1, q_2, q_3 + dq_3)} = (h_2 A_2)|_{(q_1, q_2, q_3)} + \frac{\partial (h_2 A_2)}{\partial q_3} \Big|_{(q_1, q_2, q_3)} dq_3 , \quad (3.55)$$

lo cual permite que las integrales a lo largo de C_b y C_d se combinen para dar

$$\int_{C_a+C_c} d\vec{r} \cdot \vec{A} = - \left. \frac{\partial(h_2 A_2)}{\partial q_3} \right|_{(q_1, q_2, q_3)} dq_2 dq_3 . \quad (3.56)$$

Integraciones similares pueden ser desarrolladas a lo largo de C_b y C_d . La combinación de las cuatro partes lleva a

$$\lim_{C \rightarrow 0} \oint_C \vec{A} \cdot d\vec{r} = \left[\frac{\partial(h_3 A_3)}{\partial q_2} - \frac{\partial(h_2 A_2)}{\partial q_3} \right] dq_2 dq_3 . \quad (3.57)$$

Sustituyendo las ecuaciones (3.57) y (3.51) en la ecuación (3.50) tenemos la 1-componente del rotor de \vec{A} :

$$\left[\vec{\nabla} \times \vec{A} \right]_1 = \frac{1}{h_2 h_3} \left[\frac{\partial(h_3 A_3)}{\partial q_2} - \frac{\partial(h_2 A_2)}{\partial q_3} \right] . \quad (3.58)$$

Las otras componentes de $\vec{\nabla} \times \vec{A}$ pueden ser obtenidas reorientando la superficie mostrada en la figura 3.10. Los resultados son

$$\left[\vec{\nabla} \times \vec{A} \right]_2 = \frac{1}{h_1 h_3} \left[\frac{\partial(h_1 A_1)}{\partial q_3} - \frac{\partial(h_3 A_3)}{\partial q_1} \right] \quad (3.59)$$

$$\left[\vec{\nabla} \times \vec{A} \right]_3 = \frac{1}{h_1 h_2} \left[\frac{\partial(h_2 A_2)}{\partial q_1} - \frac{\partial(h_1 A_1)}{\partial q_2} \right] , \quad (3.60)$$

Las ecuaciones (3.58)-(3.60) pueden ser usadas más compactamente usando un determinante,

$$\vec{\nabla} \times \vec{A} = \frac{1}{h_1 h_2 h_3} \begin{vmatrix} h_1 \hat{q}_1 & h_2 \hat{q}_2 & h_3 \hat{q}_3 \\ \partial/\partial q_1 & \partial/\partial q_2 & \partial/\partial q_3 \\ h_1 A_1 & h_2 A_2 & h_3 A_3 \end{vmatrix} , \quad (3.61)$$

o usando los símbolos de Levi-Civita y la notación de Einstein,

$$\vec{\nabla} \times \vec{A} = \frac{\epsilon_{ijk}}{h_j h_k} \frac{\partial(h_k A_k)}{\partial q_j} \hat{q}_i . \quad (3.62)$$

3.5. Gradiente, divergencia y rotor en sistemas cilíndricos y esféricos

3.5.1. Operaciones cilíndricas

En el sistema cilíndrico $h_1 \equiv h_\rho = 1$, $h_2 \equiv h_\phi = \rho$ y $h_3 \equiv h_z = 1$. El gradiente, la divergencia y el rotor se convierten en

$$\vec{\nabla} \Phi = \frac{\partial \Phi}{\partial \rho} \hat{q}_\rho + \frac{1}{\rho} \frac{\partial \Phi}{\partial \phi} \hat{q}_\phi + \frac{\partial \Phi}{\partial z} \hat{q}_z \quad (3.63)$$

$$\vec{\nabla} \cdot \vec{A} = \frac{1}{\rho} \frac{\partial(\rho A_\rho)}{\partial \rho} + \frac{1}{\rho} \frac{\partial A_\phi}{\partial \phi} + \frac{\partial A_z}{\partial z} \quad (3.64)$$

$$\vec{\nabla} \times \vec{A} = \left[\frac{1}{\rho} \frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z} \right] \hat{q}_\rho + \left[\frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho} \right] \hat{q}_\phi + \frac{1}{\rho} \left[\frac{\partial(\rho A_\phi)}{\partial \rho} - \frac{\partial A_\rho}{\partial \phi} \right] \hat{q}_z. \quad (3.65)$$

3.5.2. Operaciones esféricas

En el sistema esférico $h_1 \equiv h_r = 1$, $h_2 \equiv h_\theta = r$ y $h_3 \equiv h_\phi = r \operatorname{sen} \theta$. El gradiente, la divergencia y el rotor se convierten en

$$\vec{\nabla} \Phi = \frac{\partial \Phi}{\partial r} \hat{q}_r + \frac{1}{r} \frac{\partial \Phi}{\partial \theta} \hat{q}_\theta + \frac{1}{r \operatorname{sen} \theta} \frac{\partial \Phi}{\partial \phi} \hat{q}_\phi \quad (3.66)$$

$$\vec{\nabla} \cdot \vec{A} = \frac{1}{r^2} \frac{\partial(r^2 A_r)}{\partial r} + \frac{1}{r \operatorname{sen} \theta} \frac{\partial \operatorname{sen} \theta A_\theta}{\partial \theta} + \frac{1}{r \operatorname{sen} \theta} \frac{\partial A_\phi}{\partial \phi} \quad (3.67)$$

$$\begin{aligned} \vec{\nabla} \times \vec{A} = \frac{1}{r \operatorname{sen} \theta} \left[\frac{\partial(\operatorname{sen} \theta A_\theta)}{\partial \theta} - \frac{\partial A_\theta}{\partial \phi} \right] \hat{q}_r + \\ \frac{1}{r} \left[\frac{1}{\operatorname{sen} \theta} \frac{\partial A_r}{\partial \phi} - \frac{\partial(r A_\phi)}{\partial r} \right] \hat{q}_\theta + \frac{1}{r} \left[\frac{\partial(r A_\theta)}{\partial r} - \frac{\partial A_r}{\partial \theta} \right] \hat{q}_\phi. \end{aligned} \quad (3.68)$$

Capítulo 4

Introducción a tensores.

versión final 1.0-080415¹

Los tensores se encuentran en todas las ramas de la Física. En mecánica, el tensor de inercia es usado para describir la rotación de un cuerpos rígidos, y el tensor de stress-tensión describe la deformación de cuerpos rígidos. En electromagnetismo, el tensor de conductividad extiende la ley de Ohm para manejar flujos de corriente en un medio anisotrópico, y el tensor de stress de Maxwell es la forma más elegante para tratar las fuerzas electromagnéticas. El tensor de métrica de la mecánica relativista describe la extraña geometría del espacio y el tiempo.

Este capítulo presenta una introducción a tensores y sus manipulaciones, donde la forma de proceder es la siguiente: primero trataremos sólo con coordenadas cartesianas, para después generalizar a coordenadas curvilíneas. Sólo nos limitaremos a sistemas de coordenadas ortonormales. Al final de este capítulo, introduciremos unos objetos que se les suele llamar “pseudo”-objetos, los cuales surgirán de considerar las transformaciones entre sistemas de coordenadas que cumplen la ley de la mano derecha y de la izquierda.

4.1. El tensor de conductividad y la ley de Ohm.

La necesidad de tensores pueden ser fácilmente demostradas considerando la ley de Ohm. En una resistencia ideal, la ley de Ohm relaciona la corriente con el voltaje usando la expresión lineal

$$I = \frac{V}{R} . \quad (4.1)$$

En esta ecuación, I es la corriente que circula a través de la resistencia y V es el voltaje aplicado. Usando unidades MKS, I es medido en Amperes, V en Volts y R en Ohms.

La ecuación (4.1) describe el flujo de corriente a través de un elemento discreto. Para aplicar la ley de Ohm a un medio distribuido, como un sólido cristalino, una forma alternativa de esta ecuación debe ser utilizada

$$\vec{J} = \sigma \vec{E} . \quad (4.2)$$

¹Este capítulo está basado en el cuarto capítulo del libro: *Mathematical Physics* de Brusse Kusse & Erik Westwig, editorial JOHN WILEY & SONS, INC..

Aquí \vec{J} es la densidad de corriente, \vec{E} es el campo eléctrico y σ es la conductividad del material. En unidades MKS, \vec{J} es medido en Amperes por unidad de área, \vec{E} en Volts por metro y σ en Ohm-metros a la menos uno.

La ecuación (4.2) describe una simple dependencia física entre la densidad de corriente y el campo eléctrico, ya que la conductividad ha sido expresada como un escalar. Con una conductividad escalar, la cantidad de flujo de corriente es gobernado únicamente por las magnitudes σ y \vec{E} , mientras que la dirección del flujo es siempre paralela a \vec{E} . Pero en algunos materiales, esto no es así. Muchos sólidos cristalinos permiten que el flujo de corriente se desplace por una dirección más que por otra. Estos materiales anisotrópicos deben tener distintas conductividades en distintas direcciones. Además, estos cristales pueden inclusive presentar flujo de corriente de forma perpendicular a un campo eléctrico aplicado. Claramente la ecuación (4.2), con una conductividad escalar, no podrá manejar este tipo de situaciones.

Una solución es construir un arreglo de elementos de conductividad y expresar la ley de Ohm usando la notación matricial

$$\begin{pmatrix} J_1 \\ J_2 \\ J_3 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} . \quad (4.3)$$

En la ecuación (4.3), la densidad de corriente y el campo eléctrico son representados como vectores columna de un campo vectorial y la conductividad es ahora una matriz cuadrada. Esta ecuación puede ser escrita en una notación matricial más compacta

$$[J] = [\sigma][E] \quad (4.4)$$

o, en notación de Einstein

$$J_i = \sigma_{ij} E_j . \quad (4.5)$$

Todas estas expresiones producen el resultado deseado. Cualquier relación lineal entre \vec{J} y \vec{E} puede ser descrita. Por ejemplo, la componente 1 de la densidad de corriente está relacionada con la componente 1 del campo eléctrico por σ_{11} , mientras que la componente 2 de la densidad de corriente está relacionada con la componente 2 del campo eléctrico por σ_{22} . Los flujos perpendiculares están descritos por los elementos fuera de la diagonal. Por ejemplo, el elemento σ_{12} describe el flujo en la dirección 1 debido a un campo aplicado en la dirección 2.

Sin embargo, la representación matricial de la conductividad anisotrópica tiene un problema fundamental. Los elementos matriciales deben depender del sistema de coordenadas. Tal como sucede con las componentes de un vector, si reorientamos nuestro sistema de coordenadas, los valores específicos en el arreglo matricial deben cambiar. Lamentablemente, el arreglo matricial en sí no tiene la información sobre el sistema de coordenadas elegido. La manera de resolver este problema para las cantidades vectoriales fue incorporar los vectores base directamente en la notación. La misma aproximación puede ser usada para mejorar la notación para la conductividad anisotrópica. Definimos un nuevo objeto, llamado el tensor de conductividad, que notaremos $\vec{\sigma}$. Este objeto incluye tanto los elementos de matriz de la matriz de conductividad como la base de vectores en el sistema de coordenadas en cual estos elementos son válidos. Como esta notación está motivada en la notación vectorial, comenzaremos con una pequeña revisión de conceptos.

Recordemos que una cantidad vectorial, tal como el campo eléctrico, puede ser representado como un vector columna

$$\vec{E} \rightarrow \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} . \quad (4.6)$$

El vector y las cantidades matriciales no son iguales, ya que la matriz no puede reemplazar al vector \vec{E} en una ecuación vectorial y viceversa. Más aún, la base de vectores del sistema coordinado en la cual el vector está expresado debe ser incluida para formar una expresión equivalente

$$\vec{E} = E_i \hat{e}_i . \quad (4.7)$$

El tensor de conductividad anisotrópico $\vec{\sigma}$ puede ser tratado de manera similar. Puede ser representado por un arreglo matricial

$$\vec{\sigma} \rightarrow \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} , \quad (4.8)$$

pero el arreglo matricial y el tensor no son equivalentes, ya que el arreglo matricial no contiene la información sobre el sistema de coordenadas. Siguiendo el patrón usado para los vectores y la expresión para un vector dada en la ecuación (4.7), expresamos el tensor de conductividad como

$$\vec{\sigma} = \sigma_{ij} \hat{e}_i \hat{e}_j . \quad (4.9)$$

La discusión que sigue mostrará que esta es una notación muy poderosa. Soporta toda la manipulación algebraica que la notación de matrices y también podemos manejar con facilidad las transformaciones entre sistemas de coordenadas.

La expresión para el tensor de conductividad en el lado derecho de la ecuación (4.9) contiene los elementos de la matriz de conductividad y dos bases de vectores del sistema de coordenadas donde los elementos tienen validez. No hay operación entre estas bases de vectores. Ellos sirven como “cajones” en donde los elementos σ_{ij} son colocados. Hay una doble suma sobre los índices i e j , por tanto, para un sistema tridimensional, habrán 9 términos en esta suma, cada uno conteniendo dos de los vectores base. En otras palabras, podemos expandir la conductividad como

$$\vec{\sigma} = \sum_i \sum_j \sigma_{ij} \hat{e}_i \hat{e}_j = \sigma_{11} \hat{e}_1 \hat{e}_1 + \sigma_{12} \hat{e}_1 \hat{e}_2 + \sigma_{21} \hat{e}_2 \hat{e}_1 + \cdots . \quad (4.10)$$

Análogamente a cómo expandíamos un vector en términos de la base,

$$\vec{v} = \sum_i v_i \hat{e}_i = v_1 \hat{e}_1 + v_2 \hat{e}_2 + v_3 \hat{e}_3 . \quad (4.11)$$

Veamos como manejamos la ley de Ohm en esta nueva notación. Usando el tensor de conductividad, podemos escribir en un sistema de coordenadas independiente y usando notación “vector/tensor”

$$\vec{J} = \vec{\sigma} \cdot \vec{E} . \quad (4.12)$$

Notemos que usamos el producto punto entre el tensor de conductividad y el vector del campo eléctrico en el lado derecho de esta expresión. Podemos utilizar la notación de Einstein, y escribir

$$J_s \hat{e}_s = (\sigma_{jk} \hat{e}_j \hat{e}_k) \cdot (E_l \hat{e}_l) . \quad (4.13)$$

Por convención, el producto punto en la ecuación (4.13) opera entre la segunda base vectorial de $\vec{\sigma}$ y la base del vector \vec{E} . Podemos manipular la ecuación (4.13) como sigue

$$J_s \hat{e}_s = \sigma_{jk} E_l \hat{e}_j \hat{e}_k \cdot \hat{e}_l \quad (4.14)$$

$$J_s \hat{e}_s = \sigma_{jk} E_l \hat{e}_j \delta_{kl} \quad (4.15)$$

$$J_s \hat{e}_s = \sigma_{jk} E_k \hat{e}_j . \quad (4.16)$$

Las cantidades en la ecuación (4.16) son vectores. Las componentes i -ésimas de estos vectores pueden ser obtenidos aplicando producto punto con \hat{e}_i a ambos lados de la ecuación (4.16), obteniendo

$$J_i = \sigma_{ik} E_k , \quad (4.17)$$

lo cual es idéntico a las ecuaciones (4.3)-(4.5). Mantengamos en mente que hay una diferencia entre $\vec{\sigma} \cdot \vec{E}$ y $\vec{E} \cdot \vec{\sigma}$. El orden en los términos importan, ya que en general

$$\hat{e}_j \hat{e}_k \cdot \hat{e}_l \neq \hat{e}_l \cdot \hat{e}_j \hat{e}_k , \quad (4.18)$$

Las bases de vectores en esta notación cumplen variadas funciones

1. Establecen cajones para separar las componentes tensoriales.
2. Emparejan a las componentes con un sistema de coordenadas.
3. Establecen el formalismo para operaciones algebraicas entre tensores.
4. Como es mostrado en este capítulo, simplifican el formalismo para las transformaciones entre sistemas de coordenadas.

Ahora que hemos motivado nuestra investigación sobre los tensores con un ejemplo específico, procedemos a mirar algunas de sus propiedades formales.

4.2. Notación tensorial general y terminología.

El tensor de conductividad es un ejemplo específico de un tensor que usa dos bases vectoriales y cuyos elementos tienen dos subíndices. En general, un tensor puede tener una cantidad finita de subíndices, pero el número de subíndices deben ser siempre igual al número de vectores base. Por tanto, en general

$$\vec{T} = T_{ijk\dots} \hat{e}_i \hat{e}_j \hat{e}_k \dots \quad (4.19)$$

El número de vectores base determina el rango del tensor. Notemos como la notación tensorial es una generalización de la notación vectorial usada en los capítulos previos. Los vectores son simplemente tensores de rango uno. Los escalares pueden ser considerados como tensores de rango cero. Mantengamos en mente el rango del tensor con el número de vectores base en el lado derecho de la ecuación (4.19), mientras que la dimensión del sistema de coordenadas determina el número de valores diferentes que un índice en particular puede tomar. Para un sistema tridimensional, los índices (i, j, k , etc.) pueden tomar los valores (1,2,3) cada uno.

Esta notación introduce la posibilidad de una nueva operación entre los vectores, llamada el producto diadico. Este producto es escrito tanto como $\vec{A} : \vec{B}$ o simplemente $\vec{A}\vec{B}$. El producto diadico entre dos vectores crea un tensor de rango dos

$$\vec{A}\vec{B} = A_i \hat{e}_i B_j \hat{e}_j = A_i B_j \hat{e}_i \hat{e}_j \quad (4.20)$$

Este tipo de operación puede ser extendida para combinar dos tensores de un rango arbitrario. El resultado es un tensor con un rango igual a la suma de los rangos de los tensores involucrados en el producto. Usualmente esta operación es llamada un producto externo, lo cual es opuesto al producto punto, el cual es llamado producto interno.

4.3. Transformaciones entre sistemas de coordenadas.

La nueva notación tensorial de la ecuación (4.19) hace más fácil la tarea de transformar vectores entre distintos sistemas de coordenadas. De hecho, muchos textos definen formalmente un tensor como “un objeto que transforma como un tensor”. Esto parece no tener mucho sentido, como será visto en esta sección, pero es la definición correcta.

En este capítulo sólo las transformaciones entre sistemas ortonormales son considerados. Primero sólo veremos las transformaciones entre sistemas cartesianos, para luego generalizar estos resultados a sistemas curvilíneos.

4.3.1. Transformaciones vectoriales entre sistemas cartesianos.

Comenzaremos viendo las transformaciones de componentes entre dos sistemas cartesianos muy sencillos. Un sistema prima es rotado un ángulo θ_0 con respecto a un sistema sin primas, como es mostrado en la figura 4.1. Un vector \vec{v} puede ser expresado en componentes como

$$\vec{v} = v_i \hat{e}_i = v'_i \hat{e}'_i \quad (4.21)$$

De la geometría de la figura 4.1, podemos ver que las componentes vectoriales en el sistema primado están relacionadas con las componentes vectoriales del sistema no primado por las ecuaciones

$$\begin{aligned} v'_1 &= v_1 \cos \theta_0 + v_2 \sin \theta_0 \\ v'_2 &= -v_1 \sin \theta_0 + v_2 \cos \theta_0 \end{aligned} \quad (4.22)$$

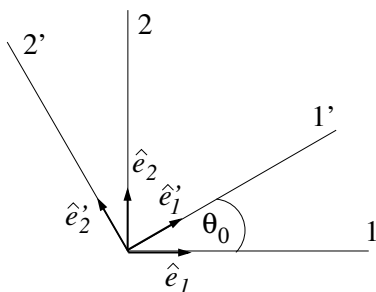


Figura 4.1: Sistemas rotados.

Estas ecuaciones pueden ser escritas en notación matricial

$$[v'] = [a][v] , \quad (4.23)$$

donde $[v']$ y $[v]$ son matrices columna que representan el vector \vec{v} con las componentes primas y no primas, y $[a]$ es la matriz cuadrada

$$[a] = \begin{pmatrix} \cos \theta_0 & \text{sen } \theta_0 \\ -\text{sen } \theta_0 & \cos \theta_0 \end{pmatrix} . \quad (4.24)$$

4.3.2. La matriz de transformación.

En general, cualquier transformación lineal de coordenadas de un vector puede ser escrita usando la notación de Einstein

$$v'_i = a_{ij}v_j , \quad (4.25)$$

donde $[a]$ es llamada la matriz de transformación. En la discusión que sigue, dos suposiciones simples son presentadas para determinar los elementos de $[a]$. La primera supone que los dos sistemas tienen vectores base conocidos. La segunda supone el conocimiento de las ecuaciones que relacionan las coordenadas. En este ejemplo utilizaremos el sistema de coordenadas cartesiano, sin embargo no es difícil generalizar a cualquier sistema de coordenadas.

Determinando $[a]$ desde la base de vectores.

Si la base de vectores de ambos sistemas coordenados son conocidos, es bastante simple determinar las componentes de $[a]$. Consideremos un vector \vec{v} expresado por componentes en

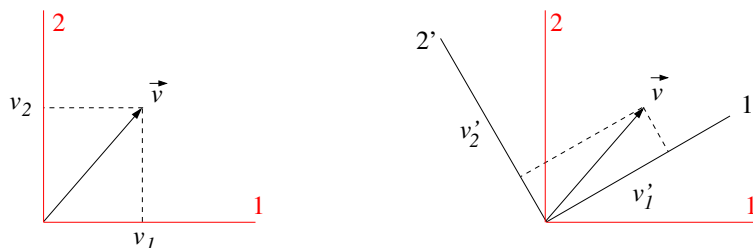


Figura 4.2: Componentes del vector.

dos sistemas cartesianos diferentes

$$\vec{v} = v_k \hat{e}_k = v'_i \hat{e}'_i . \quad (4.26)$$

Sustituyendo la expresión para v'_i dada en la ecuación (4.25) en (4.26), tenemos

$$v_k \hat{e}_k = a_{ij} v_j \hat{e}'_i . \quad (4.27)$$

Esto es verdad para cualquier \vec{v} . En particular, sea $\vec{v} = \hat{e}_m$ uno de los vectores base del sistema no primado (en otras palabras, $v_{k \neq m} = 0$ y $v_{k=m} = 1$), obtenemos

$$\hat{e}_m = a_{lm} \hat{e}'_l . \quad (4.28)$$

Aplicando producto punto por \hat{e}'_n en ambos lados, obtenemos

$$a_{nm} = (\hat{e}'_n \cdot \hat{e}_m) . \quad (4.29)$$

Notemos que los elementos de $[a]$ son sólo cosenos directores entre todos los pares de vectores base entre los sistemas primado y no primado.

Determinando $[a]$ desde las ecuaciones de coordenadas.

Si la base de vectores no es conocida explícitamente, las ecuaciones que relacionan los dos sistemas proveen el método más rápido para determinar la matriz de transformación. Comencemos considerando las expresiones para el vector desplazamiento en los dos sistemas. Como los sistemas son cartesianos,

$$d\vec{r} = dx_i \hat{e}_i = dx'_i \hat{e}'_i , \quad (4.30)$$

donde dx_i y dx'_i son los diferenciales totales de las coordenadas. Como la ecuación (4.25) representan las componentes de cualquier vector, incluido el vector de desplazamiento

$$dx'_i = a_{ij} dx_j . \quad (4.31)$$

La ecuación (4.31) provee un método general para obtener los elementos de matriz de $[a]$ usando las coordenadas primas y no primas. Trabajando en tres dimensiones, asumamos que estas ecuaciones son

$$\begin{aligned} x'_1 &= x'_1(x_1, x_2, x_3) \\ x'_2 &= x'_2(x_1, x_2, x_3) \\ x'_3 &= x'_3(x_1, x_2, x_3) , \end{aligned} \quad (4.32)$$

o en forma compacta

$$x'_i = x'_i(x_1, x_2, x_3) . \quad (4.33)$$

Expandiendo los diferenciales totales de la ecuación (4.32), tenemos

$$\begin{aligned}
dx'_1 &= \frac{\partial x'_1(x_1, x_2, x_3)}{\partial x_1} dx_1 + \frac{\partial x'_1(x_1, x_2, x_3)}{\partial x_2} dx_2 + \frac{\partial x'_1(x_1, x_2, x_3)}{\partial x_3} dx_3 \\
dx'_2 &= \frac{\partial x'_2(x_1, x_2, x_3)}{\partial x_1} dx_1 + \frac{\partial x'_2(x_1, x_2, x_3)}{\partial x_2} dx_2 + \frac{\partial x'_2(x_1, x_2, x_3)}{\partial x_3} dx_3 \\
dx'_3 &= \frac{\partial x'_3(x_1, x_2, x_3)}{\partial x_1} dx_1 + \frac{\partial x'_3(x_1, x_2, x_3)}{\partial x_2} dx_2 + \frac{\partial x'_3(x_1, x_2, x_3)}{\partial x_3} dx_3 .
\end{aligned}$$

Nuevamente, usando la notación de Einstein, podemos escribir lo anterior como

$$dx'_i = \frac{\partial x'_i(x_1, x_2, x_3)}{\partial x_j} dx_j . \quad (4.34)$$

Comparando las ecuaciones (4.31) y (4.34), podemos identificar los elementos de $[a]$

$$a_{ij} = \frac{\partial x'_i(x_1, x_2, x_3)}{\partial x_j} . \quad (4.35)$$

Propiedad ortonormal de $[a]$.

Si el sistema de coordenadas original y el primado son ambos ortonormales, podemos escribir una útil relación entre los elementos de $[a]$. Se puede derivar fácilmente aplicando producto punto con \hat{e}_k en la ecuación (4.28)

$$\begin{aligned}
\hat{e}_j &= a_{ij} \hat{e}'_i \\
\hat{e}_j \cdot \hat{e}_k &= a_{ij} (\hat{e}'_i \cdot \hat{e}_k) \\
\delta_{jk} &= a_{ij} a_{ik} .
\end{aligned} \quad (4.36)$$

La ecuación (4.36) escrita en forma matricial queda

$$[a][a]^\dagger = [1] , \quad (4.37)$$

donde $[a]^\dagger$ es la notación para la transpuesta conjugada de $[a]$, y la matriz $[1]$ es una matriz cuadrada, con 1 en la diagonal y 0 fuera de ella.

La inversa de $[a]$.

La matriz $[a]$ genera las componentes de los vectores en el sistema primado desde las componentes sin primas, como es indicado en la ecuación (4.25). Esta expresión puede ser invertida con la inversa de $[a]$, la cual es escrita como $[a]^{-1}$, y está definida por

$$[a][a]^{-1} = [a]^{-1}[a] = [1] , \quad (4.38)$$

o en notación de Einstein

$$a_{ij}^{-1} a_{jk} = a_{ij} a_{jk}^{-1} = \delta_{ik} . \quad (4.39)$$

Con la notación de Einstein, manejamos fácilmente la inversión

$$\begin{aligned}
 v'_i &= a_{ij}v_j \\
 a_{ki}^{-1}v'_i &= a_{ki}^{-1}a_{ij}v_j \\
 a_{ki}^{-1}v'_i &= \delta_{kj}v_j \\
 a_{ki}^{-1}v'_i &= v_k .
 \end{aligned} \tag{4.40}$$

Las matrices de transformación que obedecen la condición de ortonormalidad son simples de invertir. Comparando la ecuación (4.37) y (4.38) muestra que

$$[a]^{-1} = [a]^\dagger , \tag{4.41}$$

o en notación de Einstein

$$a_{ij}^{-1} = a_{ji} . \tag{4.42}$$

La relación de inversión se convierte en

$$v_i = a_{ji}v'_j . \tag{4.43}$$

Transformaciones de vectores base.

Los vectores base no primados fueron relacionados con la base del sistema primado por la ecuación (4.28)

$$\hat{e}_i = a_{ji}\hat{e}'_j . \tag{4.44}$$

Usando el hecho que la inversa de la matriz $[a]$ es su transpuesta, esta expresión puede ser invertida para obtener la base de vectores primada en términos de los no primados

$$\hat{e}'_j = a_{ij}\hat{e}_i . \tag{4.45}$$

Recordemos que estas expresiones son sólo válidas para transformaciones si ambos sistemas son ortonormales.

4.3.3. Resumen de transformaciones de coordenadas.

El siguiente cuadro resume las ecuaciones de las transformaciones entre dos sistemas cartesianos

$ \begin{aligned} v'_i &= a_{ij}v_j & \hat{e}'_i &= a_{ij}\hat{e}_j \\ v_i &= a_{ji}v'_j & \hat{e}_i &= a_{ji}\hat{e}'_j \\ a_{ij} &= (\hat{e}'_i \cdot \hat{e}_j) & &= \partial x'_i(x_1, x_2, x_3)/\partial x_j \end{aligned} $

Las funciones $x'_i = x'_i(x_1, x_2, x_3)$ relacionan el sistema de coordenadas cartesiano primado con el sistema cartesiano no primado. Para mantener las cosas ordenadas, notemos que hay un patrón para estas ecuaciones de transformación. Cada vez que convertimos del sistema no

primado con el sistema primado, estamos tratando con una base vectorial o las componentes de algún vector, sumamos sobre el segundo índice a_{ij} . Por el contrario, las conversiones desde el sistema primado al sistema no primado siempre se sumará sobre el primer índice.

4.3.4. Transformaciones tensoriales.

Para entender por qué los elementos de un tensor deben cambiar de valor cuando son expresados en distintos sistemas de coordenadas, consideremos el tensor de conductividad. Si fijamos el set de coordenadas y la corriente fluye más fácilmente en la dirección 1 que en la dirección 2, entonces $\sigma_{11} > \sigma_{22}$. Si observamos la misma situación física en un nuevo sistema de coordenadas donde la dirección 1' es equivalente a la dirección 2 y la dirección 2' es la misma que la dirección 1 original, entonces deberíamos tener que $\sigma'_{11} < \sigma'_{22}$. Claramente los elementos del tensor de conductividad deben tomar diferentes valores en los dos sistemas, aún cuando describen la misma situación Física. Esto es cierto también para una cantidad vectorial, el mismo vector velocidad tienen diferentes componentes en diferentes sistemas de coordenadas.

Las transformaciones tensoriales siguen el mismo patrón que las transformaciones vectoriales. Un vector expresado en un sistema primado y no primado seguirá siendo el mismo vector,

$$\vec{v} = v_i \hat{e}_i = v'_j \hat{e}'_j . \quad (4.46)$$

De la misma forma, siguiendo la notación de la ecuación (4.19), las expresiones para un tensor de segundo rango en los dos sistemas deben obedecer

$$\vec{T} = T_{ij} \hat{e}_i \hat{e}_j = T'_{rs} \hat{e}'_r \hat{e}'_s . \quad (4.47)$$

Aquí yace la belleza de la notación. La relación entre los elementos T_{ij} y T'_{rs} es construida desde la ecuación (4.47) y es fácilmente obtenida aplicando dos veces producto punto en ambos lados. Del primer producto punto obtenemos

$$\begin{aligned} \hat{e}_l \cdot T_{ij} \hat{e}_i \hat{e}_j &= \hat{e}_l \cdot T'_{rs} \hat{e}'_r \hat{e}'_s \\ T_{ij} (\hat{e}_l \cdot \hat{e}_i) \hat{e}_j &= T'_{rs} (\hat{e}_l \cdot \hat{e}'_r) \hat{e}'_s \\ T_{ij} \delta_{li} \hat{e}_j &= T'_{rs} a_{rl} \hat{e}'_s \\ T_{lj} \hat{e}_j &= T'_{rs} a_{rl} \hat{e}'_s . \end{aligned} \quad (4.48)$$

Aplicando un segundo producto punto y realizando el proceso análogo obtenemos

$$T_{lm} = T'_{rs} a_{rl} a_{ms} . \quad (4.49)$$

Para invertir la ecuación (4.49) usamos la matriz inversa $[a]^{-1}$ dos veces, y recordando que para sistemas de coordenadas ortonormales se cumple que $a_{ij}^{-1} = a_{ji}$, obtenemos

$$T'_{lm} = T_{rs} a_{lr} a_{ms} . \quad (4.50)$$

En general, las transformaciones tensoriales requieren un factor a_{ij} para cada subíndice en el tensor. En otras palabras, un tensor de rango r necesita r diferentes factores a_{ij} . Si

la transformación va desde el sistema sin prima al sistema prima, todos los factores a_{ij} son sumadas sobre el segundo subíndice. Para la transformación inversa, desde el sistema primado al sistema no primado, las sumas son sobre el primer subíndice. Las transformaciones tensoriales, para tensores de rango arbitrario, pueden ser resumidas como siguen

$$\begin{aligned} T'_{ijk\dots} &= T_{rst\dots} a_{ir} a_{js} a_{kt} \dots \\ T_{ijk\dots} &= T'_{rst\dots} a_{ri} a_{sj} a_{tk} \dots \end{aligned}$$

donde los elementos de la matriz $[a]$ están dados por la ecuación (4.35).

Hay otro asunto importante en la notación tensorial de la ecuación (4.19). Al contrario de la ecuación matricial, donde todos los términos deben estar en la misma base, la notación tensorial/vectorial permite que las ecuaciones estén en bases distintas. Imaginemos que los elementos de la ecuación de Ohm expresados en los sistemas primados y no primados sean los siguientes

$$\begin{aligned} \vec{J} &= J_i \hat{e}_i = J'_i \hat{e}'_i \\ \vec{E} &= E_i \hat{e}_i = E'_i \hat{e}'_i \\ \vec{\sigma} &= \sigma_{ij} \hat{e}_i \hat{e}_j = \sigma'_{ij} \hat{e}'_i \hat{e}'_j . \end{aligned} \tag{4.51}$$

La ley de Ohm queda

$$\vec{J} = \vec{\sigma} \cdot \vec{E} , \tag{4.52}$$

y cualquier combinación de las representaciones de la ecuación (4.51) pueden ser usados en la evaluación. Por ejemplo,

$$J_i \hat{e}_i = (\sigma'_{jk} \hat{e}'_j \hat{e}'_k) \cdot (E_l \hat{e}_l) = \sigma'_{jk} E_l \hat{e}'_j (\hat{e}'_k \cdot \hat{e}_l) = \sigma'_{jk} E_l \hat{e}'_j a_{kl} . \tag{4.53}$$

El hecho que los elementos de $\vec{\sigma}$ del sistema primado sean combinados con las componentes de \vec{E} del sistema no primado no representa un problema. El producto punto de las bases de los vectores toma en cuenta las representaciones mezcladas, siempre y cuando el orden de las bases de los vectores sea preservado. Esto es acompañado en la ecuación (4.53) por el hecho que $\hat{e}'_k \cdot \hat{e}_l \neq \delta_{kl}$. Este tipo de operación no puede ser hecho con la notación matricial sin antes convertir todo a una misma base.

Con esto debería quedar claro el valor de expresar un tensor de la forma como se ve en (4.19). Además de poder manejar las manipulaciones algebraicas como una matriz, también contiene toda la información necesaria para transformar los elementos de un sistema de coordenadas al otro. Por tanto, un tensor es de coordenadas independientes, y un objeto geométrico, tal como un lo vector es.

4.4. Diagonalización de tensores.

En problemas de Física a menudo necesitamos diagonalizar un tensor. Esto significa que necesitamos encontrar un sistema de coordenadas particular en el cual la representación matricial de un tensor sólo tenga elementos distintos de cero en su diagonal. Por ejemplo, un

cuerpo rígido no experimentará vibraciones cuando es rotado alrededor de cualquiera de tres ejes en un sistema de ejes donde el tensor de inercia sea diagonal. El proceso de balancear una rueda de un automóvil usa este hecho. Y cuando los ejes no coinciden con los ejes requeridos, se colocan pequeños trozos de metal en la llanta para que esto sí suceda.

Muchos estudiantes se pierden en el proceso matemático de la diagonalización y se olvidan que, en realidad, es sólo una transformación de coordenadas. En esta sección, derivamos los elementos de la matriz de transformación $[a]$ que diagonaliza un tensor dado. Comenzaremos con un tratamiento absolutamente teórico del tema. Luego veremos dos ejemplos numéricos, uno no degenerado y otro degenerado.

4.4.1. Diagonalización y problema de valores propios.

Basado en la discusión de la sección anterior, un tensor $\vec{\sigma}$ escrito en un sistema no primado debe ser equivalente a uno escrito en un sistema primado

$$\vec{\sigma} = \sigma_{ij} \hat{e}_i \hat{e}_j = \sigma'_{st} \hat{e}'_s \hat{e}'_t . \quad (4.54)$$

Estamos interesados en un sistema prima muy especial, un sistema en el cual todos los elementos no diagonales de $\vec{\sigma}$ son cero. En este caso, la ecuación (4.54) queda

$$\vec{\sigma} = \sigma_{ij} \hat{e}_i \hat{e}_j = \sigma'_{ss} \hat{e}'_s \hat{e}'_s . \quad (4.55)$$

En esta última ecuación suponemos conocidos los elementos tensoriales y la base vectorial del sistema no prima. El problema es encontrar los elementos del tensor en el sistema primado σ'_{ss} y los elementos de la base \hat{e}'_s , de tal manera que se satisfaga la ecuación (4.55). Para realizar esto, aplicamos producto punto a la ecuación (4.55) con el primer elemento de la base del sistema primado, \hat{e}'_1 , con lo cual obtenemos

$$\begin{aligned} \vec{\sigma} \cdot \hat{e}'_1 &= \sigma'_{ss} \hat{e}'_s \hat{e}'_s \cdot \hat{e}'_1 \\ &= \sigma'_{ss} \hat{e}'_s \delta_{s1} \\ &= \sigma'_{11} \hat{e}'_1 . \end{aligned} \quad (4.56)$$

La ecuación (4.56) revela una propiedad importante de la base de vectores donde el tensor es diagonal. No cambian de dirección cuando es aplicado el producto punto por el tensor. Sin embargo, ellos pueden cambiar de magnitud. Si definimos $\lambda_1 = \sigma'_{11}$, la ecuación (4.56) queda

$$\vec{\sigma} \cdot \hat{e}'_1 = \lambda_1 \hat{e}'_1 . \quad (4.57)$$

El factor λ_1 es llamado el autovalor de $\vec{\sigma}$. Un autovalor aparece cuando una operación sobre un objeto produce una constante, el autovalor, por el objeto original. El vector base del sistema primado es llamado un autovector.

Ahora introducimos el tensor unitario $\vec{\mathbb{1}}$, el cual es definido por

$$\vec{\mathbb{1}} = \delta_{ij} \hat{e}_i \hat{e}_j \quad (4.58)$$

que cumple

$$\vec{1} \cdot \vec{v} = \vec{v} . \quad (4.59)$$

Representado como matriz, el tensor unitario es

$$\vec{1} \rightarrow [1] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \quad (4.60)$$

Usando el tensor unitario, la ecuación (4.57) puede ser escrita como

$$\left(\vec{\sigma} - \lambda_1 \vec{1} \right) \cdot \hat{e}'_1 = 0 . \quad (4.61)$$

Expresando $\vec{\sigma}$ en el sistema no primado, la ecuación (4.61) puede ser reescrita en notación de Einstein

$$(\sigma_{ij} - \lambda_1 \delta_{ij}) \hat{e}_i \hat{e}_j \cdot \hat{e}'_1 = 0 . \quad (4.62)$$

Usando la ecuación (4.29) y alguna manipulación algebraica, obtenemos

$$\hat{e}_i (\sigma_{ij} - \lambda_1 \delta_{ij}) a_{1j} = 0 , \quad (4.63)$$

donde el elemento a_{1j} es uno de los tres elementos desconocidos de la matriz transformación entre el sistema original de coordenadas y el sistema donde $\vec{\sigma}$ es diagonal.

El lado izquierdo de la ecuación (4.63) es un vector, y para que sea cero, cada componente debe ser cero. Cada componente involucra una suma sobre el índice j . Por tanto, la ecuación (4.63) se convierte en tres ecuaciones, las cuales pueden ser anotadas en notación matricial

$$\begin{pmatrix} \sigma_{11} - \lambda_1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} - \lambda_1 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} - \lambda_1 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} . \quad (4.64)$$

Para que este set de ecuaciones lineales y homogéneas tengan solución, el determinante de los coeficientes debe ser cero

$$\begin{vmatrix} \sigma_{11} - \lambda_1 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} - \lambda_1 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} - \lambda_1 \end{vmatrix} = 0 . \quad (4.65)$$

Resulta una ecuación de tercer orden para λ_1 , las cuales generarán tres autovalores. De estos tres autovalores, seleccionaremos uno, el cual será llamado λ_1 , y los otros dos los usaremos luego. Reemplazando este valor en la ecuación (4.64) encontraremos una solución para a_{11} , a_{12} y a_{13} con una constante arbitraria. Estos son tres elementos de la matriz de transformación entre los sistemas primados y no primados, lo cual estamos buscando. Estos tres elementos también permitirán determinar la base vectorial \hat{e}'_1 con una constante arbitraria

$$\hat{e}'_1 = a_{1j} \hat{e}_j . \quad (4.66)$$

Imponiendo que \hat{e}'_1 sea un vector unitario, obtenemos una condición para las constantes arbitrarias asociadas con a_{11} , a_{12} y a_{13}

$$(a_{11})^2 + (a_{12})^2 + (a_{13})^2 = 1 . \quad (4.67)$$

Exceptuando un signo arbitrario global y la situación degenerada, la cual discutiremos luego, hemos determinado en forma única \hat{e}'_1 .

En forma análoga encontramos los otros elementos de la base y los elementos de la matriz de transformación. El segundo vector base del sistema primado es determinado aplicando el producto punto en la ecuación (4.56) y usando \hat{e}'_2 . Podemos escribir ecuaciones matriciales análogas a (4.64) para a_{21} , a_{22} y a_{23} . Las ecuaciones (4.65) que escribimos mediante determinante resultan idénticas para λ_2 . Seleccionamos uno de los dos autovalores restantes, y lo llamamos λ_2 , el cual usamos para determinar a_{21} , a_{22} , a_{23} y \hat{e}'_2 . Análogamente, obtenemos los elementos a_{31} , a_{32} , a_{33} y \hat{e}'_3 .

El sistema de coordenadas primado, donde $\vec{\sigma}$ es diagonal, es definido por la base vectorial \hat{e}'_1 , \hat{e}'_2 y \hat{e}'_3 . Los elementos de $\vec{\sigma}$ en este sistema son los autovalores que determinamos desde la ecuación (4.65)

$$[\sigma'] = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} . \quad (4.68)$$

Las matrices de interés en Física son Hermitianas. Si dejamos la posibilidad de elementos de matriz complejos, una matriz se dice Hermitiana si es igual a su transpuesta conjugada. Esto es, $\sigma_{ij}^\dagger = \sigma_{ij}^*$. Hay dos propiedades muy importantes en este tipo de matrices. Uno, los valores propios son números reales. Segundo, sus autovectores son siempre ortonormales. La prueba de este hecho es dejado como ejercicio.

La única complicación que puede surgir en el proceso de diagonalización es una situación degenerada, la cual ocurre cuando dos o más autovalores son idénticos. Consideremos el caso cuando $\lambda_1 \neq \lambda_2 = \lambda_3$. El autovalor λ_1 determina a_{11} , a_{12} , a_{13} y \hat{e}'_1 , como ya lo vimos. Sin embargo, los autovalores degenerados no especificarán en forma única sus autovectores. Estos autovectores pueden ser elegidos de manera infinita. Un ejemplo con este tipo de degeneración es discutido en uno de los ejemplos que a continuación siguen.

Ejemplo 1

Como un ejemplo del proceso de diagonalización, consideremos el tensor de conductividad expresado en coordenadas cartesianas

$$\vec{\sigma} = \sigma_{ij} \hat{e}_i \hat{e}_j . \quad (4.69)$$

Sea su representación matricial (ignorando las unidades)

$$[\sigma] = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 10 & 1 \\ 0 & 1 & 10 \end{pmatrix} . \quad (4.70)$$

Esta matriz es Hermitiana, por tanto podemos esperar que sus valores propios sean reales y sus autovectores ortonormales. Los autovalores para la diagonalización son generados desde la ecuación determinante

$$\begin{vmatrix} 10 - \lambda & 0 & 0 \\ 0 & 10 - \lambda & 1 \\ 0 & 1 & 10 - \lambda \end{vmatrix} = 0 . \quad (4.71)$$

La expansión del determinante nos arroja una ecuación polinomial de tercer orden

$$(10 - \lambda) [(10 - \lambda)^2 - 1] = 0 , \quad (4.72)$$

la cual tiene tres soluciones, $\lambda_1 = 9$, $\lambda_2 = 11$ y $\lambda_3 = 10$.

Los elementos a_{1j} son determinados reemplazando el valor de λ_1 en la ecuación (4.64), obtenemos

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} . \quad (4.73)$$

Esta ecuación requiere que se cumpla $a_{12} = -a_{13}$ y $a_{11} = 0$. La condición de normalización impone el contreñimiento adicional $(a_{12})^2 + (a_{13})^2 = 1$, de donde obtenemos

$$\begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} . \quad (4.74)$$

El primer autovector asociado con el sistema primado es

$$\hat{e}'_1 = \left(1/\sqrt{2}\right) \hat{e}_2 - \left(1/\sqrt{2}\right) \hat{e}_3 . \quad (4.75)$$

Las otras componentes de $[a]$ pueden ser determinadas en forma análoga. La matriz de transformación completa es

$$[a] = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \\ \sqrt{2} & 0 & 0 \end{pmatrix} . \quad (4.76)$$

Los otros dos autovectores no primados son

$$\hat{e}'_2 = \left(1/\sqrt{2}\right) \hat{e}_2 + \left(1/\sqrt{2}\right) \hat{e}_3 \quad (4.77)$$

y

$$\hat{e}'_3 = \hat{e}_1 . \quad (4.78)$$

Podemos notar que hay una ambigüedad de orden con los autovalores y en los signos asociados con cada autovector. Estas ambigüedades nos permiten fijar el sistema primado como de mano derecha. El orden y las elecciones de signo hechas en este ejemplo dan la base primada que se muestra en la figura 4.3.

Los elementos del tensor de conductividad expresados en el nuevo sistema diagonal son

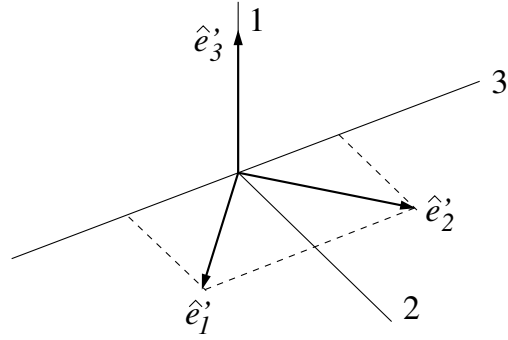


Figura 4.3: Vectores base en el sistema primado.

$$[\sigma'] = \begin{pmatrix} 9 & 0 & 0 \\ 0 & 11 & 0 \\ 0 & 0 & 10 \end{pmatrix} \quad (4.79)$$

Ejemplo 2

Este ejemplo demuestra el proceso de diagonalización cuando dos autovectores son degenerados. Consideremos nuevamente un tensor de conductividad en coordenadas cartesianas (nuevamente, ignoramos las unidades)

$$[\sigma] = \begin{pmatrix} 11 & -1 & 0 \\ -1 & 11 & 0 \\ 0 & 0 & 10 \end{pmatrix}. \quad (4.80)$$

Esta es una matriz Hermitiana, por tanto esperamos valores propios reales y vectores ortonormales. La condición del determinante queda

$$\begin{vmatrix} 11 - \lambda & -1 & 0 \\ -1 & 11 - \lambda & 0 \\ 0 & 0 & 10 - \lambda \end{vmatrix} = 0, \quad (4.81)$$

lo cual lleva a una ecuación polinomial de tercer orden

$$(10 - \lambda) [(11 - \lambda)^2 - 1]. \quad (4.82)$$

Esta ecuación de tercer orden posee tres raíces, pero sólo dos distintas, $\lambda_1 = 12$ y $\lambda_2 = \lambda_3 = 10$. La raíz λ_1 puede ser tratada como antes. Cuando es sustituida en la ecuación (4.64), la relación matricial se convierte en

$$\begin{pmatrix} -1 & -1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (4.83)$$

Cuando utilizamos esto más la condición de normalización, obtenemos

$$\begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}. \quad (4.84)$$

Estos elementos de la matriz transformación nos permiten definir el primer autovector

$$\hat{e}'_1 = \left(1/\sqrt{2}\right) \hat{e}_1 - \left(1/\sqrt{2}\right) \hat{e}_2. \quad (4.85)$$

Ahora consideremos los valores propios degenerados. Cuando sustituímos $\lambda_2 = 10$ en la ecuación (4.64), obtenemos

$$\begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_{21} \\ a_{22} \\ a_{23} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (4.86)$$

Si sustituímos λ_3 obtenemos una ecuación muy parecida

$$\begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_{31} \\ a_{32} \\ a_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (4.87)$$

La ecuación (4.86) nos requiere $a_{21} = a_{22}$, pero deja libre el factor a_{23} . La condición de normalización nos exige que se cumpla $a_{21}^2 + a_{22}^2 + a_{23}^2 = 1$. Estas condiciones pueden ser satisfechas por muchos autovectores. Como a_{23} es arbitrario, lo fijamos igual a cero. Ahora, si el segundo autovector debe ser ortonormal a \hat{e}'_1 , tenemos

$$\begin{pmatrix} a_{21} \\ a_{22} \\ a_{23} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}. \quad (4.88)$$

Con esto, escribimos el segundo autovector

$$\hat{e}'_2 = \left(1/\sqrt{2}\right) \hat{e}_1 + \left(1/\sqrt{2}\right) \hat{e}_2. \quad (4.89)$$

El autovector asociado con λ_3 está dado por la ecuación (4.87) y tiene las mismas condiciones que el autovector asociado a λ_2 , es decir, $a_{31} = a_{32}$ y a_{33} es arbitrario. Sin embargo, si queremos que los autovectores sean ortonormales, \hat{e}'_3 debe ser perpendicular a \hat{e}'_1 y \hat{e}'_2 . Los vectores base \hat{e}'_1 y \hat{e}'_2 están en el plano 1-2 de los vectores originales, por tanto si queremos que \hat{e}'_3 perpendicular a estos dos vectores, debe estar en la dirección 3. Por tanto,

$$\begin{pmatrix} a_{31} \\ a_{32} \\ a_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (4.90)$$

y para el tercer autovector

$$\hat{e}'_3 = \hat{e}_3. \quad (4.91)$$

Un chequeo rápido demostrará que estos tres autovectores son ortonormales y que definen un sistema derecho de coordenadas, en el cual los elementos del tensor de conductividad están diagonalizados.

4.5. Transformaciones tensoriales en sistemas de coordenadas curvilíneas.

Las transformaciones de las secciones previas pueden ser fácilmente generalizadas a un sistema de coordenadas curvilíneas. Consideremos el problema intermedio de una transformación entre un sistema cartesiano y uno curvilíneo.

El sistema cartesiano tendrá las coordenadas primadas (x'_1, x'_2, x'_3) y los vectores base $(\hat{e}'_1, \hat{e}'_2, \hat{e}'_3)$. Por otra parte, el sistema curvilíneo tiene las coordenadas (q_1, q_2, q_3) , los vectores base $(\hat{q}_1, \hat{q}_2, \hat{q}_3)$ y los factores de escala (h_1, h_2, h_3) . El set de ecuaciones que relacionan las coordenadas de los dos sistemas pueden ser escritas por

$$\begin{aligned}x'_1 &= x'_1(q_1, q_2, q_3) \\x'_2 &= x'_2(q_1, q_2, q_3) \\x'_3 &= x'_3(q_1, q_2, q_3) .\end{aligned}\tag{4.92}$$

Por ejemplo, las ecuaciones que relacionan el sistema de coordenadas cilíndrico con el cartesiano son

$$\begin{aligned}x'_1 &= x' = \rho \cos \theta \\x'_2 &= y' = \rho \operatorname{sen} \theta \\x'_3 &= z' = z .\end{aligned}\tag{4.93}$$

La matriz de transformación $[a]$ realiza la misma función como antes. Esto es, toma las componentes del sistema curvilíneo no primado de un vector y genera las coordenadas cartesianas en el sistema primado

$$v'_i = a_{ij}v_j .\tag{4.94}$$

Recordando del capítulo anterior que el vector desplazamiento para los dos sistemas puede ser escrito como

$$d\vec{r} = dx'_i \hat{e}'_i = h_j dq_j \hat{q}_j .\tag{4.95}$$

Las componentes del vector desplazamiento en el sistema curvilíneo no primado están dados por $h_j dq_j$, mientras que sus componentes en el sistema cartesiano primado están dados por dx'_i . Estas componentes deben estar relacionadas por la matriz transformación $[a]$. En notación de Einstein

$$dx'_i = a_{ij} h_j dq_j .\tag{4.96}$$

El diferencial total dx'_i puede ser formado desde la ecuación (4.92), de donde obtenemos

$$dx'_i = \frac{\partial x'_i(q_1, q_2, q_3)}{\partial q_j} dq_j .\tag{4.97}$$

La ecuación (4.97) puede ser colocada en la forma de la ecuación (4.96) multiplicando el lado derecho de la ecuación (4.97) por h_j/h_j

$$dx'_i = \frac{\partial x'_i(q_1, q_2, q_3)}{\partial q_j} \frac{h_j}{h_j} dq_j = \frac{\partial x'_i(q_1, q_2, q_3)}{h_j \partial q_j} h_j dq_j . \quad (4.98)$$

Comparando las ecuaciones (4.98) y (4.96) obtenemos

$$a_{ij} = \frac{\partial x'_i(q_1, q_2, q_3)}{h_j \partial q_j} \quad [\text{Curvilíneo a Cartesiano}] . \quad (4.99)$$

La generalización para la transformación entre dos sistemas curvilíneos se sigue de una manera análoga. Los elementos para la matriz transformación $[a]$ en este caso son

$$a_{ij} = \frac{h'_j \partial x'_i(q_1, q_2, q_3)}{h_j \partial q_j} \quad [\text{Curvilíneo a Curvilíneo}] . \quad (4.100)$$

Notemos que no hay suma sobre i ó j en el lado derecho de la ecuación (4.100) ya que ambos subíndices aparecen en el lado izquierdo de la expresión.

La ecuación (4.100) es la forma más general para los elementos de la matriz transformación entre dos sistemas curvilíneos. Es simplificada a la ecuación (4.99) si el sistema primado es cartesiano, ya que para este caso $h'_j \rightarrow 1$. Además se degenera a la ecuación (4.35) cuando los dos sistemas son cartesianos, ya que para este caso $h_j \rightarrow 1$.

Como antes, la matriz de transformación puede ser determinada también desde la base vectorial de los dos sistemas de coordenadas. Para el caso curvilíneo general, los elementos de $[a]$ son

$$a_{ij} = (\hat{q}'_i \cdot \hat{q}_j) . \quad (4.101)$$

La manipulación algebraica es fácil utilizando la notación de Einstein. Puede ser un ejercicio útil realizar los mismos pasos usando sólo matrices para que se convenza que es más útil.

4.6. Pseudo-objetos.

Si consideramos sólo las transformaciones que involucran traslaciones o rotaciones rígidas, no hay forma de cambiar un sistema de orientación derecha a uno de orientación izquierda, o viceversa. Para cambiar esto necesitamos una reflexión. Las transformaciones que involucran reflexiones requieren la introducción de los llamados “pseudo”-objetos. Los pseudoescalares, pseudovectores y pseudotensores son muy similares a sus contrapartes “regulares”, excepto por su comportamiento cuando son reflejados. Una forma fácil de demostrar la diferencia es examinando el producto cruz de dos vectores regulares en los sistemas derechos e izquierdos.

4.6.1. Pseudo-vectores.

Consideremos el sistema de coordenadas cartesiana derecho mostrado en la figura 4.4. La figura muestra dos vectores regulares en este sistema, orientado a lo largo de dos vectores de la base

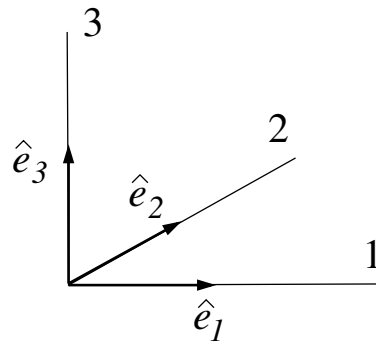


Figura 4.4: Sistema de la mano derecha.

$$\vec{A} = A_0 \hat{e}_1 \quad (4.102)$$

$$\vec{B} = B_0 \hat{e}_2 . \quad (4.103)$$

Por “regulares” nos referimos que las componentes de estos vectores obedecen la ecuación (4.25) cuando transformamos las coordenadas.

El producto cruz entre \vec{A} y \vec{B} puede ser escrito usando el determinante

$$\vec{A} \times \vec{B} = \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ A_0 & 0 & 0 \\ 0 & B_0 & 0 \end{vmatrix} = A_0 B_0 \hat{e}_3 , \quad (4.104)$$

o, usando el tensor de Levi-Civita

$$\vec{A} \times \vec{B} = A_i B_j \epsilon_{ijk} \hat{e}_k = A_0 B_0 \hat{e}_3 . \quad (4.105)$$

El vector resultante es mostrado en la figura 4.5. Notemos como la dirección de $\vec{A} \times \vec{B}$ está dada por la regla de la mano derecha. Si apuntamos los dedos de la mano en la dirección de \vec{A} y los rotamos en la dirección de \vec{B} , el pulgar apuntará en la dirección del resultado. Mantengamos en mente que el producto cruz no es conmutativo. Si el orden de la operación es invertido, es decir, si hacemos $\vec{B} \times \vec{A}$, el resultado apunta exactamente en la dirección opuesta.

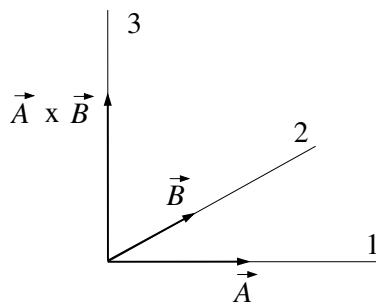


Figura 4.5: Vectores en el sistema de la mano derecha.

Consideremos ahora el sistema orientado a la izquierda, mostrado en la figura 4.6, con la base de vectores marcada con primas para diferenciarla de las coordenadas y la base del sistema de la mano derecha. Este sistema resulta de una inversión simple del eje 1 del sistema no primado. También puede ser visto como una reflexión del sistema derecho sobre el plano x_2x_3 . Las ecuaciones que relacionan los sistemas son

$$\begin{aligned}x'_1 &= -x_1 \\x'_2 &= +x_2 \\x'_3 &= +x_3\end{aligned}\tag{4.106}$$

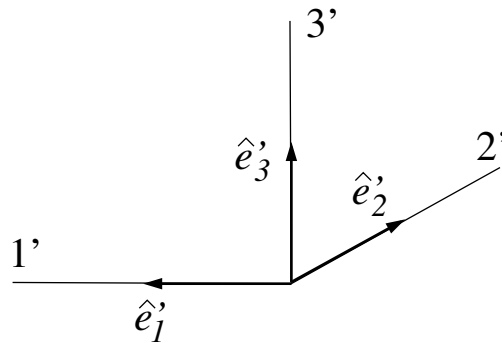


Figura 4.6: Sistema de la mano izquierda.

por tanto, la matriz transformación es

$$[a] = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .\tag{4.107}$$

Los vectores regulares \vec{A} y \vec{B} en el sistema prima son simplemente

$$\vec{A} = -A_0\hat{e}'_1\tag{4.108}$$

$$\vec{B} = B_0\hat{e}'_2 .\tag{4.109}$$

Sólo escribimos estos resultados porque son obvios. Recordemos que formalmente estos son obtenidos aplicando $[a]$ a las componentes de los vectores no primados. De la multiplicación matricial obtenemos

$$\begin{pmatrix} A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -A_0 \\ 0 \\ 0 \end{pmatrix}\tag{4.110}$$

y

$$\begin{pmatrix} B'_1 \\ B'_2 \\ B'_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ B_0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ B_0 \\ 0 \end{pmatrix} .\tag{4.111}$$

Es importante recordar que los vectores son los mismos objetos físicos en ambos sistemas de coordenadas. Están expresados en términos de distintas componentes y bases vectoriales.

Ahora formemos el producto cruz $\vec{A} \times \vec{B}$ en el sistema izquierdo. Para esto usaremos la relación de determinante

$$\vec{A} \times \vec{B} = \begin{vmatrix} \hat{e}'_1 & \hat{e}'_2 & \hat{e}'_3 \\ -A_0 & 0 & 0 \\ 0 & B_0 & 0 \end{vmatrix} = -A_0 B_0 \hat{e}'_3, \quad (4.112)$$

o, usando el tensor de Levi-Civita

$$\vec{A} \times \vec{B} = A'_i B'_j \epsilon_{ijk} \hat{e}'_k = A'_1 B'_1 \epsilon_{123} \hat{e}'_3 = -A_0 B_0 \hat{e}'_3. \quad (4.113)$$

Los vectores \vec{A} , \vec{B} y el producto cruz $\vec{A} \times \vec{B}$ para el sistema izquierdo son mostrados en la figura 4.7. Notemos como la regla de la mano derecha ya no nos sirve para encontrar la dirección del producto cruz. Si definimos el producto cruz usando el determinante en la ecuación (4.112), entonces debemos usar la regla de la mano izquierda si estamos en el sistema de coordenadas izquierdo.

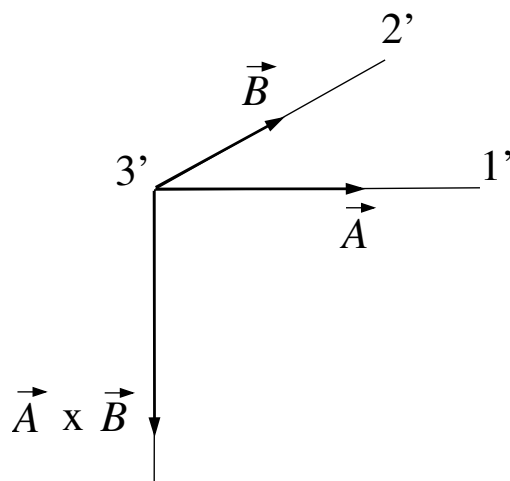


Figura 4.7: Vectores en el sistema de la mano izquierda.

Notemos algo peculiar. Comparando las figuras 4.7 y 4.5 observamos que, mientras \vec{A} y \vec{B} apuntan en las mismas direcciones, sus productos cruces no apuntan en las mismas direcciones. Cambiando la orientación del sistema de coordenadas, hemos cambiado la dirección del vector $\vec{A} \times \vec{B}$.

Miremos el producto cruz desde otro punto de vista. Si las componentes del vector $\vec{A} \times \vec{B}$ en el sistema no primado, dado por la ecuación (4.104), son transformados al sistema primado usando usando la matriz $[a]$, como lo hacemos para las componentes de los vectores regulares, obtenemos

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ A_0 B_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ A_0 B_0 \end{pmatrix}. \quad (4.114)$$

Combinando estas componentes con la base de vectores apropiada, obtenemos para el vector resultante del producto cruz

$$A_0 B_0 \hat{e}'_3 . \quad (4.115)$$

Este resultado difiere de la ecuación (4.112) por un signo menos. Para sortear esta dificultad, la cantidad formada por el producto cruz de dos vectores regulares es llamado un pseudovector. Los Pseudovectores también son llamados vectores axiales, mientras que los vectores regulares son llamados vectores polares. Si \vec{v} es un vector regular transforma de acuerdo a la ecuación (4.25). Por otra parte, si \vec{v} es un pseudovector, sus componentes transforman como

$$v'_r = |a| v_i a_{ri} . \quad (4.116)$$

De esta forma, la ecuación (4.114) se convierte en

$$\begin{pmatrix} (\vec{A} \times \vec{B})'_1 \\ (\vec{A} \times \vec{B})'_2 \\ (\vec{A} \times \vec{B})'_3 \end{pmatrix} = - \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ A_0 B_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -A_0 B_0 \end{pmatrix} \quad (4.117)$$

de donde resulta

$$\vec{A} \times \vec{B} = -A_0 B_0 \hat{e}'_3 , \quad (4.118)$$

de acuerdo con las ecuaciones (4.112) y (4.113).

En resumen, si \vec{v} es un vector regular sus componentes transforman como

$$v'_r = v_i a_{ri} . \quad (4.119)$$

En cambio, si es un pseudovector, sus componentes transforman como

$$v'_r = |a| v_i a_{ri} . \quad (4.120)$$

Si la orientación del sistema de dos sistemas de coordenadas ortonormales son el mismo, una transformación entre ellos tendrá $|a| = 1$, y los vectores y pseudovectores transformarán normalmente. Si los sistemas tienen orientación opuesta, $|a| = -1$ y los vectores transformarán normalmente, mientras que los pseudovectores cambiarán su dirección. Un vector generado por el producto cruz de dos vectores regulares es un pseudovector.

Es tentador pensar que toda esta parafernalia es un sutil error de signo embebidos en la definición de producto cruz. En algunos casos esto es correcto. Por ejemplo, cuando definimos la dirección del vector que define el campo magnético, que resulta ser un pseudovector, hemos elegido implícitamente el sentido del sistema de coordenadas que debe ser tratada de manera consistente. Otro ejemplo es el vector momento angular, el cual es definido por un producto cruz. Aunque se puede achacar este problema de pseudovectores de estos dos ejemplos es sólo un problema de definición, hay casos en que simplemente no se puede olvidar esta propiedad. Es posible diseñar una situación en la cual un experimento y su imagen especular no producen los resultados esperados, los cuales son simplemente la imagen especular una de otra. De hecho, el premio Nobel fue adjudicado a Lee y Yang por analizar estas violaciones a la conservación de paridad, lo cual va en contra de la lógica común. El experimento fue

realizado por primera vez por Wu, quien mostró este efecto con la emisión de partículas beta desde el Cobalto 60, bajo la influencia de interacciones débiles.

4.6.2. Pseudo-escalares.

Las ideas que nos dejaron los pseudovectores se aplican también a los escalares. Un escalar es invariante ante cualquier cambio del sistema de coordenadas. En cambio, un pseudoescalar cambia de signo si la orientación del sistema cambia. Un pseudoescalar involucrado en una transformación, gobernado por una matriz de transformación $[a]$, obedecerá

$$S' = |a|S . \quad (4.121)$$

Un buen ejemplo de pseudoescalar se puede derivar del comportamiento del producto cruz. El volumen de un paralelogramo tridimensional, mostrado en la figura 4.8, puede ser escrito por

$$\text{Volumen} = (\vec{A} \times \vec{B}) \cdot \vec{C} . \quad (4.122)$$

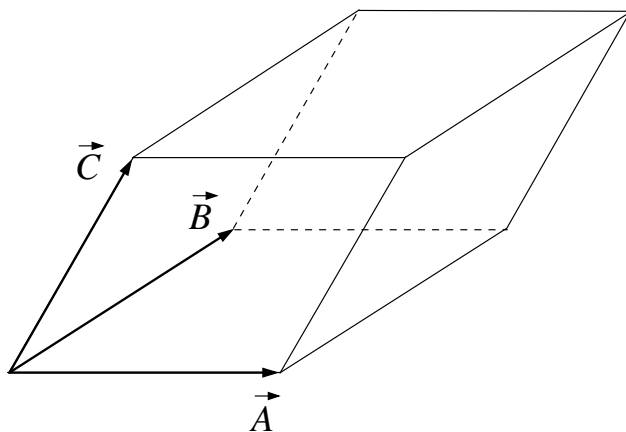


Figura 4.8: El paralelogramo.

En un sistema de coordenadas derecho, el vector formado por $\vec{A} \times \vec{B}$ apuntará hacia arriba. Por tanto, en un sistema derecho,

$$(\vec{A} \times \vec{B}) \cdot \vec{C} > 0 . \quad (4.123)$$

Pero en un sistema de coordenadas izquierdo, apunta hacia abajo, por tanto

$$(\vec{A} \times \vec{B}) \cdot \vec{C} < 0 . \quad (4.124)$$

Interpretado de esta forma, el volumen de un paralelogramo es un pseudoescalar.

4.6.3. Pseudo-tensores.

Los pseudotensores están definidos como uno espera. Bajo una transformación, las componentes de un pseudotensor obedecen

$$T'_{rst\dots} = |a|T_{ijk\dots}a_{ri}a_{sj}a_{tk}\dots, \quad (4.125)$$

la cual es igual a lo que obedece un tensor regular, salvo por el término $|a|$.

Nuevamente utilizamos el producto cruz como ejemplo. Consideremos dos sistemas de coordenadas. El sistema primado es un sistema derecho, y el otro, con el sistema no primado, es izquierdo. Usando el símbolo de Levy-Civita en los dos sistemas para generar el producto cruz $\vec{A} \times \vec{B}$ obtenemos

$$A_i B_j \epsilon_{ijk} \hat{e}_k = -A'_r B'_s \epsilon'_{rst} \hat{e}_t. \quad (4.126)$$

El signo menos aparece porque como fue mostrado antes, la dirección física del producto cruz es diferente en los dos sistemas de coordenadas. Ahora, las transformaciones de coordenadas de vectores regulares pueden ser usadas para encontrar la relación entre ϵ_{ijk} y ϵ'_{rst} . Ya que todos los vectores involucrados son regulares, es decir, \vec{A} , \vec{B} y los vectores base, estos transforman de acuerdo a la ecuación (4.25). Escribiendo las componentes primadas de estos vectores en términos de los no primados, la ecuación (4.126) se convierte en

$$A_i B_j \epsilon_{ijk} \hat{e}_k = -A_i B_j a_{ri} a_{sj} a_{tk} \epsilon'_{rst} \hat{e}_k. \quad (4.127)$$

Esta última expresión es cierta para \vec{A} y \vec{B} arbitrarios, por tanto obtenemos

$$\epsilon_{ijk} = -a_{ri} a_{sj} a_{tk} \epsilon'_{rst}. \quad (4.128)$$

Tengamos en mente que esto se aplica sólo cuando dos sistemas tienen orientaciones distintas. Si ambos sistemas tienen la misma orientación, el signo menos desaparece. Por tanto, para el caso general de una transformación arbitraria entre dos sistemas ortonormales, el símbolo de Levy-Civita obedece

$$\epsilon_{ijk} = |a| a_{ri} a_{sj} a_{tk} \epsilon'_{rst}. \quad (4.129)$$

Por tanto, el símbolo de Levy-Civita es un pseudotensor.

Capítulo 5

Sistema de coordenadas no ortogonales.

versión final 1.0-080415¹

5.1. Breve recuerdo de transformaciones tensoriales.

Ya discutimos cómo un tensor es definido por su comportamiento bajo transformaciones de coordenadas. Con una cuota de sarcasmo, la definición que dimos fue “un tensor es una cantidad que transforma como tensor”. Lo que esto significa es que las reglas de transformación son suficientes como para caracterizar a los tensores con sus propiedades especiales. Si un objeto transforma entre sistemas de coordenadas usando las reglas de transformación tensorial, podemos decir legítimamente que el objeto es un tensor.

Recordemos, los elementos de un tensor pueden transformar usando una matriz de transformaciones, cuyos elementos pueden ser obtenidos de las ecuaciones que relacionan las coordenadas de los dos sistemas. Para transformaciones entre sistemas cartesianos, los elementos de esta matriz de transformación $[a]$ están dados por

$$a_{ij} = (\hat{e}'_i \cdot \hat{e}_j) = \frac{\partial x'_i}{\partial x_j} . \quad (5.1)$$

En esta ecuación, el sistema original tiene coordenadas x_i y vectores base \hat{e}_i . El sistema es transformado al sistema primado, el cual tiene coordenadas x'_i y vectores base \hat{e}'_i . Para sistemas de coordenadas ortonormales, la inversa de esta matriz de transformación es siempre su transpuesta

$$a_{ij}^{-1} = a_{ji} . \quad (5.2)$$

Un tensor arbitrario de rango n puede ser expresado tanto en el sistema primado como en el no primado por

$$\vec{T} = T_{ijk\dots} \hat{e}_i \hat{e}_j \hat{e}_k \dots = T'_{rst\dots} \hat{e}'_r \hat{e}'_s \hat{e}'_t \dots , \quad (5.3)$$

¹Este capítulo está basado en el décimo cuarto capítulo del libro: *Mathematical Physics* de Brusse Kusse & Erik Westwig, editorial JOHN WILEY & SONS, INC..

donde hay n subíndices y n vectores base en cada término. $T_{ijk\dots}$ y $T'_{rst\dots}$ son los elementos del tensor en el sistema de coordenadas no primado y primado, respectivamente. Los dos conjuntos de elementos están relacionados por la ecuación matricial de transformación

$$T'_{rst\dots} = T_{ijk\dots} a_{ri} a_{sj} a_{tk} \dots , \quad (5.4)$$

donde la matriz $[a]$ aparece n veces. La transformación inversa es

$$T_{rst\dots} = T'_{ijk\dots} a_{ir} a_{js} a_{kt} \dots . \quad (5.5)$$

Nuestra propuesta fundamental es que cualquier cantidad que transforma en la manera descrita por la ecuación (5.4) es por definición un tensor.

Como un vector es un tensor de primer rango, las transformaciones de las componentes son descritas por las ecuaciones (5.4) y (5.5). Si escribimos un vector en dos sistemas de coordenadas como

$$\vec{v} = v_i \hat{e}_i = v'_r \hat{e}'_r , \quad (5.6)$$

la relación entre las componentes está dado por

$$v'_r = v_i a_{ri} , \quad (5.7)$$

y la inversa

$$v_r = v'_i a_{ir} . \quad (5.8)$$

Los escalares son invariantes ante una transformación de coordenadas. Podemos pensar que un escalar es un tensor de rango cero. El único elemento de un tensor de rango cero no tiene subíndices y no está combinado con una base vectorial. La ecuación (5.4) se reduce a

$$S = S' \quad (5.9)$$

donde S (o S') es el único elemento del escalar.

Ejemplo

Como un ejemplo del uso de las propiedades de las transformaciones para identificar a un objeto como tensor, consideremos la delta de Kronecker. Recordemos que este símbolo fue introducido para formar el producto punto en sistemas de coordenadas ortonormales. El producto punto entre dos vectores, \vec{A} y \vec{B} escrito en dos sistemas ortonormales de coordenadas, uno primado y otro primado, puede ser escrito por

$$\vec{A} \cdot \vec{B} = A_i B_j \delta_{ij} = A'_r B'_s \delta'_{rs} . \quad (5.10)$$

Ahora, sabemos que tanto δ_{ij} como δ'_{rs} pueden ser escritos como matrices unitarias, tal como [1]. Sin embargo, para los propósitos de esta discusión, observemos las consecuencias de imponer que las dos expresiones para el producto punto en la ecuación (5.10) sean iguales, y que A_i y B_i son componentes vectoriales, y por tanto, transforma de acuerdo a la ecuación (5.7). Sustituyendo en la ecuación (5.10), tenemos para A'_r y B'_s

$$\begin{aligned} A_i B_j \delta_{ij} &= a_{ri} A_i a_{sj} B_j \delta'_{rs} . \\ &= A_i B_j a_{ri} a_{sj} \delta'_{rs} . \end{aligned} \quad (5.11)$$

Como esta expresión debe ser verdadera para cualquier \vec{A} y \vec{B} , podemos escribir

$$\delta_{ij} = a_{ri} a_{sj} \delta'_{rs} . \quad (5.12)$$

Invirtiendo esta expresión, obtenemos

$$\delta'_{ij} = a_{ir} a_{js} \delta_{rs} . \quad (5.13)$$

Comparando las ecuaciones (5.12) y (5.13) con las ecuaciones (5.4) y (5.5), se observa que los elementos de la delta de Kronecker transforman como los elementos de un tensor de segundo rango. Por tanto, el símbolo de la delta de Kronecker es un tensor de segundo rango, el cual puede ser expresado con una base vectorial como

$$\vec{\delta} = \delta_{ij} \hat{e}_i \hat{e}_j = \delta'_{ij} \hat{e}'_i \hat{e}'_j . \quad (5.14)$$

5.2. Sistemas de coordenadas no ortogonales.

Hasta este punto, hemos tratado sólo con sistemas de coordenadas ortonormales. En sistemas cartesianos, los vectores base \hat{e}_i son independientes de la posición y ortonormales, por lo tanto $\hat{e}_i \cdot \hat{e}_j = \delta_{ij}$. En sistemas curvilíneos, los vectores base \hat{q}_i no son independientes de la posición, pero aún son ortonormales, por tanto $\hat{q}_i \cdot \hat{q}_j = \delta_{ij}$. Ahora consideraremos sistemas no ortonormales. Para distinguir estos sistemas, escribiremos los vectores base de sistemas de coordenadas no ortonormales como \hat{g}_i , y la condición de no ortonormalidad se convierte en $\hat{g}_i \cdot \hat{g}_j \neq \delta_{ij}$. Para mantener esta discusión y las derivaciones de la forma más simple, nos limitaremos a sistemas de coordenadas donde las bases vectoriales no varíen con la posición. Obviamente esto no es el caso más general de sistemas de coordenadas no ortogonales, pero es suficiente para demostrar las ideas de covarianza, contravarianza y métrica.

En física, los sistemas de coordenadas no ortonormales aparecen, por ejemplo, en relatividad (tanto especial como general). El postulado básico de la relatividad especial es que la velocidad de la luz c es la misma para todos los sistemas de referencia. Como consecuencia de este postulado, la posición y el tiempo de algún fenómeno físico (un “evento”) cambia tal como cambie el sistema de referencia. Es muy similar a cómo las componentes de un vector cambian cuando transformamos los ejes coordenados. Si restringimos el movimiento a una coordenada espacial, un evento puede ser descrito por dos coordenadas, una coordenada espacial y una temporal. Como será mostrado, la observación de un evento en dos sistemas de coordenadas distintos, uno primado y otro no primado, puede ser dibujado como un punto usando el conjunto de ejes combinados, como es mostrado en la figura 5.1. Tomando sus componentes con respecto a ambos ejes coordenados, podemos obtener todas las relaciones impuestas por la relatividad especial. Notemos cómo los ejes x y ct se intersectan en ángulos rectos, pero los ejes x' y ct' no lo hacen. Mientras el sistema no primado parece ser ortogonal, el sistema primado parece ser un sistema no ortogonal e inclinado.

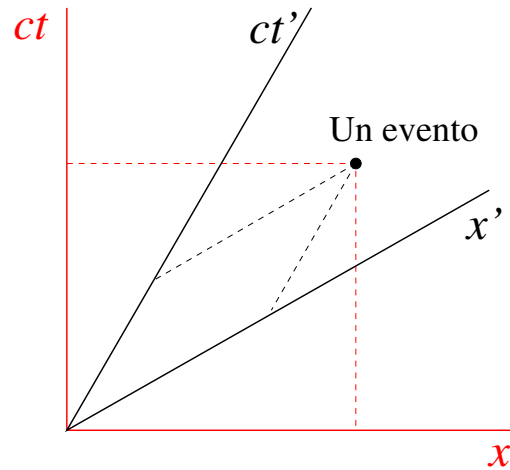


Figura 5.1: Los sistemas de coordenadas de la Relatividad Especial.

El postulado básico de la relatividad general es que la gravedad y la aceleración son equivalentes. Los eventos observados en un campo gravitacional aparecen como si estos estuviesen siendo observados en un sistema de coordenadas acelerado. Esto implica que la luz propagándose a través del campo gravitacional de un objeto masivo, como una estrella, se debería doblar, como es mostrado en la figura 5.2. Esto podría causar que la posición aparente de una estrella se desvíe de su posición actual. Este fenómeno fue observado por primera vez por Arthur Eddington, el cual midió la pequeña deflexión de las estrellas provocadas por el Sol durante el eclipse total de 1919. Los caminos que siguen los rayos de luz a través del espacio son llamados geodésicas. Una elección natural para las líneas de la grilla de un sistema de coordenadas localizado, siguen estas geodésicas, como es mostrado en la figura 5.2. No discutiremos ejemplos de este tipo de sistemas, pero si nos restringiremos a discutir los sistemas inclinados, donde las bases vectoriales no son ortonormales, pero son espacialmente invariantes.

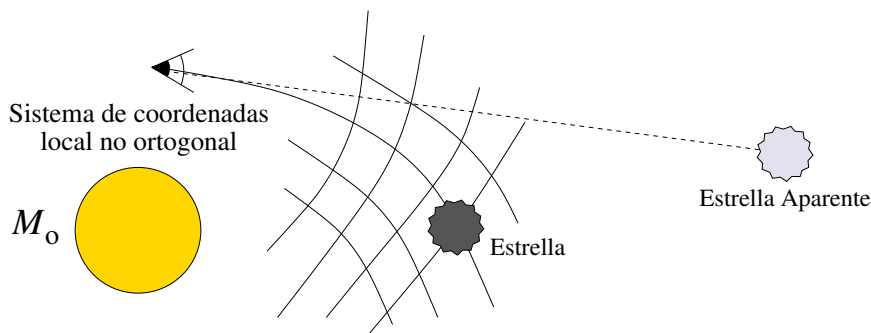


Figura 5.2: Un sistema de coordenadas de la Relatividad General.

5.2.1. Un sistema de coordenadas inclinado.

Consideremos un sistema de coordenadas cartesiano (x_1, x_2) y el sistema primado no ortonormal (x'_1, x'_2) , como es mostrado en la figura 5.3. También son representados en la

figura dos pares de vectores base y un vector arbitrario \vec{v} . La base vectorial de sistema no primado es ortonormal

$$\hat{e}_i \cdot \hat{e}_j = \delta_{ij} . \quad (5.15)$$

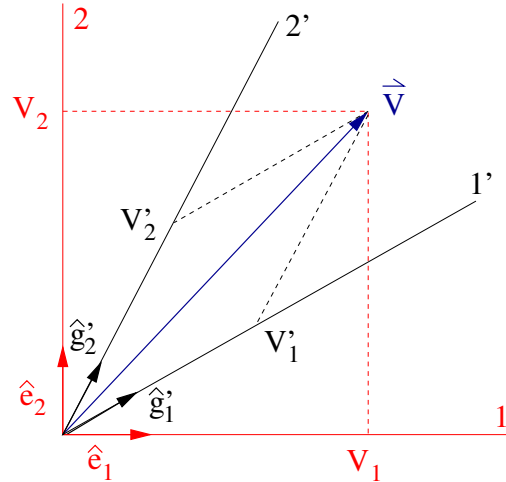


Figura 5.3: Un sistema de coordenadas ortonormal y otro inclinado.

La base vectorial del sistema inclinado son elegidos para que sean vectores unitarios,

$$\hat{g}'_1 \cdot \hat{g}'_1 = \hat{g}'_2 \cdot \hat{g}'_2 = 1 , \quad (5.16)$$

pero no son ortonormales

$$\hat{g}'_i \cdot \hat{g}'_j \neq \delta_{ij} . \quad (5.17)$$

Como podemos ver, el formalismo que es desarrollado permite que los vectores base no sean unitarios. Sin embargo, para comenzar el tratamiento de la manera más sencilla, suponemos que la base vectorial del sistema inclinado satisface la ecuación (5.16).

En el sistema ortonormal, el vector \vec{v} puede ser expresado como la suma de sus componentes proyectadas de forma paralela en los ejes, como es mostrado en la figura 5.3, junto a la correspondiente base vectorial

$$\vec{v} = v_1 \hat{e}_1 + v_2 \hat{e}_2 . \quad (5.18)$$

Estas componentes vectoriales son sólo los tamaños proyectados de \vec{v} a lo largo de los ejes del sistema no primado y pueden ser determinados con trigonometría o siguiendo la manipulación vectorial correspondiente. Una componente particular es obtenida haciendo el producto punto entre \vec{v} y el correspondiente vector base. Por ejemplo, para encontrar v_1

$$\begin{aligned} \vec{v} \cdot \hat{e}_1 &= (v_1 \hat{e}_1 + v_2 \hat{e}_2) \cdot \hat{e}_1 \\ &= v_1 (\hat{e}_1 \cdot \hat{e}_1) + v_2 (\hat{e}_2 \cdot \hat{e}_1) \\ &= v_1 \delta_{11} + v_2 \delta_{21} \\ &= v_1 . \end{aligned} \quad (5.19)$$

Esto resulta bello sólo por la ortogonalidad de los vectores base.

En el sistema primado, el mismo vector puede ser escrito en términos de las componentes proyectadas de forma paralela en los ejes y los vectores base prima, como es mostrado en la figura 5.3,

$$\vec{v} = v'_1 \hat{g}'_1 + v'_2 \hat{g}'_2 . \quad (5.20)$$

Estas componentes también pueden ser determinadas por trigonometría, pero como tenemos una geometría inclinada, no es tan sencillo como lo es en el sistema ortogonal. Como la base de vectores primados no son ortogonales, un intento por aislar una componente particular por una manipulación vectorial, similar a la desarrollada en la ecuación (5.19) falla

$$\begin{aligned} \vec{v} \cdot \hat{g}'_1 &= (v'_1 \hat{g}'_1 + v'_2 \hat{g}'_2) \cdot \hat{g}'_1 \\ &= v'_1 (\hat{g}'_1 \cdot \hat{g}'_1) + v'_2 (\hat{g}'_2 \cdot \hat{g}'_1) \\ &= v'_1 + v'_2 (\hat{g}'_2 \cdot \hat{g}'_1) \\ &\neq v_1 . \end{aligned} \quad (5.21)$$

Al parecer, las manipulaciones vectoriales en sistemas de coordenadas no ortogonales son mucho más difíciles que en los sistemas ortonormales. Afortunadamente, hay algunas técnicas formales que simplifican el proceso. En la próxima sección, introduciremos los conceptos de covarianza, contravarianza, y el tensor métrico. Usando estas herramientas, el producto punto entre dos vectores tienen la misma forma tanto en un sistema ortogonal como en un sistema no ortogonal.

5.2.2. Covarianza, contravarianza y métrica.

La complicación básica introducida por un sistema de coordenadas no-ortogonal es evidentemente en la operación producto punto. En un sistema ortonormal de dos dimensiones descrito anteriormente, el producto interno entre dos vectores es dado por

$$\begin{aligned} \vec{A} \cdot \vec{B} &= A_i \hat{e}_i \cdot B_j \hat{e}_j \\ &= A_i B_j \delta_{ij} \\ &= A_1 B_1 + A_2 B_2 . \end{aligned} \quad (5.22)$$

Si este mismo producto interno es realizado en el sistema no-ortogonal de la figura 5.3 el resultado contiene algunos términos extras:

$$\begin{aligned} \vec{A} \cdot \vec{B} &= A'_i \hat{g}'_i \cdot B'_j \hat{g}'_j \\ &= A'_i B'_j (\hat{g}'_i \cdot \hat{g}'_j) \\ &= A'_1 B'_1 + A'_2 B'_2 + (A'_1 B'_2 + A'_2 B'_1) (\hat{g}'_1 \cdot \hat{g}'_2) . \end{aligned} \quad (5.23)$$

El producto interno evaluado en el sistema no-ortonormal, expresado en (5.23), puede ser puesto en la forma de la ecuación (5.22) rearrregandolo como sigue:

$$\vec{A} \cdot \vec{B} = A'_1 (B'_1 + B'_2 (\hat{g}'_1 \cdot \hat{g}'_2)) + A'_2 (B'_1 (\hat{g}'_1 \cdot \hat{g}'_2) + B'_2) . \quad (5.24)$$

Ahora definamos un nuevo conjunto de componentes para \vec{B} como

$$\begin{aligned}\tilde{B}'_1 &= B'_1 + B'_2(\hat{g}'_1 \cdot \hat{g}'_2) \\ \tilde{B}'_2 &= B'_1(\hat{g}'_1 \cdot \hat{g}'_2) + B'_2.\end{aligned}\tag{5.25}$$

Estas cantidades son llamadas las componentes **covariantes** de \vec{B} , mientras que las componentes originales son llamadas **contravariantes**. Claramente, el vector \vec{B} no puede ser expresado por la combinación de estas nuevas componentes covariantes con los vectores bases \hat{g}'_1 y \hat{g}'_2 :

$$\vec{B} \neq \tilde{B}'_1 \hat{g}'_1 + \tilde{B}'_2 \hat{g}'_2.\tag{5.26}$$

Sin embargo, con estas componentes el producto evaluado en el sistema inclinado puede ser puesto en una forma simple

$$\begin{aligned}\vec{A} \cdot \vec{B} &= A'_i \tilde{B}'_i \\ &= A'_1 \tilde{B}'_1 + A'_2 \tilde{B}'_2.\end{aligned}\tag{5.27}$$

Notemos que el producto interno también puede ser escrito como

$$\vec{A} \cdot \vec{B} = \tilde{A}'_i B'_i,\tag{5.28}$$

con las componentes covariantes de \vec{A} definidas como

$$\begin{aligned}\tilde{A}'_1 &= A'_1 + A'_2(\hat{g}'_1 \cdot \hat{g}'_2) \\ \tilde{A}'_2 &= A'_1(\hat{g}'_1 \cdot \hat{g}'_2) + A'_2.\end{aligned}\tag{5.29}$$

El producto interno necesita estar formado con una mezcla de componentes covariantes y contravariantes, pero no importa que vector es expresado en que tipo de componentes.

Estos argumentos pueden ser extendidos a sistemas no-ortogonales de dimensión arbitraria. La restricción que los vectores bases estén normalizados a la unidad puede ser levantada. Las componentes covariantes pueden ser generadas a partir de las componentes contravariantes usando la expresión general

$$\tilde{A}'_i = A'_j(\hat{g}'_i \cdot \hat{g}'_j).\tag{5.30}$$

Hemos usado convención de Einstein, lo cual implica suma sobre j . Para un sistema de coordenada n -dimensional en cada suma habría n términos. Notemos que si el sistema de coordenadas es ortonormal la ecuación (5.30) se reduce a $\tilde{A}'_i = A'_i$ y las componentes covariante y contravariantes son iguales. En este caso, ambas ecuaciones (5.27) y (5.28) se revierten a la ecuación (5.22). Esto es importante, porque implica que esta nueva notación es suficientemente general para manejar todos nuestros previos sistemas de coordenadas Cartesianos y curvilíneos, tanto como los nuevos no-ortogonales.

Existe otra manera de expresar el producto interno entre dos vectores en un sistema no-ortogonal que hace uso de una cantidad llamada la **métrica**. Como veremos más tarde, la métrica es un tensor de rango dos. Los elementos de la métrica, en un sistema sin primas, están definidos como

$$M_{ij} = \hat{g}_i \cdot \hat{g}_j.\tag{5.31}$$

Notemos que esta definición implica que la métrica es simétrico:

$$M_{ij} = M_{ji} . \quad (5.32)$$

Usando la métrica la ecuación (5.30) puede ser escrita como

$$\tilde{A}_i = A_j M_{ij} . \quad (5.33)$$

La métrica convierte las componentes contravariantes en componentes covariantes.

Ahora el producto interno entre \vec{A} y \vec{B} pueden ser reescrito

$$\vec{A} \cdot \vec{B} = A_i B_j M_{ij} . \quad (5.34)$$

La suma sobre ambos i y j está indicada en el lado derecho de la ecuación. Si realizamos la suma sobre i primero, la ecuación (5.34) se convierte en

$$\vec{A} \cdot \vec{B} = \tilde{A}_j B_j . \quad (5.35)$$

Cuando la suma sobre j es realizada primero, la ecuación (5.34) se convierte en

$$\vec{A} \cdot \vec{B} = A_i \tilde{B}_i . \quad (5.36)$$

Cuando la ecuación (5.34) es usada para el producto interno, las componentes vectoriales no se mezclan. Las componentes contravariantes son usadas para ambos vectores. Si el sistema es ortonormal $M_{ij} = \delta_{ij}$, resultando el producto interno standard para un sistema ortonormal. Notemos que la métrica es determinada solamente por los vectores bases del sistema de coordenadas. Esto se volverá un hecho importante y nos permitirá identificar a la métrica como un tensor de rango dos.

En resumen, hay dos maneras de realizar el producto interno entre dos vectores en un sistema no-ortogonal. Una manera es usar las componentes covariantes y contravariantes, como fue hecho en las ecuaciones (5.27) y (5.28). Un método completamente equivalente es usar la métrica y las componentes regulares contravariantes del vector, como demostramos en la ecuación (5.34). Estos argumentos pueden ser naturalmente extendidos al producto interno entre cantidades tensoriales, pero esta generalización será pospuesta hasta que las ecuaciones de transformación para sistema no-ortogonales sean trabajadas.

5.2.3. Transformaciones de componentes vectoriales contravariantes.

Imaginemos dos sistemas de coordenadas inclinados diferentes, como es mostrado en la figura 5.4. Queremos encontrar como la componentes contravariantes de un vector expresadas en el primer sistema pueden ser transformadas al segundo sistema. El primer sistema tiene las coordenadas no primadas x_i y vectores base \hat{g}_i , mientras que el segundo sistema usa las coordenadas primadas x'_i y vectores base \hat{g}'_i . Recordemos que estamos limitados a sistemas de coordenadas con vectores base constantes. Sean las ecuaciones generales que relacionan los dos conjuntos de coordenadas

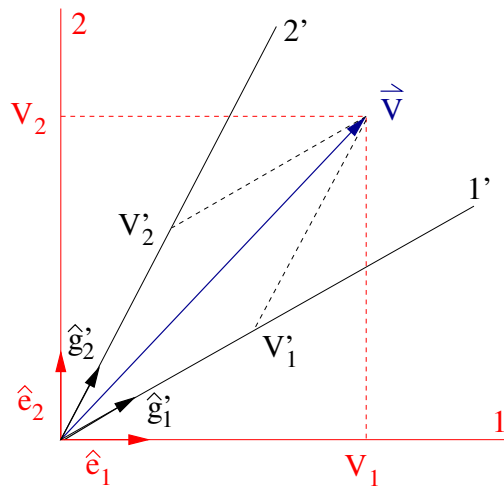


Figura 5.4: Dos sistemas de coordenadas inclinados.

$$\begin{aligned} x'_i &= x'_i(x_1, x_2, x_3) \\ x_i &= x_i(x'_1, x'_2, x'_3) . \end{aligned} \quad (5.37)$$

Habrán sólo un par de ecuaciones para cada dimensión de los sistemas.

En nuestro trabajo previo tratando con transformaciones entre sistemas de coordenadas ortonormales, fuimos capaces de relacionar las componentes vectoriales de un sistema con el otro vía la matriz de transformación $[a]$,

$$v'_i = a_{ij}v_j . \quad (5.38)$$

La restricción para sistemas ortonormales nos permitió invertir esta expresión de forma trivial, ya que se transformó en $a_{ij}^{-1} = a_{ji}$. Podemos escribir una relación similar para la ecuación (5.38) para las transformaciones entre sistemas no ortonormales, pero necesitamos tener más cuidado, ya que la inversa de la matriz transformación no es su transpuesta. Para no perder la pista entre las transformaciones que aceptan esta inversión simple y las que no, reservaremos la matriz $[a]$ para las transformaciones entre sistemas ortonormales. La matriz $[t]$ representará las transformaciones entre las coordenadas no primadas y las primadas, donde los sistemas pueden ser no ortonormales

$$v'_i = t_{ij}v_j . \quad (5.39)$$

La operación en reversa, una transformación entre las coordenadas primadas y las coordenadas no primadas, usaremos la matriz $[g]$,

$$v_i = g_{ij}v'_j , \quad (5.40)$$

donde $g_{ij} = t_{ij}^{-1} \neq t_{ji}$. Por su definición, se sigue que $t_{ij}g_{jk} = \delta_{ik}$. Discutiremos detalladamente la relación entre las matrices $[t]$ y $[g]$ más adelante. En ambas expresiones, las componentes vectoriales son componentes contravariantes regulares de \vec{v} , no las componentes covariantes que presentamos.

$$t_{ij} = \frac{\partial x'_i}{\partial x_j} ; \quad g_{jk} = \frac{\partial x_j}{\partial x'_k}$$

$$t_{ij} g_{jk} = \frac{\partial x'_i}{\partial x_j} \frac{\partial x_j}{\partial x'_k} = \delta_{ik}$$

Todos los vectores en un punto dado transforman usando la misma matriz $[t]$. Para determinar los elementos t_{ij} , es más fácil considerar el vector desplazamiento $d\vec{r}$, el cual en los dos sistemas de coordenadas está dado por

$$d\vec{r} = dx_i \hat{g}_i = dx'_i \hat{g}'_i. \quad (5.41)$$

Aplicando esta igualdad a la ecuación (5.39), tenemos

$$dx'_i = t_{ij} dx_j. \quad (5.42)$$

Refiriéndose a las ecuaciones (5.37), obtenemos la relación

$$dx'_i = \frac{\partial x'_i}{\partial x_j} dx_j, \quad (5.43)$$

y los elementos de la matriz transformación pueden ser escritos como

$$t_{ij} = \frac{\partial x'_i}{\partial x_j}. \quad (5.44)$$

Hasta ahora, estos resultados se parecen mucho a las transformaciones cartesianas que ya habíamos visto. De hecho, la ecuación para las componentes de $[t]$ dadas en la ecuación (5.44) es el mismo resultado obtenido para la matriz $[a]$ entre sistemas cartesianos. Las complicaciones aparecen cuando tratamos de invertir estas ecuaciones. Como ya hemos mencionado, la inversión de $[t]$ no es simplemente calcular la traspuesta. Una forma general de obtener $[t]^{-1}$, la cual estamos llamando $[g]$, es utilizar la expresión

$$g_{ij} = t_{ij}^{-1} = \frac{c_{ji}}{|t_{ij}|}, \quad (5.45)$$

donde c_{ji} es el cofactor ji de la matriz t_{ij} . Del álgebra de matrices, este cofactor es definido como $(-1)^{i+j}$ por el determinante de la matriz t_{ij} , con la columna j -ésima y la columna i -ésima removida. La matriz $[g]$ puede también ser obtenida desde las ecuaciones que relacionan las coordenadas, exactamente de la misma manera que se llega a la ecuación (5.44)

$$g_{ij} = t_{ij}^{-1} = \frac{\partial x_i}{\partial x'_j}. \quad (5.46)$$

Las matrices $[t]$ y $[g]$ pueden también ser usadas para relacionar una base vectorial con la otra. Usando las componentes contravariantes, cualquier vector \vec{v} puede ser expresado en el sistema primado o el no primado como

$$\vec{v} = v_j \hat{g}_j = v'_i \hat{g}'_i. \quad (5.47)$$

Sustituyendo la ecuación (5.39) en la ecuación (5.47) obtenemos

$$\vec{v} = v_j \hat{g}_j = v_j t_{ij} \hat{g}'_i. \quad (5.48)$$

Como esta expresión es válida para cualquier \vec{v} , se debe cumplir


$$\hat{g}_j = t_{ij} \hat{g}'_i. \quad (5.49)$$

Haciendo el proceso análogo, pero usando $[g]$ en vez de $[t]$, obtenemos

$$\hat{g}'_j = g_{ij} \hat{g}_i . \quad (5.50)$$

Notemos que las componentes vectoriales contravariantes son transformadas por contracciones sobre el segundo subíndice de t_{ij} o g_{ij} , mientras que las bases vectoriales son transformados contrayendo sobre el primer subíndice.

Para resumir los resultados de esta sección, la transformación entre los dos sistemas de coordenadas no ortonormales es gobernada por las relaciones

$t_{ij} = \partial x'_i / \partial x_j$	$g_{ij} = \partial x_i / \partial x'_j$	
$v'_i = t_{ij} v_j$	$v_i = g_{ij} v'_j$	
$\hat{g}_j = t_{ij} \hat{g}'_i$	$\hat{g}'_j = g_{ij} \hat{g}_i$	

5.2.4. Notación de subíndices y superíndices.

Antes de proceder con una discusión de cómo las componentes de un vector covariantes transforman, resulta conveniente introducir una notación nueva. La notación con tilde (\tilde{v}_i) que hemos usado para las componentes de los vectores covariantes es engorrosa. No es obvio que las siguientes convenciones son mucho mejores, pero proveen un mecanismo valioso para mantener la pista de cuál tipo de componentes (contravariantes o covariantes) deben ser usados en una expresión. Las componentes vectoriales proyectadas en forma paralela, las cuales hemos llamado las componentes contravariantes, serán anotadas con un superíndice, mientras que para las nuevas componentes covariantes se usará un subíndice en vez de un tilde. Por ejemplo, las componentes contravariantes del vector \vec{v} son v^i , mientras las componentes covariantes serán v_i .

Una ventaja de esta nueva notación es evidente viendo la forma del producto interno. Con la convención ya propuesta, podemos escribir el producto punto de \vec{A} y \vec{B} como

$$\vec{A} \cdot \vec{B} = A^i B_i = A_i B^i . \quad (5.51)$$

Notemos que el índice sobre el cual sumamos aparece una vez como subíndice y otra como superíndice. Esto es, por supuesto, lo mismo que decir que la suma es hecha sobre cantidades covariantes y contravariantes mezcladas. Este proceso de mezclar subíndices y superíndices persistirá sobre casi todas las contracciones sobre un índice repetido. También funcionará cuando queramos formar un vector desde sus componentes con la interpretación adecuada de los vectores base. Sabemos que el vector puede ser formado con las componentes contravariantes y la base vectorial

$$\vec{v} = v^i \hat{g}_i . \quad (5.52)$$

Para ser consistentes con la convención de subíndices y superíndices, las bases vectoriales deben ser escritas con subíndices y ser consideradas covariantes. Veremos en la próxima sección, que esta conclusión es consistente con la forma en que estas bases transforman.

Esta convención también previene que accidentalmente formemos un vector combinando sus componentes covariantes con la base vectorial \hat{g}_i

$$\vec{v} \neq v_i \hat{g}_i . \quad (5.53)$$

La notación nos advierte que esto no es correcto ya que ambos índices aparecen como subíndices.

En la sección anterior generamos varias relaciones, las cuales describieron cómo las componentes contravariantes del vector \vec{v} transforman entre dos sistemas inclinados. ¿Cómo debiesen ser modificados la presentación de estos resultados para ser consistente con la nueva convención? En la sección anterior escribimos

$$v'_i = t_{ij} v_j . \quad (5.54)$$

Ahora estas componentes vectoriales deben ser escritas de manera correcta. Para ser consistentes con esta nueva notación, uno de los índices de la matriz transformación necesita ser un subíndice y otra con superíndice,

$$v'^i = t^i_j v^j , \quad (5.55)$$

donde

$$t^i_j = \frac{\partial x'^i}{\partial x^j} . \quad (5.56)$$

De manera similar, la inversión de la ecuación (5.55) se convierte en

$$v^i = g^i_j v'^j , \quad (5.57)$$

donde

$$g^i_j = \frac{\partial x^i}{\partial x'^j} . \quad (5.58)$$

Notemos cómo en las ecuaciones (5.56) y (5.58) la componente con superíndice en el denominador de la derivada parcial resulta en un subíndice en el lado izquierdo de esta expresión. Esta es una propiedad general de las derivadas parciales con respecto a cantidades covariantes y contravariantes. Una derivada parcial con respecto a una cantidad contravariante produce un resultado covariante, mientras que una derivada parcial con respecto a una cantidad covariante da como resultado una cantidad contravariante. Probaremos este hecho más tarde en este capítulo.

Estas matrices de transformación tienen lo que es llamado una combinación de propiedades contravariantes/covariantes. Ellas son contravariantes con respecto a un índice, pero covariantes con respecto al otro.

Con la notación que usábamos hasta comenzar esta sección, la naturaleza recíproca de $[t]$ y $[g]$ era indicada por la ecuación $t_{ij} g_{jk} = \delta_{ik}$. Pero ahora, para ser consistentes con la nueva notación, anotaremos

$$t^i_j g^j_k = \delta^i_k . \quad (5.59)$$

La delta de Kronecker, escrita de esta manera, también presenta la forma mezclada de covariante y contravariante.

Las ecuaciones (5.49) y (5.50), las cuales indican cómo transforman las bases vectoriales, son escritas usando la notación de subíndices/superíndices por

$$\begin{aligned}\hat{g}_j &= t^i_j \hat{g}'_i \\ \hat{g}'_j &= g^i_j \hat{g}_i .\end{aligned}\tag{5.60}$$

Notemos la importancia de las posiciones horizontales de los índices de la matriz transformación. En las ecuaciones (5.55) y (5.57) la suma era sobre el segundo índice de la matriz, mientras que estas sumas son sobre el primer índice. Esto previene de escribir los elementos de la matriz $[t]$ como t^i_j , ya que esto podría indicar cuál índice viene primero.

Deberíamos también reescribir las relaciones que involucran la métrica usando la nueva notación. Nuestra definición previa de la métrica fue en términos de una base vectorial covariante. Por tanto, la ecuación (5.31) se mantiene

$$M_{ij} = \hat{g}_i \cdot \hat{g}_j ,\tag{5.61}$$

y ambos índices permanecen como subíndices. De esta manera, los elementos de la métrica son puramente covariantes, ya que ambos índices son subíndices. La métrica convierte las componentes contravariantes de un vector a sus componentes covariantes, dentro del mismo sistema de coordenadas. Esta operación puede ser escrita, usando la notación de subíndices/superíndices

$$v_i = M_{ij} v^j .\tag{5.62}$$

Notemos cómo la convención de sumas continúa su trabajo. La misma operación para un sistema primado, usando una métrica primada, queda

$$M'_{ij} = \hat{g}'_i \cdot \hat{g}'_j ,\tag{5.63}$$

y puede ser escrito como

$$v'_i = M'_{ij} v'^j .\tag{5.64}$$

En resumen, las ecuaciones que gobiernan las transformaciones de las componentes contravariantes de un vector \vec{v} pueden ser escritas usando la nueva notación de subíndices/superíndices

$$\begin{array}{cc} t^i_j = \partial x'^i / \partial x^j & g^i_j = \partial x^i / \partial x'^j \\ v'_i = t^i_j v^j & v^i = g^i_j v'^j \\ \hat{g}_j = t^i_j \hat{g}'_i & \hat{g}'_j = g^i_j \hat{g}_i .\end{array}$$

Las componentes covariantes de \vec{v} puede ser obtenida desde las componentes contravariantes usando la métrica

$$\begin{array}{cc} M_{ij} = \hat{g}_i \cdot \hat{g}_j & v_i = M_{ij} v^j \\ M'_{ij} = \hat{g}'_i \cdot \hat{g}'_j & v'_i = M'_{ij} v'^j .\end{array}$$

Claramente hay algunos hoyos en estos cuadros. Primero, está la pregunta de cómo las componentes covariantes de un vector transforman. Segundo, decimos que los vectores base \hat{g}_i son covariantes por naturaleza. Necesitamos probar esto. Finalmente, ¿podemos definir bases vectoriales contravariantes? Estos tópicos están íntimamente relacionados unos con los otros, y apuntamos a esa dirección.

5.2.5. Transformaciones de componentes vectoriales covariantes.

Retornemos al par de sistemas coordenados inclinados descritos en la figura 5.4. Las componentes vectoriales covariantes del vector \vec{v} transformarán de acuerdo a alguna relación lineal

$$v'_i = [?] v_j . \quad (5.65)$$

Para determinar $[?]$ de esta expresión, consideremos dos formas equivalentes del producto interno de dos vectores, uno en el sistema primado y otro en el sistema no primado

$$\vec{A} \cdot \vec{B} = A^i B_i = A'^j B'_j . \quad (5.66)$$

Las componentes contravariantes de \vec{A} transforman de acuerdo a las reglas ya determinadas

$$A'^j = t^j_i A^i . \quad (5.67)$$

Sustituyendo esta expresión en el lado derecho de la ecuación (5.66), tenemos

$$A^i B_i = t^j_i A^i B'_j . \quad (5.68)$$

Como esta ecuación debe ser válida para cualquier \vec{A} , se debe tener

$$B_i = t^j_i B'_j . \quad (5.69)$$

Esta expresión puede ser fácilmente invertida, para obtener

$$B'_i = g^j_i B_j . \quad (5.70)$$

Notemos la similaridad entre las ecuaciones (5.60), (5.69) y (5.70). Esto soporta nuestra conclusión que la base del eje paralelo es covariante.

Ahora somos capaces de combinar las componentes de los vectores base contravariantes, \hat{g}^i , las cuales pueden ser determinados con la componente del vector covariante para formar el mismo vector. Esto es,

$$\vec{v} = v^i \hat{g}_i . \quad (5.71)$$

Sería mas bonito construir un nuevo conjunto de vectores base contravariantes, \hat{g}^i , los cuales pudiesen ser combinados con las componentes covariantes para formar el mismo vector. Esto es,

$$\vec{v} = v_i \hat{g}^i . \quad (5.72)$$

De hecho, podemos usar esta expresión para definir las bases de vectores contravariantes, y ver las consecuencias.

Las propiedades básicas de las bases de vectores contravariantes pueden ser deducidas considerando nuevamente el producto interno entre dos vectores, \vec{A} y \vec{B} . Si \vec{A} es expresado usando la base de vectores covariantes y las componentes contravariantes, mientras \vec{B} es escrito con vectores base contravariantes y componentes vectoriales covariantes, el producto interno se convierte en

$$\begin{aligned}\vec{A} \cdot \vec{B} &= A^i \hat{g}_i \cdot B_j \hat{g}^j \\ &= A^i B_j \hat{g}_i \cdot \hat{g}^j .\end{aligned}\tag{5.73}$$

De acuerdo a la ecuación (5.51), esta expresión debe ser igual a $A^i B_i$, por tanto

$$\hat{g}_i \cdot \hat{g}^j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},\tag{5.74}$$

o en términos de la delta de Kronecker,

$$\hat{g}_i \cdot \hat{g}^j = \delta_i^j .\tag{5.75}$$

Esta última condición permite determinar tanto la magnitud como la dirección de la base de vectores contravariante, si la base de vectores covariante es conocida. Trabajando en dos dimensiones, $\hat{g}^1 \cdot \hat{g}_2 = 0$ y $\hat{g}^1 \cdot \hat{g}_1 = 1$. En palabras, \hat{g}^1 debe ser perpendicular a \hat{g}_2 , mientras que su proyección a lo largo del eje 1, paralelo a \hat{g}_1 , debe ser uno. Esta unicidad determina \hat{g}^1 y, por argumentos similares, \hat{g}_2 . Las condiciones de la ecuación (5.75) pueden ser vistas gráficamente como es mostrado en la figura 5.5. Las construcciones en esta figura fueron hechas suponiendo que $|\hat{g}_i| = 1$.

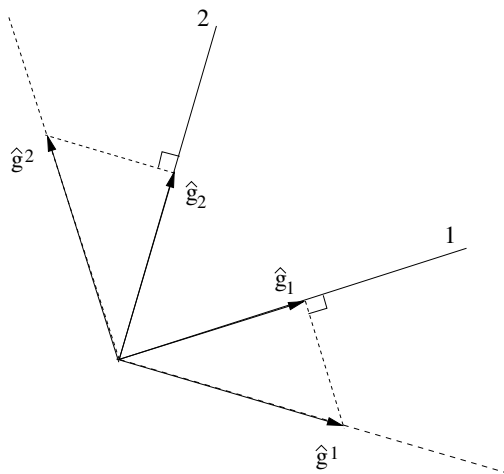


Figura 5.5: Determinación de la base de vectores contravariante.

Las componentes de los vectores covariantes y contravariantes también pueden ser interpretadas gráficamente, como es mostrado en la figura 5.6. Nuevamente, en esta figura se

ha supuesto que $|\hat{g}_i| = 1$. Las componentes de los vectores contravariantes son simplemente las magnitudes de las proyecciones paralelas a los ejes del vector sobre los ejes inclinados definidos por la base covariante de vectores. Las componentes covariantes son las magnitudes de las proyecciones del vector en el mismo eje coordenado, pero siguiendo las líneas paralelas a las nuevas base de vectores contravariante. Esto hace que las líneas de proyección para las componentes vectoriales covariantes perpendiculares a los ejes, como es mostrado en la figura. La geometría asegura que

$$\vec{v} = v^i \hat{g}_i = v_i \hat{g}^i . \quad (5.76)$$

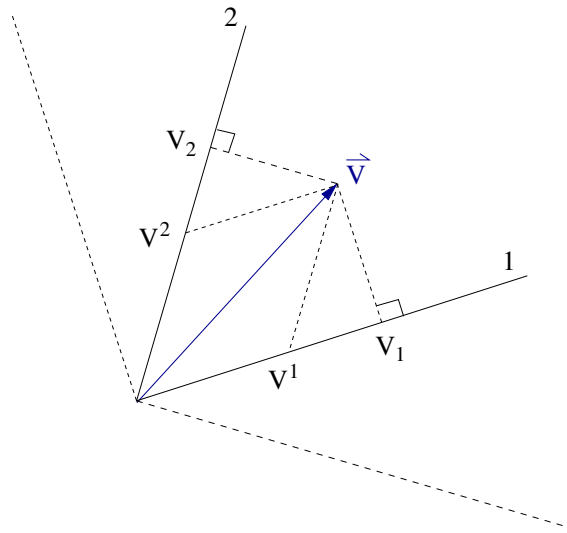


Figura 5.6: Componentes covariantes y contravariantes proyectadas de un vector.

Si la base vectorial covariante no son vectores unitarios, las construcciones de las figuras 5.5 y 5.6 deben ser ajustadas apropiadamente, siguiendo los requerimientos de las ecuaciones (5.75) y (5.76).

Las transformaciones para la base vectorial contravariante se siguen directamente de la ecuación (5.76) usando las técnicas que hemos aplicado ya varias veces

$$\hat{g}'^i = t^i_j \hat{g}^j \quad (5.77)$$

$$\hat{g}^i = g^i_j \hat{g}'^j . \quad (5.78)$$

Esto confirma nuestra clasificación de esta nueva base vectorial como contravariante, ya que transforma exactamente como las componentes contravariantes de un vector.

El conjunto completo de reglas de transformación para componentes vectoriales contravariantes y covariantes pueden ser resumidas por el conjunto simétrico de relaciones

$$\begin{array}{ll}
 v'^i = t^i_j v^j & \hat{g}'^i = t^i_j \hat{g}^j \\
 v^i = g^i_j v'^j & \hat{g}^i = g^i_j \hat{g}'^j \\
 \\
 v'_i = g^j_i v_j & \hat{g}'_i = g^j_i \hat{g}_j \\
 v_i = t^j_i v'_j & \hat{g}_i = t^j_i \hat{g}'_j
 \end{array}$$

con

$$t^i_j = \partial x'^i / \partial x^j \quad g^i_j = \partial x^i / \partial x'^j .$$

Notemos que las cantidades contravariantes siempre transforman por una suma sobre el segundo índice tanto de t^i_j y g^i_j , mientras las cantidades covariantes transforman sobre el primer índice. Para cantidades contravariantes, t^i_j es usado para ir desde el sistema no primado al sistema primado, mientras que g^i_j es usado para ir desde el sistema primado al sistema no primado. Para cantidades covariantes, los roles de t^i_j y g^i_j son los contrarios.

La nueva base de vectores contravariantes nos permiten construir otra versión del tensor métrico, esta vez con superíndices

$$M^{ij} = \hat{g}^i \cdot \hat{g}^j . \tag{5.79}$$

La aplicación de esta forma de la métrica convierte cantidades covariantes en cantidades contravariantes. Por ejemplo,

$$v^i = M^{ij} v_j . \tag{5.80}$$

Veremos en la próxima sección que las dos métricas distintas, M_{ij} y M^{ij} son simplemente representaciones del mismo objeto, el tensor métrico.

5.2.6. Covarianza y contravarianza en tensores.

Las propiedades covariantes y contravariantes discutidas en la sección anterior pueden ser fácilmente extendidas a tensores. Tal como un vector puede ser expresado con componentes contravariantes o covariantes,

$$\vec{v} = v^i \hat{g}_i = v_i \hat{g}^i , \tag{5.81}$$

un tensor puede ser expresado usando solamente componentes contravariantes o covariantes

$$\vec{T} = T^{ijk} \hat{g}_i \hat{g}_j \hat{g}_k = T_{ijk} \hat{g}^i \hat{g}^j \hat{g}^k . \tag{5.82}$$

Sin embargo, los tensores de más alto rango son más flexibles que los vectores, ya que pueden ser expresados en una forma mixta, con índices contravariantes y covariantes. Por ejemplo,

$$\vec{T} = T^i_j{}^k \hat{g}_i \hat{g}^j \hat{g}_k \tag{5.83}$$

la cual es una representación equivalente de \vec{T} .

Todas las expresiones tensoriales de las ecuaciones (5.82) y (5.83) son equivalentes, aunque los valores específicos de las componentes serán diferentes en cada caso. Así como las componentes covariantes y contravariantes de un vector están relacionadas con la métrica, las diferentes representaciones de \vec{T} pueden ser obtenidas desde una hacia otra usando la misma métrica. Por ejemplo, si las dos expresiones para \vec{T} en la ecuación (5.82) son iguales, podemos escribir

$$T^{ijk} = M^{il} M^{jm} M^{kn} T_{lmn} . \quad (5.84)$$

La expresión en la ecuación (5.83) nos arroja el mismo tensor cuando

$$T_j{}^i{}^k = M_{jm} T^{imk} . \quad (5.85)$$

Para convertir un conjunto de componentes tensoriales desde la forma puramente covariante a la forma puramente contravariante, es necesaria operación de métrica para cada índice.

Las transformaciones de sistemas de coordenadas para tensores siguen las mismas conductas que establecimos para las transformaciones vectoriales. Una matriz de transformación del tipo apropiado es usada para cada índice. Por ejemplo, podríamos escribir

$$T'{}^j{}_{i k} = g^l{}_i t^j{}_m g^n{}_k T_l{}^m{}_n \quad (5.86)$$

$$T_i{}^j{}^k = t^l{}_i g^j{}_m t^n{}_k T_l{}^m{}_n \quad (5.87)$$

Ejemplo

Hemos dicho que el tensor métrico es un tensor, pero no lo hemos probado. La demostración es directa. Consideremos los elementos de la métrica, expresado en forma covariante pura,

$$M_{ij} = \hat{g}_i \cdot \hat{g}_j . \quad (5.88)$$

El producto interno entre dos vectores, expresados en dos sistemas de coordenadas distintos, puede ser escrito como

$$\vec{A} \cdot \vec{B} = A^i B^j M_{ij} = A'^m B'^n M'_{mn} , \quad (5.89)$$

donde $M'_{mn} = \hat{g}'_m \cdot \hat{g}'_n$. Las ecuaciones de transformación pueden ser usadas para expresar las componentes del vector en el sistema primado en términos de las componentes no primados. Esto resulta ser,

$$A^i B^j M_{ij} = A^i B^j t^m{}_i t^n{}_j M'_{mn} . \quad (5.90)$$

Como la expresión debe ser válida para cualquier \vec{A} y \vec{B} , tenemos

$$M_{ij} = t^m{}_i t^n{}_j M'_{mn} , \quad (5.91)$$

lo cual es fácilmente invertido, obteniendo

$$M'_{ij} = g^m{}_i g^n{}_j M_{mn} . \quad (5.92)$$

Pero esto es exactamente como transforman los elementos de un tensor de segundo rango. Por tanto, el tensor métrico es por definición un tensor. Esto implica que podemos escribir

$$\vec{M} = M_{ij} \hat{g}^i \hat{g}^j . \quad (5.93)$$

Como la métrica es un tensor, podemos modificar su naturaleza covariante o contravariante como lo haríamos para cualquier tensor. Aunque puede parecer un poco extraña utilizar la misma métrica para modificarla, podemos cambiar una métrica puramente covariante a una métrica puramente contravariante aplicando la métrica dos veces

$$M^{ij} = M^{im} M^{jn} M_{mn} . \quad (5.94)$$

También podemos escribir la métrica en una forma mezclada escribiendo

$$M^i{}_j = M^{im} M_{mj} . \quad (5.95)$$

Usando las ecuaciones de transformación, se puede demostrar fácilmente que

$$M^i{}_j = \hat{g}^i \cdot \hat{g}_j = \delta^i{}_j . \quad (5.96)$$

Esto implica que el tensor métrico es realmente sólo una generalización del tensor Delta de Kronecker.

5.2.7. Contravarianza y covarianza de derivadas parciales.

Cuando las derivadas parciales son tomadas con respecto a las coordenadas contravariantes, el resultado es una cantidad covariante. Para ver esto, sean x^i y x'^i las coordenadas contravariantes en un par arbitrario de sistemas de coordenadas. Las reglas del cálculo requieren que se cumpla

$$\frac{\partial}{\partial x'^i} = \frac{\partial x^j}{\partial x'^i} \frac{\partial}{\partial x^j} , \quad (5.97)$$

donde hay una suma implícita sobre el índice j . Pero notemos que el término $\partial x^j / \partial x'^i$ es exactamente la definición de $g^j{}_i$. Esto nos permite escribir

$$\frac{\partial}{\partial x'^i} = g^j{}_i \frac{\partial}{\partial x^j} . \quad (5.98)$$

Comparando esta expresión con la ecuación (5.70) vemos que la operación derivada parcial transforma como una cantidad covariante. En ese caso, encontramos que

$$\frac{\partial}{\partial x'_i} = \frac{\partial x_j}{\partial x'_i} \frac{\partial}{\partial x_j} \quad (5.99)$$

$$= t^j{}_i \frac{\partial}{\partial x_j} , \quad (5.100)$$

lo cual nos dice que esta derivada parcial actúa como una cantidad contravariante. Para ser consistentes con nuestra notación, imponemos la regla que un superíndice en el “denominador” de la operación derivada actúa como un subíndice, mientras que un subíndice en el denominador actúa como un superíndice. Esta idea fue discutida brevemente en la conexión con las matrices de transformación en las ecuaciones (5.56) y (5.58).

Ejemplo

Un campo eléctrico estático es calculado usualmente tomando el gradiente de un potencial escalar

$$\vec{E} = -\vec{\nabla}\varphi . \quad (5.101)$$

En un capítulo anterior, definimos el operador gradiente por la relación

$$d\varphi = \vec{\nabla}\varphi \cdot d\vec{r} . \quad (5.102)$$

Como el vector desplazamiento puede ser descrito por

$$d\vec{r} = dx^i \hat{g}_i , \quad (5.103)$$

donde dx^i es una cantidad contravariante, es claro que el gradiente de φ puede ser escrito por

$$\vec{\nabla}\varphi = \frac{\partial\varphi}{\partial x^i} \hat{g}^i . \quad (5.104)$$

Podemos chequear la validez de esta expresión, reemplazando las ecuaciones (5.104) y (5.103) en el lado derecho de la ecuación (5.102), de donde obtenemos

$$\begin{aligned} \vec{\nabla}\varphi \cdot d\vec{r} &= \frac{\partial\varphi}{\partial x^i} \hat{g}^i \cdot dx^j \hat{g}_j \\ &= \frac{\partial\varphi}{\partial x^i} dx^j \delta^i_j \end{aligned} \quad (5.105)$$

$$\begin{aligned} &= \frac{\partial\varphi}{\partial x^i} dx^i \\ &= d\varphi \end{aligned} \quad (5.106)$$

Luego, escribimos las componentes del campo eléctrico por

$$E_i = -\frac{\partial\varphi}{\partial x^i} , \quad (5.107)$$

las cuales son covariantes y deben transformar según la relación

$$E'_i = g^j_i E_j . \quad (5.108)$$

Otro ejemplo

Un campo magnético estático \vec{B} puede ser calculado usando la ley de Àmpere

$$\oint_C d\vec{r} \cdot \vec{B} = \frac{4\pi}{c} I . \quad (5.109)$$

En esta expresión, I es la corriente total que fluye a través del camino cerrado C . Tomando $d\vec{r}$ como el vector diferencial con componentes contravariantes, como está dado en la ecuación (5.103), las componentes del campo magnético usadas en esta integración deben ser escritas en forma covariante, por tanto

$$\oint_C dx^i B_i = \frac{4\pi}{c} I . \quad (5.110)$$

Capítulo 6

Determinantes y matrices.

versión final 1.31-160401¹

6.1. Determinantes.

Comenzamos el estudio de matrices resolviendo ecuaciones lineales las cuales nos llevan a determinantes y matrices. El concepto de determinante y su notación fueron introducidos por Leibniz.

Ecuaciones lineales homogéneas.

Una de las mayores aplicaciones de los determinantes está en el establecimiento de una condición para la existencia de una solución no trivial de un conjunto de ecuaciones algebraicas lineales homogéneas. Supongamos que tenemos tres incógnitas x_1, x_2, x_3 (o n ecuaciones con n incógnitas).

$$\begin{aligned}a_1x_1 + a_2x_2 + a_3x_3 &= 0 , \\b_1x_1 + b_2x_2 + b_3x_3 &= 0 , \\c_1x_1 + c_2x_2 + c_3x_3 &= 0 .\end{aligned}\tag{6.1}$$

El problema es: ¿en qué condiciones hay alguna solución, aparte de la solución trivial $x_1 = 0, x_2 = 0, x_3 = 0$? Si usamos notación vectorial $\vec{x} = (x_1, x_2, x_3)$ para la solución y tres filas $\vec{a} = (a_1, a_2, a_3), \vec{b} = (b_1, b_2, b_3), \vec{c} = (c_1, c_2, c_3)$ para los coeficientes, tenemos que las tres ecuaciones, ecuación (6.1), se convierten en

$$\vec{a} \cdot \vec{x} = 0 , \quad \vec{b} \cdot \vec{x} = 0 , \quad \vec{c} \cdot \vec{x} = 0 .\tag{6.2}$$

Estas tres ecuaciones vectoriales tienen la interpretación geométrica obvia que \vec{x} es ortogonal a $\vec{a}, \vec{b}, \vec{c}$. Si el volumen sustentado por $\vec{a}, \vec{b}, \vec{c}$ dado por el determinante (o el producto escalar triple)

$$D_3 = (\vec{a} \times \vec{b}) \cdot \vec{c} = \det(\vec{a}, \vec{b}, \vec{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} ,\tag{6.3}$$

¹Este capítulo está basado en el tercer capítulo del libro: *Mathematical Methods for Physicists, fourth edition* de George B. Arfken & Hans J. Weber, editorial ACADEMIC PRESS.

no es cero, claramente sólo existe la solución trivial $\vec{x} = 0$.

Vice-versa, si el anterior determinante de coeficientes se anula, luego uno de los vectores columna es una combinación lineal de otros dos. Supongamos que \vec{c} está en el plano que sustenta \vec{a} , \vec{b} , *i.e.*, la tercera ecuación es una combinación lineal de las primeras dos y no es independiente. Luego \vec{x} es ortogonal a ese plano tal que $\vec{x} \sim \vec{a} \times \vec{b}$. Ya que las ecuaciones homogéneas pueden ser multiplicadas por números arbitrarios, solamente las relaciones de x_i son relevantes, para lo cual obtenemos razones de determinantes de 2×2

$$\begin{aligned} \frac{x_1}{x_3} &= \frac{(a_2 b_3 - a_3 b_2)}{(a_1 b_2 - a_2 b_1)}, \\ \frac{x_2}{x_3} &= -\frac{(a_1 b_3 - a_3 b_1)}{(a_1 b_2 - a_2 b_1)}, \end{aligned} \quad (6.4)$$

a partir de los componentes del producto cruz $\vec{a} \times \vec{b}$.

Ecuaciones lineales no homogéneas.

El caso más simple es de dos ecuaciones con dos incógnitas

$$\begin{aligned} a_1 x_1 + a_2 x_2 &= a_3, \\ b_1 x_1 + b_2 x_2 &= b_3, \end{aligned} \quad (6.5)$$

puede ser reducido al caso previo embebiéndolo en un espacio tridimensional con una solución vectorial $\vec{x} = (x_1, x_2, -1)$ y el vector fila $\vec{a} = (a_1, a_2, a_3)$, $\vec{b} = (b_1, b_2, b_3)$. Como antes, ecuación (6.5) en notación vectorial, $\vec{a} \cdot \vec{x} = 0$ y $\vec{b} \cdot \vec{x} = 0$, implica que $\vec{x} \sim \vec{a} \times \vec{b}$ tal que el análogo de la ecuación (6.4) se mantiene. Para que esto se aplique la tercera componente de $\vec{a} \times \vec{b}$ debiera ser distinta de cero, *i.e.*, $a_1 b_2 - a_2 b_1 \neq 0$, ya que la tercera componente de \vec{x} es $-1 \neq 0$. Esto produce que los x_i tengan la forma

$$x_1 = \frac{(a_3 b_2 - a_2 b_3)}{(a_1 b_2 - a_2 b_1)} = \frac{\begin{vmatrix} a_3 & a_2 \\ b_3 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}} \quad (6.6a)$$

$$x_2 = \frac{(a_1 b_3 - a_3 b_1)}{(a_1 b_2 - a_2 b_1)} = \frac{\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}}. \quad (6.6b)$$

El determinante en el numerador de $x_1(x_2)$ es obtenido a partir del determinante de los coeficientes $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$ reemplazando el primer vector columna (segundo) por el vector $\begin{pmatrix} a_3 \\ b_3 \end{pmatrix}$ del lado inhomogéneo de la ecuación (6.5).

Estas soluciones de ecuación lineal en términos de determinantes pueden ser generalizados a n dimensiones. El determinante es un arreglo cuadrado

$$D_n = \begin{vmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \\ c_1 & c_2 & \dots & c_n \\ \cdot & \cdot & \dots & \cdot \end{vmatrix}, \quad (6.7)$$

de números (o funciones), los coeficientes de n ecuaciones lineales en nuestro caso. El número n de columnas (y de filas) en el arreglo es llamado algunas veces el orden del determinante. La generalización de la expansión del producto escalar triple (de vectores fila de las tres ecuaciones lineales) tiende al siguiente valor del determinante D_n en n dimensiones,

$$D_n = \sum_{i,j,k,\dots} \varepsilon_{ijk\dots} a_i b_j c_k \dots, \quad (6.8)$$

donde $\varepsilon_{ijk\dots}$, análogo al símbolo de Levi-Civita de la ecuación (1.52), es +1 para permutaciones pares ($ijk\dots$) de $(123\dots n)$, -1 para permutaciones impares, y cero si algún índice es repetido.

Específicamente, para el determinante de orden tres D_3 de las ecuaciones (6.3) y (6.8) tenemos

$$D_3 = +a_1 b_2 c_3 - a_1 b_3 c_2 - a_2 b_1 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_3 b_2 c_1. \quad (6.9)$$

El determinante de orden tres, entonces, es esta particular combinación lineal de productos. Cada producto contiene uno y sólo un elemento de cada fila y de cada columna. Cada producto es sumado si las columnas (los índices) representan una permutación par de (123) y restando si corresponde a una permutación impar. La ecuación (6.3) puede ser considerada en notación abreviada de la ecuación (6.9). El número de términos en la suma (ecuación (6.8)) es 24 para un determinante de cuarto orden, en general $n!$ para un determinante de orden n . A causa de la aparición de signos negativos en la ecuación (6.9) pueden haber cancelaciones. Debido a ésto es muy posible que un determinante de elementos grandes tenga un valor pequeño.

Algunas propiedades útiles de los determinantes de n -ésimo orden siguen de la ecuación (6.8). De nuevo, para ser específico, la ecuación (6.9) para determinantes de orden tres es usada para ilustrar estas propiedades.

Desarrollo laplaciano por las menores.

La ecuación (6.9) puede ser reescrita

$$\begin{aligned} D_3 &= a_1(b_2 c_3 - b_3 c_2) - a_2(b_1 c_3 - b_3 c_1) + a_3(b_1 c_2 - b_2 c_1) \\ &= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}. \end{aligned} \quad (6.10)$$

En general, el determinante de orden n -ésimo puede ser expandido como una combinación lineal de productos de elementos de alguna fila (o columna) por determinantes de orden $(n-1)$ formados suprimiendo la fila y la columna del determinante original en el cual aparece el elemento. Este arreglo reducido (2×2 en el ejemplo específico) es llamado una menor. Si el elemento está en la i -ésima fila y en la j -ésima columna, el signo asociado con el producto es $(-1)^{i+j}$. La menor con este signo es llamada el cofactor. Si M_{ij} es usado para designar la menor formado omitiendo la fila i y la columna j y c_{ij} es el cofactor correspondiente, la ecuación (6.10) se convierte en

$$D_3 = \sum_{j=1}^3 (-1)^{j+1} a_j M_{1j} = \sum_{j=1}^3 a_j c_{1j}. \quad (6.11)$$

En este caso, expandiendo a lo largo de la primera fila, tenemos $i = 1$ y la suma es sobre j , las columnas.

Esta expansión de Laplace puede ser usada para sacar ventaja en la evaluación de determinantes de alto orden en el cual muchos de los elementos son nulos. Por ejemplo, para encontrar el valor de el determinante

$$D = \begin{vmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{vmatrix}, \quad (6.12)$$

expandimos a través de la fila superior para obtener

$$D = (-1)^{1+2} \cdot (1) \begin{vmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{vmatrix}. \quad (6.13)$$

Nuevamente, expandimos a través de la fila superior para obtener

$$\begin{aligned} D &= (-1) \cdot (-1)^{1+1} \cdot (-1) \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} \\ &= \begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix} = 1. \end{aligned} \quad (6.14)$$

Este determinante D (ecuación (6.12)) está formado de una de las matrices de Dirac que aparecen en la teoría relativista del electrón de Dirac.

Antisimetría.

El determinante cambia de signo si cualquier par de filas son intercambiadas o si cualquier par de columnas son intercambiadas. Esto deriva del carácter par-impar del Levi-Civita ε en la ecuación (6.8) o explícitamente de la forma de las ecuaciones (6.9) y (6.10).

Esta propiedad es frecuentemente usada en Mecánica Cuántica para la construcción de una función de onda de muchas partículas que, en concordancia con el principio de exclusión de Pauli, será antisimétrica bajo el intercambio de cualquier par de partículas idénticas con spin 1/2 (electrones, protones, neutrones, etc).

Como un caso especial de antisimetría, cualquier determinante con dos filas iguales o dos columnas iguales es nulo.

Si cada elemento en una fila o de una columna es cero el determinante completo es nulo.

Si cada elemento en una fila o de una columna es multiplicado por una constante, el determinante completo es multiplicado por esa constante.

El valor de un determinante es inalterado si un múltiplo de una fila es añadido (columna por columna) a otra fila o si un múltiplo de una columna es añadido (fila por fila) a otra columna. Tenemos

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 + ka_2 & a_2 & a_3 \\ b_1 + kb_2 & b_2 & b_3 \\ c_1 + kc_2 & c_2 & c_3 \end{vmatrix}. \quad (6.15)$$

Usando el desarrollo de Laplace sobre el lado derecho, obtenemos

$$\begin{vmatrix} a_1 + ka_2 & a_2 & a_3 \\ b_1 + kb_2 & b_2 & b_3 \\ c_1 + kc_2 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} + k \begin{vmatrix} a_2 & a_2 & a_3 \\ b_2 & b_2 & b_3 \\ c_2 & c_2 & c_3 \end{vmatrix}, \quad (6.16)$$

entonces por la propiedad de antisimetría el segundo determinante del lado derecho se anula, verificando la ecuación (6.15).

Un caso especial, un determinante es igual a cero, si cualquier par de filas o columnas son proporcionales.

Volviendo a las ecuaciones homogéneas (6.1) y multiplicando el determinante de los coeficientes por x_1 , y luego sumando x_2 veces la segunda columna y x_3 veces la tercera columna, podemos establecer directamente la condición para la presencia de una solución no trivial para la ecuación (6.1):

$$x_1 \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} x_1 a_1 & a_2 & a_3 \\ x_1 b_1 & b_2 & b_3 \\ x_1 c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 x_1 + a_2 x_2 + a_3 x_3 & a_2 & a_3 \\ b_1 x_1 + b_2 x_2 + b_3 x_3 & b_2 & b_3 \\ c_1 x_1 + c_2 x_2 + c_3 x_3 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} 0 & a_2 & a_3 \\ 0 & b_2 & b_3 \\ 0 & c_2 & c_3 \end{vmatrix} = 0. \quad (6.17)$$

Por lo tanto x_1 (x_2 y x_3) deberían ser cero a menos que el determinante de los coeficientes sea nulo. Podemos mostrar que si el determinante de los coeficientes es nulo, existe realmente una solución no trivial.

Si nuestras ecuaciones lineales son inhomogéneas, esto es, como en la ecuación (6.5) o si los ceros en el lado derecho de la ecuación (6.1) fueran reemplazados por a_4 , b_4 , c_4 respectivamente, luego de la ecuación (6.17) obtenemos,

$$x_1 = \frac{\begin{vmatrix} a_4 & a_2 & a_3 \\ b_4 & b_2 & b_3 \\ c_4 & c_2 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}}, \quad (6.18)$$

la cual generaliza la ecuación (6.6a) a la dimensión $n = 3$. Si el determinante de los coeficientes se anula, el conjunto de ecuaciones no homogéneas no tiene solución a menos que el numerador también se anule. En este caso las soluciones pueden existir pero ellas no son únicas.

Para el trabajo numérico, esta solución del determinante, ecuación (6.18), es enormemente difícil de manejar. El determinante puede involucrar grandes números con signos alternados, y en la resta de dos números grandes el error relativo podría remontarse al punto que hace que el resultado no tenga valor. También, aunque el método del determinante es ilustrado aquí con tres ecuaciones y tres incógnitas, podríamos fácilmente tener 200 ecuaciones con 200 incógnitas las cuales, involucran sobre 200! términos por determinante, lo que pone un desafío muy alto a la velocidad computacional. Debería haber una mejor manera. En efecto, hay una mejor manera. Una de las mejores es un proceso a menudo llamado eliminación de Gauss. Para ilustrar esta técnica, consideremos el siguiente conjunto de ecuaciones.

Resolvamos

$$\begin{aligned} 3x + 2y + z &= 11 \\ 2x + 3y + z &= 13 \\ x + y + 4z &= 12 . \end{aligned} \tag{6.19}$$

El determinante de la ecuación lineal no homogénea ecuación (6.19) es 18, por lo tanto existe una solución.

Por conveniencia y para una óptima precisión numérica, las ecuaciones son reordenadas tal que los coeficientes mayores corran a lo largo de la diagonal principal (superior izquierda a inferior derecha). Esto ha sido hecho en el conjunto anterior.

La técnica de Gauss es usar la primera ecuación para eliminar la primera incógnita x de las ecuaciones restantes. Entonces la (nueva) segunda ecuación es usada para eliminar y de la última ecuación. En general, descendemos poco a poco a través del conjunto de ecuaciones, y luego, con una incógnita determinada, avanzamos gradualmente para resolver cada una de las otras incógnitas en sucesión.

Dividiendo cada fila por su coeficiente inicial, vemos que las ecuaciones (6.19) se convierten en

$$\begin{aligned} x + \frac{2}{3}y + \frac{1}{3}z &= \frac{11}{3} \\ x + \frac{3}{2}y + \frac{1}{2}z &= \frac{13}{2} \\ x + y + 4z &= 12 . \end{aligned} \tag{6.20}$$

Ahora, usando la primera ecuación, eliminamos x de la segunda y la tercera:

$$\begin{aligned} x + \frac{2}{3}y + \frac{1}{3}z &= \frac{11}{3} \\ \frac{5}{6}y + \frac{1}{6}z &= \frac{17}{6} \\ \frac{1}{3}y + \frac{11}{3}z &= \frac{25}{3} , \end{aligned} \tag{6.21}$$

y

$$\begin{aligned} x + \frac{2}{3}y + \frac{1}{3}z &= \frac{11}{3} \\ y + \frac{1}{5}z &= \frac{17}{5} \\ y + 11z &= 25 . \end{aligned} \tag{6.22}$$

Repitiendo la técnica, usamos la segunda ecuación para eliminar y a partir de la tercera ecuación:

$$\begin{aligned} x + \frac{2}{3}y + \frac{1}{3}z &= \frac{11}{3} \\ y + \frac{1}{5}z &= \frac{17}{5} \\ 54z &= 108 , \end{aligned} \tag{6.23}$$

o

$$z = 2 .$$

Finalmente, al reemplazar obtenemos

$$y + \frac{1}{5} \times 2 = \frac{17}{5} ,$$

o

$$y = 3 .$$

Luego con z e y determinados,

$$x + \frac{2}{3} \times 3 + \frac{1}{3} \times 2 = \frac{11}{3} ,$$

y

$$x = 1 .$$

La técnica podría parecer no tan elegante como la ecuación (6.17), pero está bien adaptada a los computadores modernos y es más rápida que el tiempo gastado con los determinantes.

Esta técnica de Gauss puede ser usada para convertir un determinante en una forma triangular:

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & 0 & c_3 \end{vmatrix} ,$$

para un determinante de tercer orden cuyos elementos no deben ser confundidos con aquellos en la ecuación (6.3). De esta forma $D = a_1 b_2 c_3$. Para un determinante de n -ésimo orden la evaluación de una forma triangular requiere solamente $n - 1$ multiplicaciones comparadas con las $n!$ requeridas para el caso general.

Una variación de esta eliminación progresiva es conocida como eliminación de Gauss-Jordan. Comenzamos como si fuera el procedimiento de Gauss, pero cada nueva ecuación considerada es usada para eliminar una variable de todas las “otras” ecuaciones, no sólo de aquellas bajo ella. Si hemos usado esta eliminación de Gauss-Jordan, la ecuación (6.23) llegaría a ser

$$\begin{aligned} x + \frac{1}{5}z &= \frac{7}{5} \\ y + \frac{1}{5}z &= \frac{17}{5} \\ z &= 2 , \end{aligned} \tag{6.24}$$

usando la segunda ecuación de la ecuación (6.22) para eliminar y de ambas, la primera y tercera ecuaciones. Entonces la tercera ecuación de la ecuación (6.24) es usada para eliminar z de la primera y segunda ecuaciones, dando

$$\begin{aligned} x &= 1 \\ y &= 3 \\ z &= 2 , \end{aligned} \tag{6.25}$$

Volveremos a la técnica de Gauss-Jordan cuando invertamos matrices.

Otra técnica disponible para el uso computacional es la técnica de Gauss-Seidel. Cada técnica tiene sus ventajas y desventajas. Los métodos de Gauss y Gauss-Jordan pueden tener problemas de precisión para un determinante grande. Esto también es un problema para la inversión de matrices. El método de Gauss-Seidel, como un método iterativo, puede tener problemas de convergencia.

6.2. Matrices.

El análisis matricial pertenece al álgebra lineal ya que las matrices son operadores o mapas lineales tales como rotaciones. Supongamos, por ejemplo, que rotamos las coordenadas cartesianas de un espacio bidimensional tal que, en notación vectorial,

$$\begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \begin{pmatrix} x_1 \cos \varphi & x_2 \operatorname{sen} \varphi \\ -x_2 \sin \varphi & x_1 \cos \varphi \end{pmatrix} = (\sum_j a_{ij} x_j) . \quad (6.26)$$

Etiquetamos el arreglo de elementos a_{ij} por la matriz \mathbf{A} de 2×2 consistente de dos filas y dos columnas, además, consideramos los vectores x , x' como matrices de 2×1 . Tomemos la suma de productos de la ecuación (6.26) como una definición de la multiplicación matricial que involucra el producto escalar de cada uno de los vectores fila de \mathbf{A} con el vector columna x . Así en notación matricial la ecuación (6.26) se convierte en

$$x' = \mathbf{A}x . \quad (6.27)$$

Para extender esta definición de multiplicación de una matriz por un vector columna a el producto de dos matrices de 2×2 , consideremos la rotación de coordenada seguida por una segunda rotación dada por la matriz \mathbf{B} tal que

$$x'' = \mathbf{B}x' . \quad (6.28)$$

Por componentes

$$x''_i = \sum_j b_{ij} x'_j = \sum_j b_{ij} \sum_k a_{jk} x_k = \sum_k \left(\sum_j b_{ij} a_{jk} \right) x_k . \quad (6.29)$$

La suma sobre j es la multiplicación matricial definiendo una matriz $\mathbf{C} = \mathbf{B}\mathbf{A}$ tal que

$$x''_i = \sum_k c_{ik} x_k , \quad (6.30)$$

o $x'' = \mathbf{C}x$ en notación matricial. Nuevamente, esta definición involucra el producto escalar de vectores filas de \mathbf{B} con vectores columnas de \mathbf{A} . Esta definición de multiplicación matricial se puede generalizar a matrices de $m \times n$ y es útil, realmente “su utilidad es la justificación de su existencia”. La interpretación física es que el producto matricial de dos matrices, $\mathbf{B}\mathbf{A}$, es la rotación que conduce del sistema sin prima directamente al sistema de coordenadas con doble prima. Antes de pasar a la definición formal, podemos notar que el operador \mathbf{A} está descrito

por sus efectos sobre las coordenadas o vectores base. Los elementos de matriz a_{ij} constituyen una representación del operador, una representación que depende de la elección de una base.

El caso especial donde una matriz tiene una columna y n filas es llamada un vector columna, $|x\rangle$, con componentes x_i , $i = 1, 2, \dots, n$. Si A es una matriz de $n \times n$, $A|x\rangle$ es un vector columna de n componentes, $A|x\rangle$ está definida como en la ecuación (6.27) y (6.26). Similarmente, si una matriz tiene una fila y n columnas, es llamada un vector fila, $\langle x|$ con componentes x_i , $i = 1, 2, \dots, n$. Claramente, $\langle x|$ resulta de $|x\rangle$ por el intercambio de filas y columnas, una operación matricial llamada *transposición*, y para cualquier matriz A , \tilde{A} es llamada ² la transpuesta de A con elementos de matriz $(\tilde{A})_{ik} = A_{ik}$. Transponiendo un producto de matrices AB se invierte el orden y da BA ; similarmente, $A|x\rangle$ se transpone como $\langle x|A$. El producto escalar toma la forma $\langle x|y\rangle$.

Definiciones básicas.

Una matriz puede ser definida como un arreglo cuadrado o rectangular de números o funciones que obedecen ciertas leyes. Esto es una extensión perfectamente lógica de los conceptos matemáticos familiares. En aritmética tratamos con números simples. En la teoría de variable compleja tratamos con pares ordenados de números, $(1, 2) = 1 + 2i$, en el cual el orden es importante. Ahora consideremos números (o funciones) ordenados en un arreglo cuadrados o rectangular. Por conveniencia en el trabajo posterior los números son distinguidos por dos subíndices, el primero indica la fila (horizontal) y el segundo indica la columna (vertical) en la cual aparecen los números. Por ejemplo, a_{13} es el elemento de matriz en la primera fila y tercera columna. De este modo, si A es una matriz con m filas y n columnas,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \quad (6.31)$$

Quizás el hecho más importante a notar es que los elementos a_{ij} no están combinados unos con otros. Una matriz no es un determinante. Es un arreglo ordenado de números, no un simple número.

La matriz A hasta ahora de sólo es un arreglo de números que tiene las propiedades que le asignamos. Literalmente, esto significa construir una nueva forma de matemáticas. Postulamos que las matrices A , B y C , con elementos a_{ij} , b_{ij} y c_{ij} , respectivamente, combinan de acuerdo a las siguientes reglas.

Igualdad.

Matriz $A =$ Matriz B si y sólo si $a_{ij} = b_{ij}$ para todos los valores de i y j . Esto, por su puesto, requiere que A y B sean cada uno arreglos de $m \times n$ (m filas y n columnas).

²Algunos textos denotan A transpuesta por A^T .

Suma.

$A + B = C$ si y sólo si $a_{ij} + b_{ij} = c_{ij}$ para todos los valores de i y j , los elementos se combinan de acuerdo a las leyes del álgebra lineal (o aritmética si hay números simples). Esto significa que $A + B = B + A$, la conmutación. También, se satisface la ley de asociatividad $(A + B) + C = A + (B + C)$. Si todos los elementos son cero, la matriz es llamada matriz nula y se denota por 0 . Para todo A ,

$$A + 0 = 0 + A = A ,$$

con

$$0 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} . \quad (6.32)$$

Tal que las matrices de $m \times n$ forman un espacio lineal con respecto a la suma y la resta.

Multiplicación (por un escalar).

La multiplicación de la matriz A por una cantidad escalar α está definida como

$$\alpha A = (\alpha A) , \quad (6.33)$$

en la cual los elementos de αA son αa_{ij} ; esto es, cada elemento de la matriz A es multiplicado por el factor escalar. Esto contrasta con el comportamiento de los determinantes en el cual el factor α multiplica solamente una columna o una fila y no cada elemento del determinante. Una consecuencia de esta multiplicación por escalar es que

$$\alpha A = A\alpha , \quad \text{conmutación.} \quad (6.34)$$

Multiplicación (multiplicación matricial) producto interno.

$$AB = C \quad \text{si y solo si} \quad c_{ij} = \sum_k a_{ik} b_{kj} . \quad (6.35)$$

Los elementos i y j de C están formados como un producto escalar de la i -ésima fila de A con el j -ésima columna de B (el cual demanda que A tenga el mismo número de columnas como B tiene de filas). El índice mudo k toma los valores $1, 2, \dots, n$ en sucesión, esto es,

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} , \quad (6.36)$$

para $n = 3$. Obviamente, el índice mudo k puede ser reemplazado por algún otro símbolo que no esté en uso sin alterar la ecuación (6.35). Quizás la situación puede ser aclarada afirmando que la ecuación (6.35) defina el método de combinar ciertas matrices. Este método de combinación, es llamado multiplicación matricial. Para ilustrar, consideremos dos matrices (matrices de Pauli)

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \quad (6.37)$$

El elemento $_{11}$ del producto, $(\sigma_1\sigma_3)_{11}$ está dado por la suma de productos de elementos de la primera fila de σ_1 con el correspondiente elemento de la primera columna de σ_3 : Aquí

$$(\sigma_1\sigma_3)_{ij} = \sigma_{1i1}\sigma_{31j} + \sigma_{1i2}\sigma_{32j} .$$

Una aplicación directa de la multiplicación de matrices muestra que

$$\sigma_3\sigma_1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (6.38)$$

y por la ecuación (6.35)

$$\sigma_1\sigma_3 = -\sigma_1\sigma_3 . \quad (6.39)$$

Excepto en casos especiales, la multiplicación de matrices no es conmutativa.³

$$AB \neq BA . \quad (6.40)$$

Sin embargo, de la definición de multiplicación de matrices podemos mostrar que se mantiene una ley de asociatividad, $(AB)C = A(BC)$. También se satisface una ley de distributividad, $A(B + C) = AB + AC$. La matriz unidad tiene elementos δ_{ij} , la delta de Kronecker, y la propiedad de que $1A = A1 = A$ para toda A ,

$$1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} . \quad (6.41)$$

Notamos que es posible que el producto de dos matrices sea una matriz nula sin ser ninguna de ellas una matriz nula. Por ejemplo, si

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} .$$

$AB = 0$. Esto difiere de la multiplicación de números reales o complejos los cuales forman un *campo*, mientras que la estructura aditiva y multiplicativa de las matrices es llamada *anillo* por los matemáticos.

Si A es una matriz de $n \times n$ con determinante $|A| \neq 0$, luego tiene una única inversa A^{-1} tal que $AA^{-1} = A^{-1}A = 1$. Si B es también una matriz de $n \times n$ con inversa B^{-1} , luego el producto de AB tiene la inversa

$$(AB)^{-1} = B^{-1}A^{-1} , \quad (6.42)$$

ya que $ABB^{-1}A^{-1} = 1 = B^{-1}A^{-1}AB$.

El *teorema del producto* el cual dice que el determinante de un producto, $|AB|$, de dos matrices de $n \times n$ A y B es igual al producto de los determinantes, $|A||B|$, uniendo matrices con determinantes. El anterior teorema puede ser fácilmente probado.

³La pérdida de la propiedad conmutativa es descrita por el conmutador $[A, B] = AB - BA$. La no conmutatividad se expresa por $[A, B] \neq 0$.

Producto directo.

Un segundo procedimiento para multiplicar matrices, conocido como el tensor producto directo o de Kronecker. Si A es una matriz de $m \times m$ y B una matriz de $n \times n$, luego el producto directo es

$$A \otimes B = C . \quad (6.43)$$

C es una matriz de $mn \times mn$ con elementos

$$C_{\alpha\beta} = A_{ij}B_{kl} , \quad (6.44)$$

con

$$\alpha = n(i - 1) + k , \quad \beta = n(j - 1) + l .$$

Por ejemplo, si A y B ambas son matrices de 2×2 ,

$$\begin{aligned} A \otimes B &= \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix} \\ &= \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix} . \end{aligned} \quad (6.45)$$

El producto directo es asociativo pero no conmutativo. Como un ejemplo de producto directo, las matrices de Dirac pueden ser desarrolladas como productos directos de las matrices de Pauli y de la matriz unidad. Otros ejemplos aparecen en la construcción de grupos en teoría de grupos y en espacios de Hilbert en teoría cuántica.

El producto directo definido aquí es algunas veces llamado la forma *standard* y es denotado por \otimes . Otros tres tipos de producto directo de matrices existe como posibilidades o curiosidades matemáticas pero tienen muy poca o ninguna aplicación en física matemática.

Matrices diagonales.

Un tipo especial muy importante de matrices es la matriz cuadrada en la cual todos los elementos no diagonales son cero. Específicamente, si una matriz A de 3×3 es diagonal,

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} .$$

Una interpretación física de tales matrices diagonales y el método de reducir matrices a esta forma diagonal son considerados en la sección 6.5. Aquí nos limitamos a notar la importante propiedad de que la multiplicación de matrices es conmutativa, $AB = BA$, si A y B son cada una diagonales.

Traza.

En cualquiera matriz cuadrada la suma de los elementos diagonales es llamada la *traza*. Claramente la traza es una operación lineal:

$$\text{traza}(\mathbf{A} - \mathbf{B}) = \text{traza}(\mathbf{A}) - \text{traza}(\mathbf{B}) .$$

Una de sus interesantes y útiles propiedades es que la traza de un producto de dos matrices \mathbf{A} y \mathbf{B} es independiente del orden de la multiplicación:

$$\begin{aligned} \text{traza}(\mathbf{AB}) &= \sum_i (\mathbf{AB})_{ii} = \sum_i \sum_j a_{ij} b_{ji} \\ &= \sum_i \sum_j b_{ji} a_{ij} = \sum_j (\mathbf{BA})_{jj} \\ &= \text{traza}(\mathbf{BA}) . \end{aligned} \tag{6.46}$$

Esto se mantiene aún cuando $\mathbf{AB} \neq \mathbf{BA}$. La ecuación (6.46) significa que la traza de cualquier conmutador, $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$, es cero. De la ecuación (6.46) obtenemos

$$\text{traza}(\mathbf{ABC}) = \text{traza}(\mathbf{BCA}) = \text{traza}(\mathbf{CAB}) ,$$

lo cual muestra que la traza es invariante bajo permutaciones cíclicas de la matriz en un producto.

Para una matriz simétrica o una matriz Hermítica compleja la traza es la suma, y el determinante el producto, de sus autovalores, y ambos son coeficientes del polinomio característico. La traza servirá una función similar para las matrices como la ortogonalidad sirve para los vectores y funciones.

En términos de tensores la traza es una contracción y como el tensor de segundo orden contraído es un escalar (invariante).

Las matrices son usadas ampliamente para representar los elementos de grupos. La traza de las matrices representando los elementos de grupo es conocido en teoría de grupos como el *carácter*. La razón de este nombre especial y espacial atención es que mientras las matrices pueden variar la traza o carácter se mantiene invariante.

Inversión de matriz.

Al comienzo de esta sección la matriz \mathbf{A} fue presentada como la representación de un operador que (linealmente) transforma los ejes de coordenadas. Una rotación podría ser un ejemplo de tal transformación lineal. Ahora buscaremos la transformación inversa \mathbf{A}^{-1} que restablecerá los ejes de coordenadas originales. Esto significa, ya sea como una ecuación matricial o de operador⁴,

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{1} . \tag{6.47}$$

Podemos probar (ejercicio) que

$$a_{ij}^{-1} = \frac{C_{ji}}{|\mathbf{A}|} , \tag{6.48}$$

⁴Aquí y a través de todo el capítulo nuestras matrices tienen rango finito.

con la suposición que el determinante de \mathbf{A} ($|\mathbf{A}| \neq 0$). Si es cero, etiquetaremos a \mathbf{A} como singular. No existe la inversa. Como fue explicado en la sección 6.1 esta forma con determinante es *totalmente inapropiado para el trabajo numérico* con grandes matrices.

Hay una amplia variedad de técnicas alternativas. Una de las mejores y más comúnmente usada es la técnica de inversión de matrices de Gauss-Jordan. La teoría está basada en los resultados que muestran que existen matrices \mathbf{M}_L tal que el producto $\mathbf{M}_L\mathbf{A}$ será \mathbf{A} pero con

- una fila multiplicada por una constante, o
- una fila reemplazada por la fila original menos un múltiplo de otra fila, o
- filas intercambiadas.

Otras matrices \mathbf{M}_R operando sobre la derecha de $(\mathbf{A}\mathbf{M}_R)$ puede llevar a las mismas operaciones sobre las columnas de \mathbf{A} .

Esto significa que las filas y las columnas de la matriz pueden ser alteradas (por multiplicación de matrices) como si estuviéramos tratando con determinantes, así podemos aplicar las técnicas de eliminación de Gauss-Jordan a los elementos de matriz. Por tanto existe una matriz \mathbf{M}_L (o \mathbf{M}_R) tal que⁵

$$\mathbf{M}_L\mathbf{A} = \mathbf{1} . \quad (6.49)$$

La $\mathbf{M}_L = \mathbf{A}^{-1}$. Determinamos \mathbf{M}_L realizando las operaciones de eliminación idénticas sobre la matriz unidad. Luego

$$\mathbf{M}_L\mathbf{1} = \mathbf{M}_L . \quad (6.50)$$

Para clarificar ésto consideremos un ejemplo específico.

Deseamos invertir la matriz

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix} . \quad (6.51)$$

Por conveniencia escribimos \mathbf{A} y $\mathbf{1}$ lado a lado realizando operaciones idénticas sobre cada una de ellas

$$\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} . \quad (6.52)$$

Para ser sistemáticos, multiplicamos cada fila para obtener $a_{k1} = 1$,

$$\begin{pmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ 1 & \frac{3}{2} & \frac{1}{2} \\ 1 & 1 & 4 \end{pmatrix} \quad \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} . \quad (6.53)$$

Restando la primera fila de la segunda y tercera, obtenemos

$$\begin{pmatrix} 1 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & \frac{1}{3} & \frac{11}{3} \end{pmatrix} \quad \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ -\frac{1}{3} & \frac{1}{2} & 0 \\ -\frac{1}{3} & 0 & 1 \end{pmatrix} . \quad (6.54)$$

⁵Recordemos que $\det(\mathbf{A}) \neq 0$.

Entonces dividimos la segunda fila (de ambas matrices) por $5/6$ y sustrayéndola $2/3$ veces de la primera fila, y $1/3$ veces de la tercera fila. Los resultados para ambas matrices son

$$\begin{pmatrix} 1 & 0 & \frac{1}{5} \\ 0 & 1 & \frac{1}{5} \\ 0 & 0 & \frac{18}{5} \end{pmatrix} \quad \begin{pmatrix} \frac{3}{5} & -\frac{2}{5} & 0 \\ -\frac{2}{5} & \frac{3}{5} & 0 \\ -\frac{1}{5} & -\frac{1}{5} & 1 \end{pmatrix}. \quad (6.55)$$

Dividimos la tercera fila (de ambas matrices) por $18/5$. Luego como último paso $1/5$ veces la tercera fila es sustraída de cada una de las dos primeras filas (de ambas matrices). Nuestro par final es

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} \frac{11}{8} & -\frac{7}{18} & -\frac{1}{18} \\ -\frac{7}{18} & \frac{11}{18} & -\frac{1}{18} \\ -\frac{1}{18} & -\frac{1}{18} & \frac{5}{18} \end{pmatrix}. \quad (6.56)$$

El chequeo es multiplicar la original A por la calculada A^{-1} para ver si realmente obtuvimos la matriz unidad 1 .

Como con la solución de Gauss-Jordan de ecuaciones algebraicas simultáneas, esta técnica está bien adaptada para computadores.

6.3. Matrices ortogonales.

El espacio de tres dimensiones ordinario puede ser descrito con las coordenadas cartesianas (x_1, x_2, x_3) . Consideremos un segundo conjunto de coordenadas cartesianas (x'_1, x'_2, x'_3) cuyo origen y sentido coinciden con el primero pero su orientación es diferente (figura 6.1). Podemos

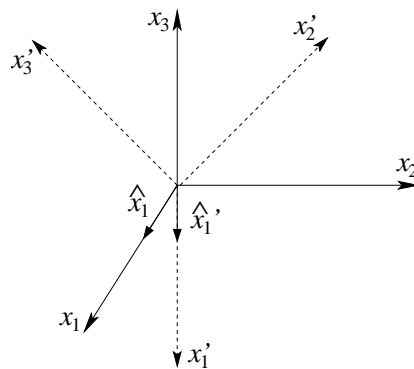


Figura 6.1: Sistemas de coordenadas cartesianos.

decir que el sistema de ejes prima ha sido rotado respecto al inicial sistema de coordenadas sin prima. Ya que esta rotación es una operación lineal, esperamos una ecuación matricial que relaciones la base con primas con la sin primas.

Cosenos directores.

Un vector unitario a lo largo del eje x'_1 (\hat{x}'_1) puede ser resuelto en sus componentes a lo largo de los ejes x_1 , x_2 y x_3 por las usuales técnicas de proyección.

$$\hat{x}'_1 = \hat{x}_1 \cos(x'_1, x_1) + \hat{x}_2 \cos(x'_1, x_2) + \hat{x}_3 \cos(x'_1, x_3) . \quad (6.57)$$

Por conveniencia estos cosenos, los cuales son los cosenos directores, son etiquetados

$$\begin{aligned} \cos(x'_1, x_1) &= \hat{x}'_1 \cdot \hat{x}_1 = a_{11} , \\ \cos(x'_1, x_2) &= \hat{x}'_1 \cdot \hat{x}_2 = a_{12} , \\ \cos(x'_1, x_3) &= \hat{x}'_1 \cdot \hat{x}_3 = a_{13} . \end{aligned} \quad (6.58)$$

Continuando, tenemos

$$\begin{aligned} \cos(x'_2, x_1) &= \hat{x}'_2 \cdot \hat{x}_1 = a_{21} , & (a_{21} \neq a_{12}) , \\ \cos(x'_2, x_2) &= \hat{x}'_2 \cdot \hat{x}_2 = a_{22} , & \text{y así sucesivamente.} \end{aligned} \quad (6.59)$$

Ahora la ecuación (6.57) puede ser reescrita como

$$\hat{x}'_1 = \hat{x}_1 a_{11} + \hat{x}_2 a_{12} + \hat{x}_3 a_{13}$$

y también

$$\begin{aligned} \hat{x}'_2 &= \hat{x}_1 a_{21} + \hat{x}_2 a_{22} + \hat{x}_3 a_{23} \\ \hat{x}'_3 &= \hat{x}_1 a_{31} + \hat{x}_2 a_{32} + \hat{x}_3 a_{33} . \end{aligned} \quad (6.60)$$

También podemos ir de la otra manera resolviendo \hat{x}_1 , \hat{x}_2 y \hat{x}_3 en sus componentes en el sistema con primas. Entonces

$$\begin{aligned} \hat{x}_1 &= \hat{x}'_1 a_{11} + \hat{x}'_2 a_{21} + \hat{x}'_3 a_{31} \\ \hat{x}_2 &= \hat{x}'_1 a_{12} + \hat{x}'_2 a_{22} + \hat{x}'_3 a_{32} \\ \hat{x}_3 &= \hat{x}'_1 a_{13} + \hat{x}'_2 a_{23} + \hat{x}'_3 a_{33} . \end{aligned} \quad (6.61)$$

Aplicaciones a vectores.

Si consideramos un vector cuyas componentes son funciones de la posición, entonces

$$\begin{aligned} \vec{V}(x_1, x_2, x_3) &= \hat{x}_1 V_1 + \hat{x}_2 V_2 + \hat{x}_3 V_3 \\ &= \vec{V}'(x'_1, x'_2, x'_3) = \hat{x}'_1 V'_1 + \hat{x}'_2 V'_2 + \hat{x}'_3 V'_3 , \end{aligned} \quad (6.62)$$

ya que el punto puede ser dado en cualquiera de los dos sistema de coordenadas (x_1, x_2, x_3) o (x'_1, x'_2, x'_3) . Notemos que \vec{V} y \vec{V}' son geoméricamente el mismo vector (pero con diferentes componentes). Si los ejes de coordenadas son rotados, el vector se mantiene fijo. Usando la ecuación (6.60) para eliminar \hat{x}_1 , \hat{x}_2 , \hat{x}_3 , podemos separar la ecuación (6.62) en tres ecuaciones escalares

$$\begin{aligned} V'_1 &= a_{11} V_1 + a_{12} V_2 + a_{13} V_3 \\ V'_2 &= a_{21} V_1 + a_{22} V_2 + a_{23} V_3 \\ V'_3 &= a_{31} V_1 + a_{32} V_2 + a_{33} V_3 . \end{aligned} \quad (6.63)$$

En particular, estas relaciones se mantendrán para las coordenadas de un punto (x_1, x_2, x_3) y (x'_1, x'_2, x'_3) , dando

$$\begin{aligned}x'_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\x'_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\x'_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 ,\end{aligned}\tag{6.64}$$

y similarmente para las coordenadas primas. En esta notación el conjunto de tres ecuaciones (6.64) pueden ser escritas como

$$x'_i = \sum_{j=1}^3 a_{ij}x_j ,\tag{6.65}$$

donde i toma los valores 1, 2 y 3 y el resultado son tres ecuaciones separadas.

De la ecuación anterior podemos derivar interesante información sobre los a_{ij} los cuales describen la orientación del sistema de coordenadas (x'_1, x'_2, x'_3) relativa al sistema (x_1, x_2, x_3) . La distancia respecto al origen es la misma en ambos sistemas. Elevando al cuadrado,

$$\begin{aligned}\sum_i x_i^2 &= \sum_i x_i'^2 \\&= \sum_i \left(\sum_j a_{ij}x_j \right) \left(\sum_k a_{ik}x_k \right) \\&= \sum_{j,k} x_j x_k \sum_i a_{ij}a_{ik} .\end{aligned}\tag{6.66}$$

Esto sólo puede ser cierto para todos los puntos si y sólo si

$$\sum_i a_{ij}a_{ik} = \delta_{jk} , \quad j, k = 1, 2, 3 .\tag{6.67}$$

La ecuación (6.67) es una consecuencia de requerir que la longitud permanezca constante (invariante) bajo rotaciones del sistema de coordenadas, es llamada la *condición de ortogonalidad*. Los a_{ij} escritos como una matriz \mathbf{A} , forman una matriz ortogonal. Notemos que la ecuación (6.67) no es una multiplicación matricial.

En notación matricial la ecuación (6.65) llega a ser

$$|x'\rangle = \mathbf{A}|x\rangle .\tag{6.68}$$

Condiciones de ortogonalidad, caso bidimensional.

Podemos ganar un mejor entendimiento de los a_{ij} y de la condición de ortogonalidad considerando con detalle rotaciones en dos dimensiones. Esto lo podemos pensar como un sistema tridimensional con los ejes x_1 y x_2 rotados respecto a x_3 . De la figura 6.2,

$$\begin{aligned}x'_1 &= x_1 \cos \varphi + x_2 \sin \varphi , \\x'_2 &= -x_1 \sin \varphi + x_2 \cos \varphi .\end{aligned}\tag{6.69}$$

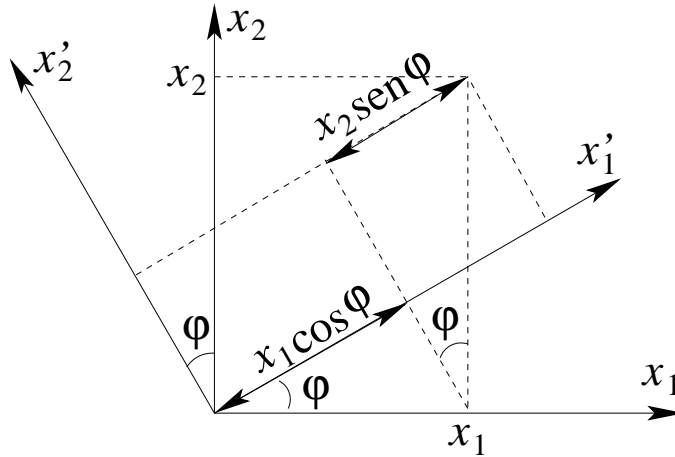


Figura 6.2: Sistemas de coordenadas rotados en dos dimensiones.

Por lo tanto por la ecuación (6.68)

$$A = \begin{pmatrix} \cos \varphi & \text{sen } \varphi \\ -\text{sen } \varphi & \cos \varphi \end{pmatrix}. \quad (6.70)$$

Notemos que A se reduce a la matriz unidad para $\varphi = 0$. La rotación cero significa que nada ha cambiado. Es claro a partir de la figura 6.2 que

$$\begin{aligned} a_{11} &= \cos \varphi = \cos(x'_1, x_1), \\ a_{12} &= \text{sen } \varphi = \cos\left(\frac{\pi}{2} - \varphi\right) = \cos(x'_1, x_2), \quad \text{y así sucesivamente,} \end{aligned} \quad (6.71)$$

de este modo identificamos los elementos de matriz a_{ij} con los cosenos directores. La ecuación (6.67), la condición de ortogonalidad, llega a ser

$$\begin{aligned} \text{sen}^2 \varphi + \cos^2 \varphi &= 1, \\ \text{sen } \varphi \cos \varphi - \text{sen } \varphi \cos \varphi &= 0. \end{aligned} \quad (6.72)$$

la extensión a tres dimensiones (rotación de las coordenadas a lo largo del eje z en un ángulo φ en el sentido de los punteros del reloj) es simplemente

$$A = \begin{pmatrix} \cos \varphi & \text{sen } \varphi & 0 \\ -\text{sen } \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (6.73)$$

El $a_{33} = 1$ expresa el hecho que $x'_3 = x_3$, ya que la rotación ha sido en torno al eje x_3 . Los ceros garantizan que x'_1 y x'_2 no dependen de x_3 y que x'_3 no depende de x_1 y x_2 . En un lenguaje más sofisticado, x_1 y x_2 se extienden sobre un subespacio invariante, mientras que x_3 forma un subespacio invariante por si solo. La forma de A es reducible. La ecuación (6.73) da una posible descomposición.

Matriz inversa \mathbf{A}^{-1} .

Volviendo a la matriz de transformación general \mathbf{A} , la matriz inversa \mathbf{A}^{-1} es definida tal que

$$|x\rangle = \mathbf{A}^{-1}|x'\rangle . \quad (6.74)$$

Esto es, \mathbf{A}^{-1} describe el inverso de la rotación dada por \mathbf{A} y retorna el sistema de coordenadas a su posición original. Simbólicamente, las ecuaciones (6.68) y (6.74) combinadas dan

$$|x\rangle = \mathbf{A}^{-1}\mathbf{A}|x\rangle , \quad (6.75)$$

y ya que $|x\rangle$ es arbitrario,

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{1} , \quad (6.76)$$

la matriz unidad, Similarmente,

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{1} . \quad (6.77)$$

usando las ecuaciones (6.68) y (6.74) y eliminando $|x'\rangle$ en vez de $|x\rangle$.

Matriz transpuesta, $\tilde{\mathbf{A}}$.

Podemos determinar los elementos de nuestra postulada matriz inversa \mathbf{A}^{-1} empleando la condición de ortogonalidad. La ecuación (6.67), la condición de ortogonalidad, no está de acuerdo con nuestra definición de multiplicación matricial, pero la podemos definir de acuerdo a una nueva matriz $\tilde{\mathbf{A}}$ tal que

$$\tilde{a}_{ji} = a_{ij} . \quad (6.78)$$

La ecuación (6.67) llega a ser

$$\tilde{\mathbf{A}}\mathbf{A} = \mathbf{1} . \quad (6.79)$$

Esta es una reformulación de la condición de ortogonalidad y puede ser tomada como una definición de ortogonalidad. Multiplicando (6.79) por \mathbf{A}^{-1} por la derecha y usando la ecuación (6.77), tenemos

$$\tilde{\mathbf{A}} = \mathbf{A}^{-1} . \quad (6.80)$$

Este importante resultado que la inversa es igual a la transpuesta se mantiene sólo para matrices ortogonales y puede ser tomado como una reformulación de la condición de ortogonalidad.

Multiplicando la ecuación (6.80) por \mathbf{A} por la izquierda, obtenemos

$$\mathbf{A}\tilde{\mathbf{A}} = \mathbf{1} , \quad (6.81)$$

o

$$\sum_i a_{ji}a_{ki} = \delta_{jk} , \quad (6.82)$$

lo cual es otra forma más de la condición de ortogonalidad.

Resumiendo, la condición de ortogonalidad puede ser enunciada de varias maneras equivalentes:

$$\sum_i a_{ij}a_{ik} = \delta_{jk} \quad (6.83a)$$

$$\sum_i a_{ji}a_{ki} = \delta_{jk} \quad (6.83b)$$

$$\tilde{A}A = A\tilde{A} = 1 \quad (6.83c)$$

$$\tilde{A} = A^{-1} . \quad (6.83d)$$

Cualquiera de estas relaciones es condición necesaria y suficiente para que A sea ortogonal.

Es posible ahora ver y entender por qué el nombre *ortogonal* es apropiado para estas matrices. Tenemos la forma general

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} ,$$

de una matriz de cosenos directores en la cual a_{ij} es el coseno del ángulo entre x'_i y x_j . Por lo tanto, a_{11} , a_{12} , a_{13} son los cosenos directores de x'_1 relativo a x_1 , x_2 , x_3 . Estos tres elementos de A definen una unidad de longitud a lo largo de x'_1 , esto es, un vector unitario \hat{x}'_1 ,

$$\hat{x}'_1 = \hat{x}_1 a_{11} + \hat{x}_2 a_{12} + \hat{x}_3 a_{13} .$$

La relación de ortogonalidad (ecuación (6.82)) es simplemente una declaración que los vectores unitarios \hat{x}'_1 , \hat{x}'_2 , y \hat{x}'_3 son mutuamente perpendiculares o ortogonales. Nuestra matriz de transformación ortogonal A transforma un sistema ortogonal en un segundo sistema ortogonal de coordenadas por rotación y/o reflexión.

Ángulos de Euler.

Nuestra matriz de transformación A contiene nueve cosenos directores. Claramente, sólo tres de ellos son independientes, la ecuación (6.67) proveen seis restricciones. De otra manera, uno puede decir que necesita dos parámetros (θ y φ en coordenadas polares esféricas) para fijar el eje de rotación, más uno adicional para describir la magnitud de la rotación en torno a ese eje. En la formulación Lagrangiana de la mecánica es necesario describir A usando algún conjunto de tres parámetros independientes más que los redundantes cosenos directores. La elección usual de estos parámetros es la de los ángulos de Euler⁶

El objetivo de describir la orientación de un sistema final rotado (x'''_1, x'''_2, x'''_3) relativo a algún sistema de coordenadas inicial (x_1, x_2, x_3). El sistema final es desarrollado en tres pasos cada paso involucra una rotación descrita por un ángulo de Euler (figura 6.3):

1. Los ejes x'_1 , x'_2 , y x'_3 son rotados respecto al eje x_3 en un ángulo α en el sentido horario relativo a x_1 , x_2 y x_3 . (Los ejes x_3 y x'_3 coinciden.)

⁶No hay una única manera de definir los ángulos de Euler. Usamos la elección usual en Mecánica Cuántica de momento angular.

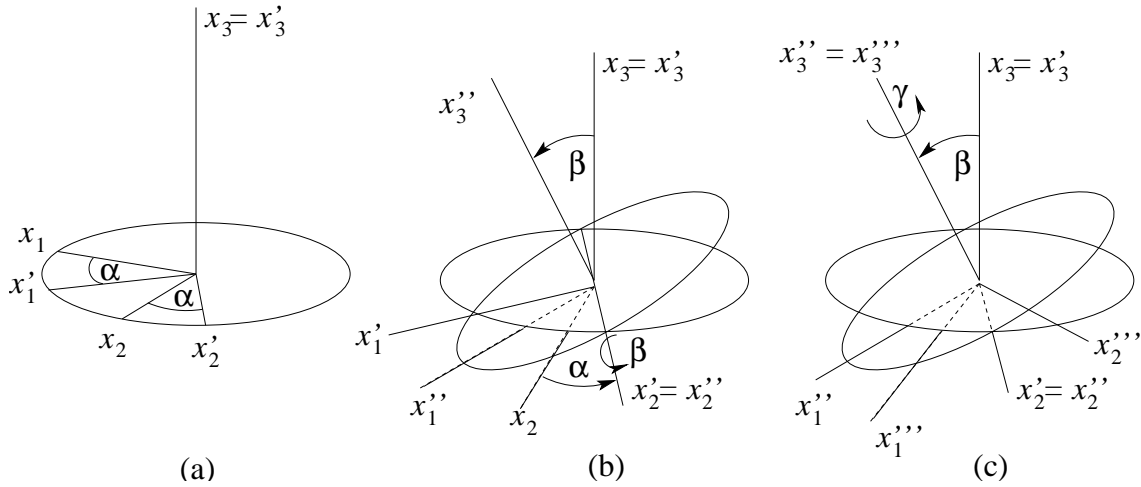


Figura 6.3: (a) Rotación respecto al eje x_3 en un ángulo α ; (b) Rotación respecto a un eje x_2' en un ángulo β ; (c) Rotación respecto a un eje x_3''' en un ángulo γ .

2. los ejes x_1'' , x_2'' , y x_3'' son rotados respecto al eje x_2' en un ángulo β en el sentido horario relativo a x_1' , x_2' y x_3' . (Los ejes x_2' y x_2'' coinciden.)
3. la tercera y final rotación es en un ángulo γ en sentido horario respecto al eje x_3'' produciendo el sistema (x_1''', x_2''', x_3''') . (Los ejes x_3'' y x_3''' coinciden.)

Las tres matrices que describen estas rotaciones son:

$$R_z(\alpha) = \begin{pmatrix} \cos \alpha & \text{sen } \alpha & 0 \\ -\text{sen } \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (6.84)$$

exactamente como en la ecuación (6.73,

$$R_y(\beta) = \begin{pmatrix} \cos \beta & 0 & -\text{sen } \beta \\ 0 & 1 & 0 \\ \text{sen } \beta & 0 & \cos \beta \end{pmatrix} \quad (6.85)$$

y

$$R_z(\gamma) = \begin{pmatrix} \cos \gamma & \text{sen } \gamma & 0 \\ -\text{sen } \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (6.86)$$

La rotación total es descrita por el producto matricial triple,

$$A(\alpha, \beta, \gamma) = R_z(\gamma)R_y(\beta)R_z(\alpha). \quad (6.87)$$

Notemos el orden: $R_z(\alpha)$ opera primero, entonces $R_y(\beta)$, y finalmente $R_z(\gamma)$. La multiplicación da

$$A = \begin{pmatrix} \cos \gamma \cos \beta \cos \alpha - \text{sen } \gamma \text{sen } \alpha & \cos \gamma \cos \beta \text{sen } \alpha - \text{sen } \gamma \cos \alpha & -\cos \gamma \text{sen } \beta \\ -\text{sen } \gamma \cos \beta \cos \alpha - \cos \gamma \text{sen } \alpha & -\text{sen } \gamma \cos \beta \text{sen } \alpha + \cos \gamma \cos \alpha & \text{sen } \gamma \text{sen } \beta \\ \text{sen } \beta \cos \alpha & \text{sen } \beta \text{sen } \alpha & \cos \beta \end{pmatrix}. \quad (6.88)$$

Comparando $A(a_{ij})$ con $A(\alpha, \beta, \gamma)$ elemento por elemento, nos produce los cosenos directores en términos de los ángulos de Euler.

Propiedades de simetría.

Nuestra descripción matricial conduce al grupo de rotaciones $SO(3)$ en el espacio tridimensional \mathbb{R}^3 , y la descripción en términos de ángulos de Euler de las rotaciones forman una base para desarrollar el grupo de rotaciones. Las rotaciones pueden también ser descritas por el grupo unitario $SU(2)$ en el espacio bidimensional \mathbb{C}^2 .

La matriz transpuesta es útil en la discusión de las propiedades de simetría. Si

$$A = \tilde{A}, \quad a_{ij} = a_{ji}, \quad (6.89)$$

la matriz es llamada *simétrica*, mientras que si

$$A = -\tilde{A}, \quad a_{ij} = -a_{ji}, \quad (6.90)$$

es llamada *antisimétrica*. Los elementos de la diagonal son nulos. Es fácil mostrar que cualquier matriz cuadrada puede ser escrita como la suma de una matriz simétrica y una antisimétrica. Consideremos la identidad

$$A = \frac{1}{2} [A + \tilde{A}] + \frac{1}{2} [A - \tilde{A}]. \quad (6.91)$$

$A + \tilde{A}$ es claramente simétrica, mientras que $A - \tilde{A}$ es claramente antisimétrica.

Hasta ahora hemos interpretado las matrices ortogonales como rotaciones del sistema de coordenadas. Estas cambian las componentes de un vector fijo. Sin embargo, una matriz ortogonal A puede ser interpretada igualmente bien como una rotación del vector en la dirección opuesta (figura 6.4).

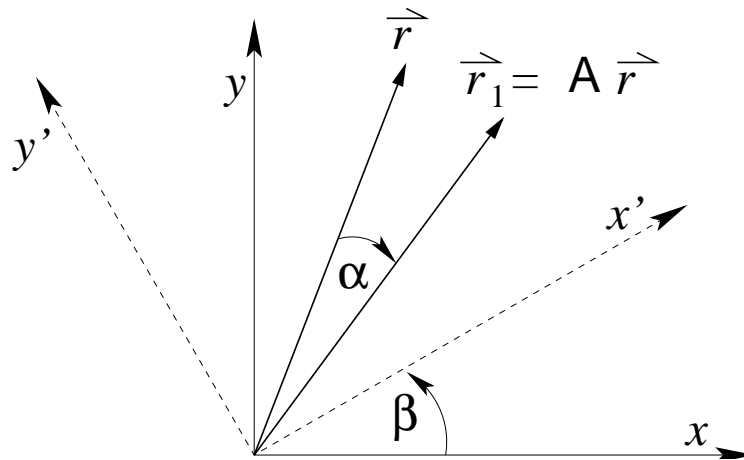


Figura 6.4: Vector fijo con coordenadas rotadas.

Estas dos posibilidades, (1) rotar el vector manteniendo la base fija y (2) rotar la base (en el sentido opuesto) manteniendo el vector fijo.

Supongamos que interpretamos la matriz \mathbf{A} como rotar un vector \vec{r} en una nueva posición \vec{r}_1 , *i.e.*, en un particular sistema de coordenadas tenemos la relación

$$\vec{r}_1 = \mathbf{A}\vec{r} . \quad (6.92)$$

Ahora rotemos las coordenadas aplicando una matriz \mathbf{B} , la cual rota (x, y, z) en (x', y', z') ,

$$\begin{aligned} \vec{r}'_1 &= \mathbf{B}\vec{r}_1 = \mathbf{B}\mathbf{A}\vec{r} = (\mathbf{A}\vec{r})' \\ &= \mathbf{B}\mathbf{A}(\mathbf{B}^{-1}\vec{r}) \\ &= (\mathbf{B}\mathbf{A}\mathbf{B}^{-1})\vec{r} = (\mathbf{B}\mathbf{A}\mathbf{B}^{-1})\vec{r}' . \end{aligned} \quad (6.93)$$

$\mathbf{B}\vec{r}_1$ es justo \vec{r}'_1 en el nuevo sistema de coordenadas con una interpretación similar se mantiene para $\mathbf{B}\vec{r}$. Ya que en este nuevo sistema $(\mathbf{B}\vec{r})$ es rotado a la posición $(\mathbf{B}\vec{r}_1)$ por la matriz $\mathbf{B}\mathbf{A}\mathbf{B}^{-1}$.

$$\begin{array}{ccc} \mathbf{B}\vec{r}_1 &= & (\mathbf{B}\mathbf{A}\mathbf{B}^{-1}) \mathbf{B}\vec{r} \\ \downarrow & & \downarrow \quad \downarrow \\ \vec{r}'_1 &= & \mathbf{A}' \quad \vec{r}' . \end{array}$$

En el nuevo sistema las coordenadas han sido rotadas por la matriz \mathbf{B} , \mathbf{A} tiene la forma \mathbf{A}' , en la cual

$$\mathbf{A}' = \mathbf{B}\mathbf{A}\mathbf{B}^{-1} . \quad (6.94)$$

\mathbf{A}' opera en el espacio x', y', z' como \mathbf{A} opera en el espacio x, y, z .

La transformación definida por la ecuación (6.94) con \mathbf{B} cualquier matriz, no necesariamente ortogonal, es conocida como transformación de similaridad. Por componentes la ecuación (6.94) llega a ser

$$a'_{ij} = \sum_{k,l} b_{ik} a_{kl} (\mathbf{B}^{-1})_{lj} . \quad (6.95)$$

Ahora si \mathbf{B} es ortogonal,

$$(\mathbf{B}^{-1})_{lj} = (\tilde{\mathbf{B}})_{lj} = b_{jl} , \quad (6.96)$$

y tenemos

$$a'_{ij} = \sum_{k,l} b_{ik} b_{jl} a_{kl} . \quad (6.97)$$

La matriz \mathbf{A} es la representación de un mapeo lineal en un sistema de coordenadas dado o base. Pero hay direcciones asociadas con \mathbf{A} , ejes cristalinos, ejes de simetría en un sólido rotando y etc. tal que la representación depende de la base. La transformación de similaridad muestran justo como la representación cambia con un cambio de base.

6.4. Matrices Hermíticas, matrices unitarias.

Definiciones.

Hasta aquí hemos generalmente supuesto que nuestro espacio vectorial es un espacio real y que los elementos de las matrices (la representación de los operadores lineales) son reales. Para muchos cálculos en Física Clásica los elementos de matriz reales serán suficientes. Sin

embargo, en Mecánica Cuántica las variables complejas son inevitables por la forma de las reglas de conmutación básicas (o la ecuación tiempo dependiente de Schödinger). Con esto en mente, generalizamos al caso de matrices con elementos complejos. Para manejar estos elementos, definamos algunas propiedades.

1. Compleja conjugada, A^* , formada por tomar los complejos conjugados ($i \rightarrow -i$) de cada elemento, donde $i = \sqrt{-1}$.
2. Adjunta, A^\dagger , formada por transponer A^* ,

$$A^\dagger = \widetilde{A^*} = \widetilde{A}^* . \quad (6.98)$$

3. Matriz hermítica: La matriz es etiquetada como hermítica (o autoadjunta) si

$$A = A^\dagger . \quad (6.99)$$

Si A es real, entonces $A^\dagger = \widetilde{A}$, y las matrices hermíticas reales son matrices reales y simétricas. En Mecánica Cuántica las matrices son hermíticas o unitarias.

4. Matriz unitaria: La matriz U es etiquetada como unitaria si

$$U^\dagger = U^{-1} . \quad (6.100)$$

Si U es real, entonces $U^{-1} = \widetilde{U}$, tal que las matrices reales unitarias son matrices ortogonales. Este representa una generalización del concepto de matriz ortogonal.

5. $(AB)^* = B^*A^*$, $(AB)^\dagger = B^\dagger A^\dagger$.

Si los elementos son complejos, a la Física casi siempre le interesan las matrices adjuntas, hermíticas y unitarias. Las matrices unitarias son especialmente importantes en Mecánica Cuántica porque ellos dejan el largo de un vector (complejo) inalterado, análoga a la operación de una matriz ortogonal sobre un vector real. Una importante excepción a este interés en las matrices unitarias es el grupo de matrices de Lorentz.

En un espacio n -dimensional complejo el cuadrado del largo de un punto $\tilde{x} = (x_1, x_2, \dots, x_n)$, o el cuadrado de su distancia al origen, es definido como $x^\dagger x = \sum_i x_i^* x_i = \sum_i |x_i|^2$. Si una transformación de coordenadas $y = Ux$ deja la distancia inalterada, entonces $x^\dagger x = y^\dagger y = (Ux)^\dagger Ux = x^\dagger U^\dagger Ux$. Ya que x es arbitrario concluimos que $U^\dagger U = \mathbf{1}_n$, *i.e.*, U es una matriz unitaria de $n \times n$. Si $x' = Ax$ es un mapa lineal, entonces su matriz en las nuevas coordenadas llega a ser una transformación unitaria (análogo de una de similaridad)

$$A' = UAU^\dagger ,$$

porque $Ux' = y' = UAx = UAU^{-1}y = UAU^\dagger y$.

Matrices de Pauli y de Dirac.

El conjunto de tres matrices de Pauli de 2×2 σ ,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (6.101)$$

fueron introducidas por W. Pauli para describir una partícula de spin $\frac{1}{2}$ en Mecánica Cuántica no relativista. Se puede demostrar que las σ satisfacen

$$\sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij} \mathbf{1}_2, \quad \text{anticonmutación} \quad (6.102)$$

$$\sigma_i \sigma_j = i\sigma_k, \quad \text{permutación cíclica de los índices} \quad (6.103)$$

$$(\sigma_i)^2 = \mathbf{1}_2, \quad (6.104)$$

donde $\mathbf{1}_2$ es la matriz unidad de 2×2 . Así, el vector $\vec{\sigma}/2$ satisface las mismas reglas de conmutación

$$[\sigma_i, \sigma_j] \equiv \sigma_i \sigma_j - \sigma_j \sigma_i = 2i\varepsilon_{ijk} \sigma_k, \quad (6.105)$$

que el momento angular \vec{L} .

Las tres matrices de Pauli $\vec{\sigma}$ y la matriz unitaria forman un conjunto completo tal que cualquier matriz de 2×2 \mathbf{M} puede ser expandida como

$$\mathbf{M} = m_0 \mathbf{1} + m_1 \sigma_1 + m_2 \sigma_2 + m_3 \sigma_3 = m_0 \mathbf{1} + \vec{m} \cdot \vec{\sigma}, \quad (6.106)$$

donde los m_i son constantes. Usando $\sigma_i^2 = 1$ y $\text{tr}(\sigma_i) = 0$ nosotros obtenemos a partir de la ecuación (6.106) los coeficientes m_i de la expansión formando las trazas,

$$2m_0 = \text{tr}(\mathbf{M}), \quad 2m_i = \text{tr}(\mathbf{M} \sigma_i), \quad i = 1, 2, 3. \quad (6.107)$$

En 1927 P.A.M. Dirac extendió este formalismo para partículas de spin $\frac{1}{2}$ moviéndose a velocidades cercana a la de la luz tales como electrones Para incluir la relatividad especial su punto de partida es la ecuación de Einstein para la energía $E^2 = \vec{p}^2 c^2 + m^2 c^4$ en vez de la energía cinética y potencial no relativista $E = \vec{p}^2/2m + V$. La clave para la ecuación de Dirac es factorizar

$$E^2 - \vec{p}^2 c^2 = E^2 - (c\vec{\sigma} \cdot \vec{p})^2 = (E - c\vec{\sigma} \cdot \vec{p})(E + c\vec{\sigma} \cdot \vec{p}) = m^2 c^4, \quad (6.108)$$

usando la identidad matricial en 2×2

$$(c\vec{\sigma} \cdot \vec{p})^2 = \vec{p}^2 \mathbf{1}_2. \quad (6.109)$$

La matriz unidad de 2×2 $\mathbf{1}_2$ no es escrita explícitamente en la ecuación (6.108) y (6.109). Podemos presentar las matrices γ_0 y γ para factorizar $E^2 - \vec{p}^2 c^2$ directamente,

$$(\gamma_0 E - \gamma c\vec{\sigma} \cdot \vec{p})^2 = \gamma_0^2 E^2 + \gamma^2 c^2 (\vec{\sigma} \cdot \vec{p})^2 - E c\vec{\sigma} \cdot \vec{p} (\gamma_0 \gamma + \gamma \gamma_0) = E^2 - \vec{p}^2 c^2 = m^2 c^4. \quad (6.110)$$

Si reconocemos

$$\gamma_0 E - \gamma c\vec{\sigma} \cdot \vec{p} = \gamma \cdot p = (\gamma_0, \gamma \vec{\sigma}) \cdot (E, c\vec{p}), \quad (6.111)$$

como el producto escalar de dos cuadvectores γ^μ y p^μ , entonces la ecuación (6.110) con $p^2 = p \cdot p = E^2 - \vec{p}^2 c^2$ puede ser visto como una generalización cuadvectorial de la ecuación (6.109). Claramente, para que la ecuación (6.110) mantenga las condiciones

$$\gamma_0^2 = 1 = -\gamma^2, \quad \gamma_0 \gamma + \gamma \gamma_0 = 0, \quad (6.112)$$

debe satisfacerse que las cuatro matrices γ^μ anticonmuten, justo como las tres matrices de Pauli. Ya que estas últimas son un conjunto completo de matrices anticonmutantes de 2×2 , la condición (6.112) no puede ser satisfacerse para matrices de 2×2 , pero ella puede ser satisfecha para matrices de 4×4

$$\begin{aligned} \gamma_0 = \gamma^0 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_2 & 0 \\ 0 & -\mathbf{1}_2 \end{pmatrix}, \\ \gamma &= \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{1}_2 \\ -\mathbf{1}_2 & 0 \end{pmatrix}. \end{aligned} \quad (6.113)$$

Alternativamente, el vector de matrices de 4×4

$$\gamma = \begin{pmatrix} 0 & \vec{\sigma} \\ -\vec{\sigma} & 0 \end{pmatrix} = \gamma \vec{\sigma} = \sigma_1 \times \vec{\sigma}, \quad (6.114)$$

puede obtenerse como el producto directo en el mismo sentido de la sección 6.2 de las matrices de 2×2 de Pauli. De la misma manera, $\gamma_0 = \sigma_3 \times \mathbf{1}_2$ y $\mathbf{1}_4 = \mathbf{1}_2 \times \mathbf{1}_2$.

Resumiendo el tratamiento relativista de una partícula de spin $\frac{1}{2}$, produce matrices de 4×4 , mientras que las partículas no relativistas de spin $\frac{1}{2}$ son descritas por las matrices de Pauli σ de 2×2 .

6.5. Diagonalización de matrices.

Momento de la matriz de inercia .

En muchos problemas en Física que involucran matrices reales simétricas o complejas hermíticas es deseable llevar a cabo una real transformación de similaridad ortogonal o una transformación unitaria (correspondiente a una rotación del sistema de coordenadas) para reducir la matriz a una forma diagonal, con todos los elementos no diagonales nulos. Un ejemplo particularmente directo de esto es la matriz del momento de inercia I de un cuerpo rígido. A partir de la definición del momento angular \vec{L} tenemos

$$\vec{L} = I \vec{\omega}, \quad (6.115)$$

donde $\vec{\omega}$ viene a ser la velocidad angular. La matriz de inercia I tiene elementos diagonales

$$I_{xx} = \sum_i m_i (r_i^2 - x_i^2), \text{ y así sucesivamente,} \quad (6.116)$$

el subíndice i referencia la masa m_i localizada en $\vec{r}_i = (x_i, y_i, z_i)$. Para las componentes no diagonales tenemos

$$I_{xy} = - \sum_i m_i x_i y_i = I_{yx} . \quad (6.117)$$

Por inspección la matriz I es simétrica. También, ya que I aparece en una ecuación física de la forma (6.115), la cual se mantiene para todas las orientaciones del sistema de coordenadas, esta puede ser considerada un tensor (regla del cociente).

La clave ahora es la orientación de los ejes (a lo largo de un cuerpo fijo) tal que I_{xy} y los otros elementos no diagonales desaparezcan. Como una consecuencia de esta orientación y una indicación de ella, si la velocidad angular está a lo largo de tales realineados ejes, la velocidad angular y el momento angular serán paralelos.

Autovectores y autovalores (eigenvector y eigenvalues).

Es quizás instructivo considerar un cuadro geométrico asociado a este problema. Si la matriz de inercia I es multiplicada a cada lado por un vector unitario cuya dirección es variable, $\hat{n} = (\alpha, \beta, \gamma)$, entonces en notación de Dirac

$$\langle \hat{n} | I | \hat{n} \rangle = I , \quad (6.118)$$

donde I es el momento de inercia respecto a la dirección \hat{n} y es un número positivo (escalar). Llevando a cabo la multiplicación, obtenemos

$$I = I_{xx}\alpha^2 + I_{yy}\beta^2 + I_{zz}\gamma^2 + 2I_{xy}\alpha\beta + 2I_{xz}\alpha\gamma + 2I_{yz}\beta\gamma . \quad (6.119)$$

Si introducimos

$$\vec{n} = \frac{\hat{n}}{\sqrt{I}} = (n_1, n_2, n_3) , \quad (6.120)$$

la cual es variable en dirección y magnitud entonces la ecuación (6.119) llega a ser

$$1 = I_{xx}n_1^2 + I_{yy}n_2^2 + I_{zz}n_3^2 + 2I_{xy}n_1n_2 + 2I_{xz}n_1n_3 + 2I_{yz}n_2n_3 , \quad (6.121)$$

una forma cuadrática positiva la cual debe ser un elipsoide (ver figura 6.5).

A partir de la geometría analítica es sabido que los ejes de coordenadas pueden ser rotados para coincidir con los ejes de nuestro elipsoide. En muchos casos elementales, espacialmente cuando hay simetría, estos nuevos ejes, llamados ejes principales, pueden ser encontrados por inspección. Ahora nosotros procederemos a desarrollar un método general de hallazgo de los elementos diagonales y los ejes principales.

Si $R^{-1} = \tilde{R}$ es la correspondiente matriz ortogonal real tal que $\vec{n}' = R\vec{n}$, o $|n'\rangle = R|n\rangle$ en la notación de Dirac, son las nuevas coordenadas, luego obtenemos usando $\langle n' | R = \langle n |$ en la ecuación (6.121)

$$\langle n | I | n \rangle = \langle n' | R I \tilde{R} | n' \rangle = I'_1 n_1'^2 + I'_2 n_2'^2 + I'_3 n_3'^2 , \quad (6.122)$$

donde los $I'_i > 0$ son los momentos de inercia principales. La matriz de inercia I' en la ecuación (6.122) es diagonal en las nuevas coordenadas,

$$I' = R I \tilde{R} = \begin{pmatrix} I'_1 & 0 & 0 \\ 0 & I'_2 & 0 \\ 0 & 0 & I'_3 \end{pmatrix} . \quad (6.123)$$

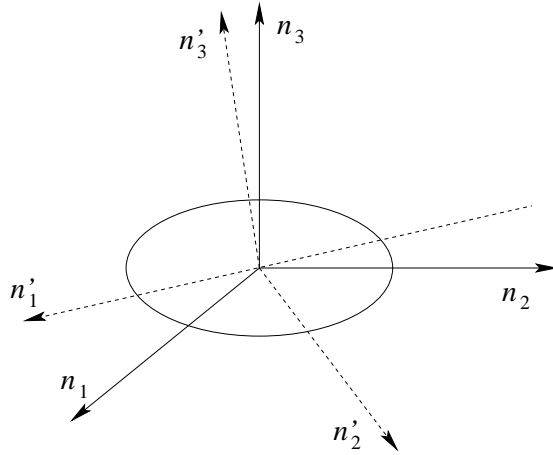


Figura 6.5: Elipsoide del momento de inercia.

Si reescribimos la ecuación (6.123) usando $R^{-1} = \tilde{R}$

$$\tilde{R}' = I\tilde{R} , \quad (6.124)$$

y tomando $\tilde{R} = (\vec{v}_1, \vec{v}_2, \vec{v}_3)$ compuesto de tres vectores columnas, entonces la ecuación (6.124) se separa en tres ecuaciones de autovalores

$$I\vec{v}_i = I'_i\vec{v}_i , \quad i = 1, 2, 3 , \quad (6.125)$$

con autovalores I'_i y autovectores \vec{v}_i . Como estas ecuaciones son lineales y homogéneas (para un i fijo), por la sección 6.1 los determinantes tienen que anularse:

$$\begin{vmatrix} I_{11} - I'_1 & I_{12} & I_{13} \\ I_{21} & I_{22} - I'_2 & I_{23} \\ I_{31} & I_{32} & I_{33} - I'_3 \end{vmatrix} = 0 . \quad (6.126)$$

Reemplazando los autovalores I'_i por una variable λ veces la matriz unidad $\mathbf{1}$, podríamos reescribir la ecuación (6.125) como

$$(I - \lambda\mathbf{1})v = 0 , \quad (6.127)$$

cuyo determinante

$$|I - \lambda\mathbf{1}| = 0 , \quad (6.128)$$

es un polinomio cúbico en λ ; sus tres raíces, por supuesto, son los I'_i . Sustituyendo una raíz de regreso en la ecuación (6.125), podemos encontrar los correspondientes autovectores. La ecuación (6.126) (o la (6.128)) es conocida como la *ecuación secular*. El mismo tratamiento se aplica a una matriz simétrica real I , excepto que sus autovalores no necesitan ser todos positivos. También, la condición de ortogonalidad en la ecuación (6.83a-6.83d) para R dice que, en términos geométricos, los autovectores \vec{v}_i son vectores mutuamente ortogonales unitarios. Por cierto ellos forman las nuevas coordenadas del sistema. El hecho que cualquier par de

autovectores \vec{v}_i, \vec{v}_j son ortogonales si $I'_i \neq I'_j$ se deduce de la ecuación (6.125) en conjunción con la simetría de \mathbf{I} multiplicando con \vec{v}_i y \vec{v}_j , respectivamente,

$$\langle v_j | \mathbf{I} | v_i \rangle = I'_i \langle v_j | v_i \rangle = \langle v_i | \mathbf{I} | v_j \rangle = I'_j \langle v_j | v_i \rangle . \quad (6.129)$$

Ya que $I'_i \neq I'_j$ y la ecuación (6.129) implica que $(I'_i - I'_j) \vec{v}_i \cdot \vec{v}_j = 0$, por lo tanto $\vec{v}_i \cdot \vec{v}_j = 0$.

Matrices hermíticas.

Para espacios vectoriales complejos las matrices unitarias y hermíticas juegan el mismo rol como las matrices ortogonales y simétricas sobre los espacios vectoriales reales, respectivamente. Primero, generalicemos el importante teorema acerca de los elementos diagonales y los ejes principales para la ecuación de autovalores

$$\mathbf{A}|r\rangle = \lambda|r\rangle . \quad (6.130)$$

Ahora mostramos que si \mathbf{A} es una matriz hermítica, sus autovalores son reales y sus autovectores ortogonales.

Sean λ_i y λ_j dos autovalores y $|r_i\rangle$ y $|r_j\rangle$, los correspondientes autovectores de \mathbf{A} , una matriz hermítica. Entonces

$$\mathbf{A}|r_i\rangle = \lambda_i|r_i\rangle \quad (6.131)$$

$$\mathbf{A}|r_j\rangle = \lambda_j|r_j\rangle . \quad (6.132)$$

La ecuación (6.131) es multiplicada por $|r_j\rangle$

$$\langle r_j | \mathbf{A} | r_i \rangle = \lambda_i \langle r_j | r_i \rangle . \quad (6.133)$$

La ecuación (6.132) es multiplicada por $|r_i\rangle$ para dar

$$\langle r_i | \mathbf{A} | r_j \rangle = \lambda_j \langle r_i | r_j \rangle . \quad (6.134)$$

Tomando la adjunta conjugada de esta ecuación, tenemos

$$\langle r_j | \mathbf{A}^\dagger | r_i \rangle = \lambda_j^* \langle r_j | r_i \rangle \quad (6.135)$$

o

$$\langle r_j | \mathbf{A} | r_i \rangle = \lambda_j^* \langle r_j | r_i \rangle , \quad (6.136)$$

ya que \mathbf{A} es hermítica. Sustrayendo la ecuación (6.136) de la ecuación (6.133), obtenemos

$$(\lambda_i - \lambda_j^*) \langle r_j | r_i \rangle . \quad (6.137)$$

Este es un resultado general para todas las combinaciones posibles de i y j . Primero, sea $j = i$. Luego la ecuación (6.137) se convierte en

$$(\lambda_i - \lambda_i^*) \langle r_i | r_i \rangle = 0 . \quad (6.138)$$

Ya que $\langle r_i | r_i \rangle = 0$ sería una solución trivial de la ecuación (6.138), concluimos que

$$\lambda_i = \lambda_i^* , \quad (6.139)$$

es decir, λ_i es real, para todo i .

Segundo, para $i \neq j$ y $\lambda_i \neq \lambda_j$,

$$(\lambda_i - \lambda_j) \langle r_i | r_j \rangle = 0 \quad (6.140)$$

o

$$\langle r_i | r_j \rangle = 0 \quad (6.141)$$

lo cual significa que los autovectores de distintos autovalores son ortogonales, la ecuación (6.141) siendo la generalización de ortogonalidad en este espacio complejo.

Si $\lambda_i = \lambda_j$ (caso degenerado), $\langle r_i |$ no es automáticamente ortogonal a $|r_j\rangle$, pero podría hacerse ortogonal. Consideremos el problema físico de la matriz del momento de inercia nuevamente. Si x_i es un eje de simetría rotacional, entonces encontraremos que $\lambda_2 = \lambda_3$. Los autovectores $|r_2\rangle$ y $|r_3\rangle$ son cada uno perpendiculares al eje de simetría, $|r_1\rangle$, pero ellos yacen en alguna parte en el plano perpendicular a $|r_1\rangle$; esto es, alguna combinación lineal de $|r_2\rangle$ y $|r_3\rangle$ es también un autovector. Considere $(a_2|r_2\rangle + a_3|r_3\rangle)$ con a_2 y a_3 constantes. Entonces

$$\begin{aligned} A(a_2|r_2\rangle + a_3|r_3\rangle) &= a_2\lambda_2|r_2\rangle + a_3\lambda_3|r_3\rangle \\ &= \lambda_2(a_2|r_2\rangle + a_3|r_3\rangle) , \end{aligned} \quad (6.142)$$

como es esperado, para x_1 un eje de simetría rotacional. Por lo tanto, si $|r_1\rangle$ y $|r_2\rangle$ son fijos, $|r_3\rangle$, puede simplemente escogerse yaciendo en el plano perpendicular a $|r_1\rangle$ y también perpendicular a $|r_2\rangle$. Un método general para ortogonalizar soluciones conocido como proceso de Gram-Schmidt, es aplicado a funciones más adelante.

El conjunto de n autovectores ortogonales de nuestra matriz hermítica de $n \times n$ forma un conjunto completo, generando el espacio de n dimensiones complejo. Este hecho es útil en un cálculo variacional de los autovalores. Los autovalores y los autovectores no están limitados a las matrices hermíticas. Todas las matrices tienen autovalores y autovectores. Por ejemplo, la matriz T de población estocástica satisface una ecuación de autovalores

$$T\vec{P}_{\text{equilibrio}} = \lambda\vec{P}_{\text{equilibrio}} ,$$

con $\lambda = 1$. Sin embargo, solamente las matrices hermíticas tienen todos los autovectores ortogonales y todos sus autovalores reales.

Matrices antihermíticas.

Ocasionalmente, en Mecánica Cuántica encontramos matrices antihermíticas:

$$A^\dagger = -A .$$

Siguiendo el análisis de la primera porción de esta sección, podemos mostrar que

- a. Los autovalores son imaginarios puros (o cero).
- b. Los autovectores correspondientes a autovalores distintos son ortogonales.

La matriz R formada de los autovectores normalizados es unitaria. Esta propiedad anti-hermítica es preservada bajo transformaciones unitarias.

Ejemplo: Autovalores y autovectores de una matriz real simétrica.

Sea

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} . \quad (6.143)$$

La ecuación secular es

$$\begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{vmatrix} = 0 , \quad (6.144)$$

o

$$-\lambda(\lambda^2 - 1) = 0 , \quad (6.145)$$

expandiéndolo por las menores. Las raíces son $\lambda = -1, 0, 1$. Para encontrar el autovector correspondiente a $\lambda = -1$, sustituimos este valor de vuelta en la ecuación de autovalores, ecuación (6.130),

$$\begin{pmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} . \quad (6.146)$$

Con $\lambda = -1$, esto produce

$$x + y = 0 , \quad z = 0 . \quad (6.147)$$

Dentro de un factor de escala arbitrario, y un signo arbitrario (factor de fase), $\langle r_1 | = (1, -1, 0)$. Notemos que (para el real $|r\rangle$ en el espacio ordinario) el autovector define una línea en el espacio. El sentido positivo o negativo no está determinado. Esta indeterminación puede ser entendida si notamos que la ecuación (6.130) es homogénea en $|r\rangle$. Por conveniencia requeriremos que los autovectores estén normalizados a la unidad, $\langle r_1 | r_1 \rangle = 1$. Con esta elección de signo

$$\langle r_1 | = \vec{r}_1 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0 \right) , \quad (6.148)$$

está fijo. Para $\lambda = 0$, la ecuación (6.130) produce

$$y = 0 , \quad x = 0 , \quad (6.149)$$

$\langle r_2 |$ o $\vec{r}_2 = (0, 0, 1)$ es un autovector aceptable. Finalmente, para $\lambda = 1$, tenemos

$$-x + y = 0 , \quad z = 0 , \quad (6.150)$$

o

$$\langle r_3 | = \vec{r}_3 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0 \right) . \quad (6.151)$$

La ortogonalidad de \vec{r}_1 , \vec{r}_2 y \vec{r}_3 , correspondientes a los tres autovalores distintos, puede ser fácilmente verificada.

Ejemplo: Autovalores degenerados.

Consideremos

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \quad (6.152)$$

La ecuación secular es

$$\begin{vmatrix} 1 - \lambda & 0 & 0 \\ 0 & -\lambda & 1 \\ 0 & 1 & -\lambda \end{vmatrix} = 0, \quad (6.153)$$

o

$$(1 - \lambda)(\lambda^2 - 1) = 0, \quad \lambda = -1, 1, 1, \quad (6.154)$$

un caso degenerado. Si $\lambda = -1$, la ecuación de autovalores (6.130) produce

$$2x = 0, \quad y + z = 0. \quad (6.155)$$

Un autovector normalizado adecuado es

$$\langle r_1 | = \vec{r}_1 = \left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right). \quad (6.156)$$

para $\lambda = 1$, tenemos

$$-y + z = 0. \quad (6.157)$$

Cualquier autovector que satisface la ecuación (6.157) es perpendicular a \vec{r}_1 . Tenemos infinito número de opciones. Tomemos una elección posible tomando

$$\langle r_2 | = \vec{r}_2 = \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \quad (6.158)$$

la cual claramente satisface la ecuación (6.157). Entonces \vec{r}_3 debe ser perpendicular a \vec{r}_1 y puede ser escogido perpendicular a \vec{r}_2 por⁷

$$\vec{r}_3 = \vec{r}_1 \times \vec{r}_2 = (1, 0, 0). \quad (6.159)$$

Funciones de matrices.

Polinomios con uno o más argumentos matriciales están bien definidos y ocurren a menudo. Series de potencias de una matriz también pueden estar definidas para dar la convergencia de la serie para cada uno de los elementos de matriz. Por ejemplo, si \mathbf{A} es cualquiera matriz de $n \times n$ entonces la serie de potencia

$$\exp(\mathbf{A}) = \sum_{i=0}^{\infty} \frac{\mathbf{A}^i}{i!}, \quad (6.160a)$$

$$\text{sen}(\mathbf{A}) = \sum_{i=0}^{\infty} (-1)^i \frac{\mathbf{A}^{2i+1}}{(2i+1)!}, \quad (6.160b)$$

$$\text{cos}(\mathbf{A}) = \sum_{i=0}^{\infty} (-1)^i \frac{\mathbf{A}^{2i}}{(2i)!}, \quad (6.160c)$$

⁷El uso del producto cruz es limitado a tres dimensiones.

son matrices de $n \times n$ bien definidas. Para todas las matrices de Pauli σ_k la identidad de Euler para θ real y $k=1, 2$ o 3

$$\exp(i\sigma_k\theta) = \mathbf{1}_2 \cos(\theta) + i\sigma_k \text{sen}(\theta) , \quad (6.161)$$

sale a partir de coleccionar las potencias pares e impares en series separadas usando $\sigma_k^2 = \mathbf{1}$. Para las matrices de Dirac σ^{ij} de 4×4 con $(\sigma^{ij})^2 = \mathbf{1}$, si $j \neq k = 1, 2$ o 3 , obtenemos de manera similar (sin escribir las obvias matrices $\mathbf{1}_4$ nunca más)

$$\exp(i\sigma^{jk}\theta) = \cos(\theta) + i\sigma^{jk} \text{sen}(\theta) , \quad (6.162)$$

mientras

$$\exp(i\sigma^{0k}\zeta) = \cosh(\zeta) + i\sigma^{0k} \text{senh}(\zeta) , \quad (6.163)$$

manteniendo ζ real porque $(i\sigma^{0k})^2 = \mathbf{1}$ para $k = 1, 2$ o 3 .

Para una matriz hermítica A hay una matriz unitaria U que la diagonaliza, es decir, $UAU^\dagger = [a_1, a_2, \dots, a_n]$. Entonces la *fórmula de la traza*

$$\det(\exp(A)) = \exp(\text{tr}(A)) \quad (6.164)$$

Puede ser fácilmente demostrado.

Otra importante relación es la de *fórmula de Baker-Hausdorff*

$$\exp(iG)H \exp(-iG) = H + [iG, H] + \frac{[iG, [iG, H]]}{2!} + \dots \quad (6.165)$$

lo cual resulta de multiplicar la serie de potencia para $\exp(iG)$ y recolectar los términos de la misma potencia en iG . Aquí definimos

$$[G, H] = GH - HG$$

como el conmutador de G con H .

6.6. Matrices normales.

En la sección 6.5 nos concentramos principalmente en matrices hermíticas o reales simétricas y en el proceso de encontrar autovalores y autovectores. En esta sección generalizaremos a matrices normales con matrices hermítica y unitario como casos especiales. Consideramos los casos físicamente importantes como el problema de los modos de vibraciones y el problema numérico importante de matrices patológicas.

Una matriz normal es una matriz que conmuta con su adjunta,

$$[A, A^\dagger] = 0 .$$

Ejemplos obvios e importante son las matrices hermíticas y unitarias. Mostraremos que las matrices normales tienen autovectores (ver tabla 6.1)

I. Sea $|x\rangle$ un autovector de A con correspondiente autovalor λ . Entonces

$$A|x\rangle = \lambda|x\rangle \quad (6.166)$$

Matriz	Autovalores	Autovectores (para diferentes autovalores)
Hermítica	Real	Ortogonal
Antihermítica	Imaginaria puro (o cero)	Ortogonal
Unitaria	Magnitud uno	Ortogonal
Normal	Si A tiene autovalor λ A^\dagger tiene autovalor λ^*	Ortogonal A y A^\dagger tienen los mismos autovectores

Cuadro 6.1:

o

$$(A - \lambda 1)|x\rangle = 0 . \quad (6.167)$$

Por conveniencia la combinación $A - \lambda 1$ la etiquetamos B . Tomando la adjunta de la ecuación (6.167), obtenemos

$$\langle x|(A - \lambda 1)^\dagger = 0 = \langle x|B^\dagger . \quad (6.168)$$

Porque

$$[(A - \lambda 1), (A - \lambda 1)^\dagger] = [A, A^\dagger] = 0 ,$$

tenemos

$$[B, B^\dagger] = 0 . \quad (6.169)$$

La matriz B es también normal.

A partir de las ecuaciones (6.167) y (6.168) formamos

$$\langle x|B^\dagger B|x\rangle = 0 . \quad (6.170)$$

Usando (6.169)

$$\langle x|BB^\dagger|x\rangle = 0 . \quad (6.171)$$

Ahora la ecuación (6.171) puede ser rescrita como

$$(B^\dagger|x\rangle)^\dagger(B^\dagger|x\rangle) = 0 . \quad (6.172)$$

Así

$$B^\dagger|x\rangle = (A^\dagger - \lambda^* 1)|x\rangle = 0 . \quad (6.173)$$

Vemos que para matrices normales, A^\dagger tiene los mismos autovectores que A pero los autovalores son los complejos conjugados.

II. Ahora, consideremos más que uno autovector-autovalor, tenemos

$$A|x_i\rangle = \lambda_i|x_i\rangle , \quad (6.174)$$

$$A|x_j\rangle = \lambda_j|x_j\rangle . \quad (6.175)$$

Multiplicando la ecuación (6.175) por la izquierda por $\langle x_i|$ produce

$$\langle x_i|A|x_j\rangle = \lambda_j\langle x_i|x_j\rangle . \quad (6.176)$$

Operando sobre el lado izquierdo de la ecuación (6.176), obtenemos

$$\langle x_i | \mathbf{A} = (\mathbf{A}^\dagger | x_i \rangle)^\dagger . \quad (6.177)$$

A partir de la ecuación (6.173) sabemos que \mathbf{A}^\dagger tiene los mismos autovectores que \mathbf{A} pero con los complejos conjugados de los autovalores

$$(\mathbf{A}^\dagger | x_i \rangle)^\dagger = (\lambda_i^* | x_i \rangle)^\dagger = \lambda_i \langle x_i | . \quad (6.178)$$

Sustituyendo en la ecuación (6.176) tenemos

$$\lambda_i \langle x_i | x_j \rangle = \lambda_j \langle x_i | x_j \rangle \quad (6.179)$$

o

$$(\lambda_i - \lambda_j) \langle x_i | x_j \rangle = 0 . \quad (6.180)$$

Esta es la misma que la ecuación (6.140).

Para $\lambda_i \neq \lambda_j$

$$\langle x_i | x_j \rangle = 0 .$$

Los autovectores correspondientes a diferentes autovalores de una matriz normal son *ortogonales*. Esto significa que una matriz normal puede ser diagonalizada por una transformación unitaria. La matriz unitaria requerida puede ser construida a partir de los vectores ortonormales como se mostró en la sección anterior.

El converso también es válido. Si \mathbf{A} puede ser diagonalizada por una transformación unitaria, entonces \mathbf{A} es normal.

Modos normales de vibración.

Consideremos las vibraciones de un modelo clásico de la molécula de CO_2 . Esta es una ilustración de la aplicación de las técnicas matriciales a un problema que no parte como un problema de matrices. También provee un ejemplo de autovalores y autovectores de una matriz real asimétrica.

Ejemplo: Modos Normales.

Consideremos tres masas sobre el eje x unidas por resortes como muestra la figura 6.6. Las fuerzas de los resortes se suponen lineales (para pequeños desplazamientos, ley de Hooke) y las masas se restringen a mantenerse sobre el eje x .

Usando una coordenada diferente para cada masa la segunda ley de Newton produce el conjunto de ecuaciones

$$\begin{aligned} \ddot{x}_1 &= -\frac{k}{M}(x_1 - x_2) \\ \ddot{x}_2 &= -\frac{k}{M}(x_2 - x_1) - \frac{k}{m}(x_2 - x_3) \\ \ddot{x}_3 &= -\frac{k}{M}(x_3 - x_2) . \end{aligned} \quad (6.181)$$

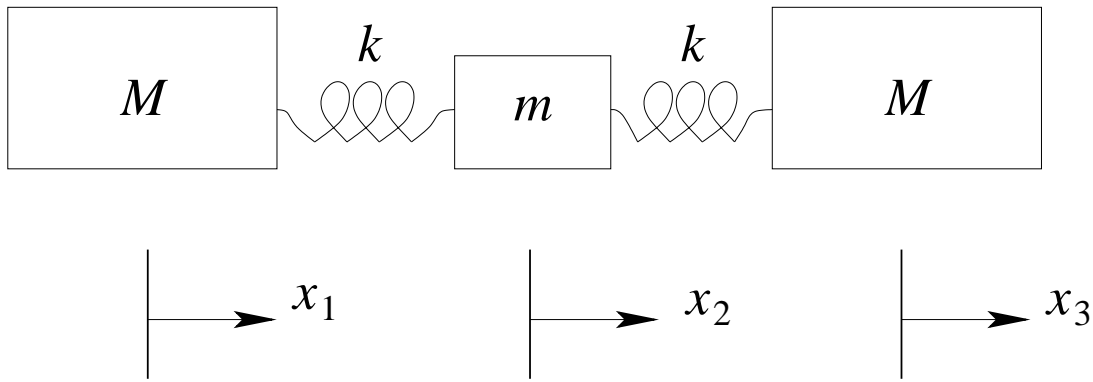


Figura 6.6: Vector fijo con coordenadas rotada.

El sistema de masa está vibrando. Buscamos las frecuencias comunes, ω tal que todas las masas vibren en esta misma frecuencia. Estos son los modos normales. Sea

$$x_i = x_{i0}e^{i\omega t}, \quad i = 1, 2, 3.$$

Substituyendo en la ecuación (6.181), podemos escribir este conjunto como

$$\begin{pmatrix} \frac{k}{M} & -\frac{k}{M} & 0 \\ -\frac{k}{m} & \frac{2k}{m} & -\frac{k}{m} \\ 0 & -\frac{k}{M} & \frac{k}{M} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \omega^2 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad (6.182)$$

dividiendo por el factor común $e^{i\omega t}$. Tenemos una ecuación matricial de autovalores con la matriz asimétrica. La ecuación secular es

$$\begin{vmatrix} \frac{k}{M} - \omega^2 & -\frac{k}{M} & 0 \\ -\frac{k}{m} & \frac{2k}{m} - \omega^2 & -\frac{k}{m} \\ 0 & -\frac{k}{M} & \frac{k}{M} - \omega^2 \end{vmatrix} = 0. \quad (6.183)$$

Esto conduce a

$$\omega^2 \left(\frac{k}{M} - \omega^2 \right) \left(\omega^2 - \frac{2k}{m} - \frac{k}{M} \right) = 0$$

Los autovalores son

$$\omega^2 = 0, \quad \frac{k}{M}, \quad \text{y} \quad \frac{k}{M} + \frac{2k}{m},$$

todas reales.

Los correspondientes autovectores son determinados sustituyendo los autovalores de regreso en la ecuación (6.182) un autovalor a la vez. Para $\omega^2 = 0$, ecuación (6.182) produce

$$\begin{aligned}x_1 - x_2 &= 0 \\ -x_1 + 2x_2 - x_3 &= 0 \\ -x_2 + x_3 &= 0 .\end{aligned}$$

Entonces, tenemos

$$x_1 = x_2 = x_3 .$$

Esto describe una translación pura sin movimiento relativo de las masas y sin vibración.

Para $\omega^2 = \frac{k}{M}$, la ecuación (6.182) produce

$$x_1 = -x_3 , \quad x_2 = 0 . \quad (6.184)$$

Las masas exteriores se mueven en direcciones opuestas. El masa del centro está estacionaria.

Para $\omega^2 = \frac{k}{M} + \frac{2k}{M}$, las componentes de los autovectores son

$$x_1 = x_3 , \quad x_2 = -\frac{2M}{m}x_1 .$$

Las dos masas exteriores se están moviendo juntas. La masa del centro se está moviendo opuesta a las otras dos. El momentum neto es cero.

Cualquier desplazamiento de estas tres masas a lo largo del eje x puede ser descrito como una combinación lineal de estos tres tipos de movimiento: translación más dos formas de vibración.

Sistemas con condiciones patológicas.

Un sistema lineal de ecuaciones puede ser escrito como

$$\mathbf{A}|x\rangle = |y\rangle \quad \text{o} \quad \mathbf{A}^{-1}|y\rangle = |x\rangle , \quad (6.185)$$

con \mathbf{A} y $|y\rangle$ conocido y $|x\rangle$ desconocido. Podemos encontrar ejemplos en los cuales un pequeño error en $|y\rangle$ resulta en un gran error en $|x\rangle$. En este caso la matriz \mathbf{A} es llamada de condición patológica. Si $|\delta x\rangle$ es el error en $|x\rangle$ y $|\delta y\rangle$ es el error en $|y\rangle$, entonces los errores relativos pueden ser escritos como

$$\left[\frac{\langle \delta x | \delta x \rangle}{\langle x | x \rangle} \right]^{1/2} \leq K(\mathbf{A}) \left[\frac{\langle \delta y | \delta y \rangle}{\langle y | y \rangle} \right]^{1/2} . \quad (6.186)$$

Aquí $K(\mathbf{A})$, una propiedad de la matriz \mathbf{A} , es etiquetado la *condición de número*. Para \mathbf{A} hermítica una forma de la condición de número es dada por

$$K(\mathbf{A}) = \frac{|\lambda|_{\max}}{|\lambda|_{\min}} . \quad (6.187)$$

Una forma aproximada debido a Turing es

$$K(\mathbf{A}) = n[A_{ij}]_{\max}[A_{ij}^{-1}]_{\max}, \quad (6.188)$$

en la cual n es el orden de la matriz y $[A_{ij}]_{\max}$ es el máximo elemento en \mathbf{A} .

Ejemplo: Una matriz patológica.

Un ejemplo común de una matriz con condición patológica es la matriz de Hilbert, la matriz de Hilbert de orden 4 es $H_{ij} = (i + j - 1)^{-1}$,

$$\mathbf{H}_4 = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix}. \quad (6.189)$$

Los elementos de la matriz inversa (orden n) son dados por

$$(\mathbf{H}_n^{-1})_{ij} = \frac{(-1)^{i+j}}{i + j \pm 1} \cdot \frac{(n + i - 1)!(n + j - 1)!}{[(i - 1)!(j - 1)!]^2(n - i)!(n - j)!}. \quad (6.190)$$

Para $n = 4$

$$\mathbf{H}_4^{-1} = \begin{pmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -140 & 1680 & -4200 & 2800 \end{pmatrix}. \quad (6.191)$$

A partir de la ecuación (6.188) la estimación de Turing de la condición de número para \mathbf{H}_4 llega a ser

$$K_{\text{Turing}} = 4 \times 1 \times 6480 \\ 2.59 \times 10^4 .$$

Esto es una advertencia de que un error en la entrada puede ser multiplicado por 26000 en el cálculo del resultado de salida. Esto sentencia que \mathbf{H}_4 tiene condición patológica. Si usted encuentra un sistema altamente patológico tiene un par de alternativas (además de abandonar el problema).

- Tratar un ataque matemático diferente.
- Hacer arreglos para llevar más cifras significativas y a costa de fuerza bruta empujar de principio a fin.

Capítulo 7

Teoría de grupo.

versión final 1.2-090502¹

Disciplined judgment about what is neat and symmetrical and elegant has time and time again proved an excellent guide to how nature work.

MURRAY GELL-MANN

7.1. Introducción.

En mecánica clásica la *simetría* de un sistema físico conduce a una *ley de conservación*. La conservación del momentum angular es una consecuencia directa de la simetría rotacional, lo cual significa *invariancia* bajo rotaciones espaciales. A principios del siglo pasado, Wigner y otros comprendieron que la invariancia era el concepto clave en el entendimiento de los nuevos fenómenos y en el desarrollo de teorías apropiadas. Así, en mecánica cuántica los conceptos de momento angular y spin han llegado a ser aún más centrales. Sus generalizaciones, el *isospin* en física nuclear y la simetría de *sabor* en física de partículas, son herramientas indispensables en la construcción teórica y en sus soluciones. Las generalizaciones del concepto de invariancia de *gauge* de la electrodinámica clásica para la simetría del isospin conduce a la teoría de gauge electrodébil.

En cada caso el conjunto de estas operaciones de simetría forman un *grupo*. La teoría de grupo es la herramienta matemática para tratar las invariancias y las simetrías. Ella trae consigo unificación y formalización de principios tales como reflexión espacial, o paridad, momento angular, y geometría que son ampliamente usados por los físicos.

En geometría el rol fundamental de la teoría de grupo fue reconocido hace mucho tiempo por los matemáticos. En geometría euclídea la distancia entre dos puntos, el producto escalar de dos vectores o métrica, no cambia bajo rotaciones o translaciones. Estas simetrías son características de esta geometría. En relatividad especial la métrica, o producto escalar de cuadvectores, difiere del de la geometría euclídea en que ya no es más positivo definido y es invariante ante transformaciones de Lorentz.

¹Este capítulo está basado en el cuarto capítulo del libro: *Mathematical Methods for Physicists, fourth edition* de George B. Arfken & Hans J. Weber, editorial ACADEMIC PRESS.

Para un cristal el grupo de simetría contiene sólo un número finito de rotaciones en valores discretos del ángulo y reflexiones. La teoría de tales grupos discretos o finitos, desarrollada inicialmente como una rama de las matemáticas pura, ahora es una útil herramienta para el desarrollo de la cristalografía y la física de la materia condensada. Haremos una breve introducción a ellos. Cuando las rotaciones dependen de un ángulo continuo el grupo de rotaciones tiene un número infinito de elementos. Estudiaremos tales grupos continuos o de Lie.

Definición de grupo.

Un grupo G puede ser definido como un conjunto de objetos u operaciones, llamados los elementos de G , que pueden ser combinados o “multiplicados” para formar un producto bien definido en G el cual satisface las siguientes cuatro condiciones.

1. Si a y b son cualquier par de elementos de G , entonces el producto ab es también elemento de G ; o $(a, b) \rightarrow ab$ mapea $G \times G$ sobre G .
2. Esta multiplicación es asociativa, $(ab)c = a(bc)$.
3. Hay un elemento unidad o neutro I en G tal que $Ia = aI = a$ para cada elemento a de G .²
4. Debe haber un inverso o recíproco de cada elemento a de G , etiquetado a^{-1} , tal que $aa^{-1} = a^{-1}a = I$.

Un ejemplo de grupo es el conjunto de rotaciones de coordenadas en el sentido del puntero del reloj,

$$R(\varphi) = \begin{pmatrix} \cos \varphi & \text{sen } \varphi \\ -\text{sen } \varphi & \cos \varphi \end{pmatrix} \quad (7.1)$$

en un ángulo φ del sistema de coordenadas xy a una nueva orientación. El producto de dos rotaciones $R(\varphi_1)R(\varphi_2)$ es definida como una rotación primero en un ángulo φ_2 y entonces en un ángulo φ_1 . De acuerdo a la ecuación (6.29), esto corresponde al producto de las dos matrices ortogonales de 2×2

$$\begin{pmatrix} \cos \varphi_1 & \text{sen } \varphi_1 \\ -\text{sen } \varphi_1 & \cos \varphi_1 \end{pmatrix} \begin{pmatrix} \cos \varphi_2 & \text{sen } \varphi_2 \\ -\text{sen } \varphi_2 & \cos \varphi_2 \end{pmatrix} = \begin{pmatrix} \cos(\varphi_1 + \varphi_2) & \text{sen}(\varphi_1 + \varphi_2) \\ -\text{sen}(\varphi_1 + \varphi_2) & \cos(\varphi_1 + \varphi_2) \end{pmatrix}, \quad (7.2)$$

usando las fórmulas de adición de funciones trigonométricas. El producto es claramente una rotación representada por una matriz ortogonal con un ángulo $\varphi_1 + \varphi_2$. El producto es la multiplicación asociativa de matrices. Es *conmutativo* o *abeliano* porque el orden en el cual estas rotaciones son realizadas no importa. El inverso de la rotación con ángulo φ es una con ángulo $-\varphi$. La unidad o neutro corresponde al ángulo $\varphi = 0$. El nombre del grupo es $SO(2)$, si el ángulo varía continuamente desde 0 a 2π . Claramente, $SO(2)$ tiene infinitos elementos. La unidad con ángulo $\varphi = 0$ y la rotación con $\varphi = \pi$ forman un *subgrupo* finito. Un subgrupo G' de un grupo G consiste de elementos de G tal que el producto de cualquiera de sus elementos

²También etiquetan al elemento unidad como E .

está de nuevo en el subgrupo G' , *i.e.*, G' es cerrado bajo la multiplicación de G . Si $gg'g^{-1}$ es un elemento de G' para cualquier g de G y g' de G' , entonces G' es llamado un *subgrupo invariante* de G .

Las matrices ortogonales $n \times n$ forman el grupo $O(n)$, y también $SO(n)$ si sus determinantes son $+1$ (S por eSpecial). Si $\tilde{O}_i = O_i^{-1}$ para $i = 1$ y 2 , entonces el producto

$$\widetilde{O_1 O_2} = \tilde{O}_2 \tilde{O}_1 = O_1^{-1} O_2^{-1} = (O_1 O_2)^{-1}$$

es también una matriz ortogonal en $SO(n)$. La inversa es la matriz (ortogonal) transpuesta. La unidad del grupo es 1_n . Una matriz real ortogonal de $n \times n$ tiene $n(n-1)/2$ parámetros independientes. Para $n = 2$ hay sólo un parámetro: un ángulo en la ecuación (7.1). Para $n = 3$, hay tres parámetros independientes: los tres ángulos de Euler de la sección 6.3.

De la misma manera, las matrices unitarias de $n \times n$ forman el grupo $U(n)$, y también $SU(n)$ si sus determinantes son $+1$. Si $U_i^\dagger = U_i^{-1}$, entonces

$$(U_1 U_2)^\dagger = U_2^\dagger U_1^\dagger = U_2^{-1} U_1^{-1} = (U_1 U_2)^{-1},$$

tal que el producto es unitario y un elemento de $SU(n)$. Cada matriz unitaria tiene una inversa la cual es también unitaria.

Homomorfismo, isomorfismo.

Puede haber una correspondencia entre los elementos de dos grupos (o entre dos representaciones), uno a uno, dos a uno, o muchos a uno. Si esta correspondencia preserva la multiplicación del grupo, diremos que los dos grupos son *homomórficos*. Una de las más importantes correspondencias homomórficas entre el grupo de rotaciones $SO(3)$ y el grupo de matrices unitarias $SU(2)$ será desarrollado en la próxima sección. Si la correspondencia es uno a uno, y aún preserva la multiplicación del grupo,³ entonces los grupos son *isomórficos*. Un ejemplo es las rotaciones de coordenadas a través de un ángulo finito φ en el sentido horario respecto al eje z en el espacio tridimensional descrito por

$$R_z(\varphi) = \begin{pmatrix} \cos \varphi & \text{sen } \varphi & 0 \\ -\text{sen } \varphi & \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (7.3)$$

El grupo de rotaciones R_z es isomórfico al grupo de rotaciones en la ecuación (7.1).

Representaciones matriciales, reducibles e irreducibles.

La representación de los elementos de un grupo por matrices es una técnica muy poderosa y ha sido casi universalmente adoptada por los físicos. El uso de matrices no impone restricciones significativas. Puede mostrarse que los elementos de cualquier grupo finito y de grupos continuos pueden ser representados por matrices. Ejemplos son las rotaciones descritas en la ecuación (7.1) y (7.3).

³Supongamos que los elementos del primer grupo son etiquetados por g_i y los elementos del segundo grupo por h_i . Entonces $g_i \leftrightarrow h_i$ en una correspondencia uno a uno para todos los valores de i . Si $g_i g_j = g_k$ y $h_i h_j = h_k$, entonces g_k y h_k deben ser los elementos correspondientes del grupo.

Para ilustrar como las representaciones matriciales surgen a partir de una simetría, consideremos la ecuación estacionaria de Schrödinger (o algún otra ecuación de autovalores tal como $Iv_i = I_i v_i$ para los momentos principales de inercia de un cuerpo rígido en mecánica clásica)

$$H\psi = E\psi . \quad (7.4)$$

Supongamos que la ecuación (7.4) se mantiene invariante bajo la acción de un grupo G de transformaciones R en G (rotaciones de coordenadas, por ejemplo, para un potencial central $V(r)$ en el Hamiltoniano H), *i.e.*,

$$H_R = RHR^{-1} = H . \quad (7.5)$$

Ahora tomamos una solución ψ de la ecuación (7.4) y la “rotamos”: $\psi \rightarrow R\psi$. Entonces $R\psi$ tiene la misma energía E porque multiplicando la ecuación (7.4) por R y usando (7.5) produce

$$RH\psi = E(R\psi) = (RHR^{-1})R\psi = H(R\psi) . \quad (7.6)$$

En otras palabras, todas las soluciones rotadas $R\psi$ son *degeneradas* en energía o forman lo que los físicos llaman un *multiplete*. Supongamos que este espacio vectorial V_ψ de soluciones transformadas tiene una dimensión finita n . Sean $\psi_1, \psi_2, \dots, \psi_n$ una base. Ya que $R\psi_j$ es un miembro del multiplete, podemos expandirlo en términos de esta base

$$R\psi_j = \sum_k r_{jk} \psi_k . \quad (7.7)$$

Así, cada R en G puede ser asociado a una matriz (r_{jk}) , y este mapeo $R \rightarrow (r_{jk})$ es llamada una representación de G . Si podemos tomar cualquier elemento de V_ψ y por rotaciones con todos los elementos de R de G transforman en todos los otros elementos de V_ψ entonces la representación es *irreducible*. Si todos los elementos de V_ψ no son alcanzados, entonces V_ψ se separa en una suma directa de dos o más subespacios vectoriales, $V_\psi = V_1 + V_2 + \dots$, los cuales son mapeados dentro de ellos mismos por rotación de sus elementos. En este caso la representación es llamada *reducible*. Entonces podemos encontrar una base en V_ψ (*i.e.*, hay una matriz unitaria U) tal que

$$U(r_{jk})U^\dagger = \begin{pmatrix} r_1 & 0 & \dots \\ 0 & r_2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (7.8)$$

para todos los R de G y todas las matrices (r_{jk}) . Aquí r_1, r_2, \dots , son matrices de menor dimensiones que (r_{jk}) que están alineadas a lo largo de la diagonal y los 0 son matrices de ceros. podemos decir que R ha sido descompuestas en $r_1 + r_2 + \dots$ en paralelo con $V_\psi = V_1 \oplus V_2 \oplus \dots$

Las representaciones irreducibles juegan un rol en teoría de grupo que es aproximadamente análogo a los vectores unitarios en el análisis vectorial. Ellas son las representaciones más simples, toda otra puede ser construida desde ellas.

7.2. Generadores de grupos continuos.

Una característica de los grupos continuos conocidos como grupos de Lie es que los parámetros de un producto de elementos son funciones analíticas de los parámetros de los factores. La naturaleza analítica de las funciones nos permite desarrollar el concepto de *generador* y reduce el estudio del grupo completo a un estudio de los elementos del grupo en la vecindad del elemento identidad.

La idea esencial de Lie fue el estudio de elementos R en un grupo G que estén infinitesimalmente cercanos a la unidad de G . Consideremos el grupo $SO(2)$ como un ejemplo simple. Las matrices de rotación de 2×2 en la ecuación (7.1) puede ser escrita en forma exponencial usando la identidad de Euler ecuación (6.161) como

$$R(\varphi) = \begin{pmatrix} \cos \varphi & \text{sen } \varphi \\ -\text{sen } \varphi & \cos \varphi \end{pmatrix} = \mathbf{1}_2 \cos \varphi + i\sigma_2 \text{sen } \varphi = \exp(i\sigma_2 \varphi) . \quad (7.9)$$

A partir de la forma exponencial es obvio que la multiplicación de estas matrices es equivalente a la suma de los argumentos

$$R(\varphi_2)R(\varphi_1) = \exp(i\sigma_2 \varphi_2) \exp(i\sigma_2 \varphi_1) = \exp(i\sigma_2(\varphi_1 + \varphi_2)) = R(\varphi_1 + \varphi_2) .$$

Por supuesto las rotaciones cercanas a 1 tienen un ángulo pequeño $\varphi \sim 0$.

Esto sugiere que busquemos una representación exponencial

$$R = \exp(i\varepsilon S) , \quad \varepsilon \rightarrow 0 , \quad (7.10)$$

para elementos del grupo $R \in G$ cercanos a la 1. Las transformaciones infinitesimales S son llamadas los *generadores* de G . Ellos forman un espacio lineal cuya dimensión es el *orden* de G porque la multiplicación de los elementos R del grupo se traduce en la suma de los generadores S .

Si R no cambia el elemento de volumen, *i.e.*, $\det(R) = 1$, nosotros usamos la ecuación (6.164) para ver que

$$\det(R) = \exp(\text{tr}(\ln R)) = \exp(i\varepsilon \text{tr}(S)) = 1$$

implica que los generadores son de traza nula,

$$\text{tr}(S) = 0 . \quad (7.11)$$

Este es el caso para el grupo de rotaciones $SO(n)$ y el grupo unitario $SU(n)$, como veremos más adelante.

Si R de G en la ecuación (7.1) es unitario, entonces $S^\dagger = S$ es hermítica, lo cual también es el caso para $SO(n)$ y $SU(n)$. Ya que hay un i extra en la ecuación (7.10).

Expandamos los elementos del grupo

$$\begin{aligned} R_i &= \exp(i\varepsilon_i S_i) = \mathbf{1} + i\varepsilon_i S_i - \frac{1}{2}\varepsilon_i^2 S_i^2 + \dots , \\ R_i^{-1} &= \exp(-i\varepsilon_i S_i) = \mathbf{1} - i\varepsilon_i S_i - \frac{1}{2}\varepsilon_i^2 S_i^2 + \dots , \end{aligned} \quad (7.12)$$

a segundo orden en el pequeño parámetro del grupo ε_i porque los términos lineales y varios términos cuadráticos se cancelan en el producto (figura 7.1)

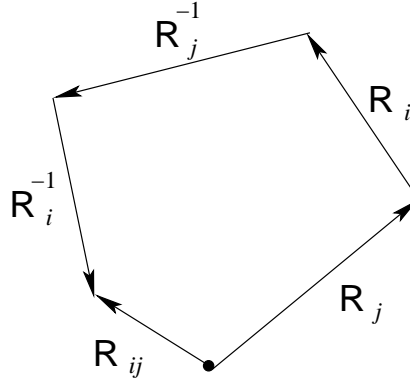


Figura 7.1: Ilustración de la ecuación (7.13).

$$\begin{aligned} R_i^{-1}R_j^{-1}R_iR_j &= 1 + \varepsilon_i\varepsilon_j[S_j, S_i] + \dots, \\ &= 1 + \varepsilon_i\varepsilon_j \sum_k c_{ji}^k S_k + \dots, \end{aligned} \quad (7.13)$$

cuando las ecuaciones (7.12) son sustituidas dentro de la ecuación (7.13). La última línea es debido a que el producto en la ecuación (7.13) es nuevamente un elemento, R_{ij} cercano a la unidad en el grupo G . Por tanto su exponente debe ser una combinación lineal de los generadores S_k y sus parámetros infinitesimales del grupo tienen que ser proporcionales al producto $\varepsilon_i\varepsilon_j$. Comparando ambas líneas (7.13) encontramos la relación de *clausura* de los generadores del grupo de Lie G ,

$$[S_i, S_j] = \sum_k c_{ij}^k S_k \quad (7.14)$$

Los coeficientes c_{ij}^k son las *constantes de estructura* del grupo G . Ya que el conmutador en la ecuación (7.14) es antisimétrico en i y en j , también lo son las constantes de estructura en los índices inferiores,

$$c_{ij}^k = -c_{ji}^k. \quad (7.15)$$

Si el conmutador en la ecuación (7.14) es tomado como la *regla de multiplicación de los generadores*, vemos que el espacio vectorial de los generadores llega a ser un álgebra, el álgebra de Lie G del grupo G . Para $SU(l+1)$ el álgebra de Lie es llamada A_l , para $SO(2l+1)$ es B_l y para $SO(2l)$ es D_l , donde $l = 1, 2, \dots$ es un entero positivo, esto será llamado el *rango* de grupo de Lie G o de su álgebra G .

Finalmente, la *identidad de Jacobi* se satisface para los dobles conmutadores

$$[[S_i, S_j], S_k] + [[S_j, S_k], S_i] + [[S_k, S_i], S_j] = 0, \quad (7.16)$$

lo cual es fácilmente verificable usando la definición de conmutador. Cuando la ecuación (7.14) es substituida en (7.16) encontramos otra restricción sobre las constantes de estructura,

$$\sum_m \{c_{ij}^m [S_m, S_k] + c_{jk}^m [S_m, S_i] + c_{ki}^m [S_m, S_j]\} = 0. \quad (7.17)$$

Usando de nuevo la ecuación (7.14) en la ecuación (7.17) implica que

$$\sum_{mn} \{c_{ij}^m c_{mk}^n S_n + c_{jk}^m c_{mi}^n S_n + c_{ki}^m c_{mj}^n S_n\} = 0, \quad (7.18)$$

donde el factor común S_n (y la suma sobre n) pueden eliminarse por que los generadores son linealmente independientes. Por tanto

$$\sum_m \{c_{ij}^m c_{mk}^n + c_{jk}^m c_{mi}^n + c_{ki}^m c_{mj}^n\} = 0. \quad (7.19)$$

Las relaciones (7.14), (7.15) y (7.19) forman la base de las álgebras de Lie desde la cual los elementos finitos del grupo de Lie cerca de su unidad puede ser reconstruído.

Volviendo a la ecuación (7.5), el inverso de R es exactamente $R^{-1} = \exp(-i\varepsilon S)$. expandimos H_R de acuerdo a la fórmula de Baker-Haudorff, ecuación (6.17),

$$H = H_R = \exp(i\varepsilon S)H \exp(-i\varepsilon S) = H + i\varepsilon[S, H] - \varepsilon^2 \frac{[S, [iS, H]]}{2!} + \dots. \quad (7.20)$$

Al simplificar H de la ecuación (7.20), dividiendo por ε y haciendo $\varepsilon \rightarrow 0$. Entonces la ecuación (7.20) implica que para cualquier rotación cercana a 1 en G el conmutador

$$[S, H] = 0. \quad (7.21)$$

Si S y H son matrices hermíticas, la ecuación (7.21) dice que S y H pueden ser simultaneamente diagonalizados. Si S y H son operadores diferenciales como el Hamiltoniano y el momento angular orbital en mecánica cuántica, entoces la ecuación (7.21) dice que S y H tienen autofunciones en común y que los autovalores degenerados de H pueden ser distinguidos por los autovalores de los generadores S . Esta es con mucho la más importante aplicación de teoría de grupos a mecánica cuántica.

A continuación, estudiaremos los grupos ortogonales y unitarios como ejemplos.

Grupos de rotaciones $SO(2)$ y $SO(3)$.

Para $SO(2)$ definido por la ecuación (7.1) hay sólo un generador linealmente independiente, σ_2 y el orden de $SO(2)$ es 1. Obtenemos σ_2 a partir de diferenciar la ecuación (7.9) y evaluarla en cero,

$$-i \frac{dR(\varphi)}{d\varphi} \Big|_{\varphi=0} = -i \begin{pmatrix} -\text{sen } \varphi & \cos \varphi \\ -\cos \varphi & -\text{sen } \varphi \end{pmatrix} \Big|_{\varphi=0} = -i \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \sigma_2. \quad (7.22)$$

Para las rotaciones $R_z(\varphi)$ sobre el eje z descritas por la ecuación (7.3), el generador es dado por

$$-i \frac{dR(\varphi)}{d\varphi} \Big|_{\varphi=0} = S_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (7.23)$$

donde el factor extra i es insertado para hacer S_z hermítica. La rotación $R_z(\delta\varphi)$ en un ángulo infinitesimal $\delta\varphi$ puede ser escrita como

$$R_z(\delta\varphi) = 1_3 + i\delta\varphi S_z, \quad (7.24)$$

Una expansión de Maclaurin-Taylor de R_z cerca de la unidad $\varphi = 0$ con términos hasta orden $(\delta\varphi)^2$ y los superiores son despreciados. Una rotación finita puede ser compuesta por sucesivas rotaciones infinitesimales

$$R_z(\delta\varphi_1 + \delta\varphi_2) = (\mathbf{1}_3 + i\delta\varphi_1\mathbf{S}_z)(\mathbf{1}_3 + i\delta\varphi_2\mathbf{S}_z) . \quad (7.25)$$

Sea $\delta\varphi = \varphi/N$ para N rotaciones, con $N \rightarrow \infty$. Entonces,

$$R_z(\varphi) = \lim_{N \rightarrow \infty} \left[\mathbf{1}_3 + i\frac{\varphi}{N}\mathbf{S}_z \right]^N = \exp(i\mathbf{S}_z) . \quad (7.26)$$

Esta forma identifica \mathbf{S}_z como el generador del grupo R_z , un subgrupo abeliano de $SO(3)$, el grupo de rotaciones en tres dimensiones con determinante +1. Cada matriz de 3×3 $R_z(\varphi)$ es ortogonal, por lo tanto unitaria, y la $\text{tr}(\mathbf{S}_z) = 0$ de acuerdo con la ecuación (7.11).

Por diferenciación de las rotaciones de coordenadas

$$R_x(\psi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & \text{sen } \psi \\ 0 & -\text{sen } \psi & \cos \psi \end{pmatrix} , \quad R_y(\theta) = \begin{pmatrix} \cos \theta & 0 & -\text{sen } \theta \\ 0 & 1 & 0 \\ \text{sen } \theta & 0 & \cos \theta \end{pmatrix} , \quad (7.27)$$

obtenemos los generadores

$$\mathbf{S}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} , \quad \mathbf{S}_y = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} , \quad (7.28)$$

de R_x y R_y , los subgrupos de rotaciones en torno a los ejes x e y respectivamente.

Rotaciones de funciones y momento angular orbital.

En la discusión precedente los elementos del grupos son matrices que rotan las coordenadas. Cualquier sistema físico que esta siendo descrito se mantiene fijo. Ahora mantengamos fijas las coordenadas y rotemos una función $\psi(x, y, z)$ relativa a nuestras coordenadas fijas. Con R para rotar las coordenadas,

$$\vec{x}' = R\vec{x} , \quad (7.29)$$

definimos R por

$$R\psi(x, y, z) = \psi'(x, y, z) \rightarrow \psi(\vec{x}') . \quad (7.30)$$

En palabras, la matriz R opera sobre la función ψ , creando una *nueva función* ψ' que es numéricamente igual a $\psi(\vec{x}')$, donde \vec{x}' son las coordenadas rotadas por R . Si R rota las coordenadas en el sentido horario, el efecto de la matriz R es rotar el modelo de la función ψ en el sentido horario.

Volviendo a las ecuaciones (7.3) y (7.28), consideremos una rotación infinitesimal, $\varphi \rightarrow \delta\varphi$. Luego, usando R_z , ecuación (7.3), obtenemos

$$R_z(\delta\varphi)\psi(x, y, z) = \psi(x + y\delta\varphi, y - x\delta\varphi, z) . \quad (7.31)$$

El lado derecho puede ser expandido como una serie de Taylor de primer orden en $\delta\varphi$ para dar

$$\begin{aligned} R_z(\delta\varphi)\psi(x, y, z) &= \psi(x, y, z) - \delta\varphi \left\{ x \frac{\partial\psi}{\partial y} - y \frac{\partial\psi}{\partial x} \right\} + \mathcal{O}(\delta\varphi)^2 \\ &= (1 - i\delta\varphi L_z)\psi(x, y, z) , \end{aligned} \quad (7.32)$$

la expresión diferencial en el paréntesis de llave es iL_z . Ya que una rotación primero en φ y luego en $\delta\varphi$ alrededor del eje z está dado por

$$R_z(\varphi + \delta\varphi)\psi(x, y, z) = R_z(\delta\varphi)R_z(\varphi)\psi(x, y, z) = (1 - i\delta\varphi L_z)R_z(\varphi)\psi(x, y, z) , \quad (7.33)$$

tenemos (como una ecuación de operadores)

$$\frac{R_z(\varphi + \delta\varphi) - R_z(\varphi)}{\delta\varphi} = -iL_z R_z(\varphi) . \quad (7.34)$$

El lado izquierdo es justo $dR_z(\varphi)/d\varphi$ (para $\delta\varphi \rightarrow 0$). En esta forma la ecuación (7.34) se integra inmediatamente a

$$R_z(\varphi) = \exp(-i\varphi L_z) . \quad (7.35)$$

Note cuidadosamente que $R_z(\varphi)$ rota funciones (en el sentido horario) relativa a las coordenadas fijadas y que L_z es la componente z del momento angular orbital \vec{L} . La constante de integración está fijada por la condición de borde $R_z(0) = 1$.

Si reconocemos que los elementos de matriz

$$L_z = (x, y, z) \mathbf{S}_z \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} , \quad (7.36)$$

claramente L_x, L_y, L_z satisface la misma relación de conmutación

$$[L_i, L_j] = i\varepsilon_{ijk} L_k \quad (7.37)$$

que S_x, S_y, S_z y tienen a la misma constantes de estructura $i\varepsilon_{ijk}$ de $\text{SO}(3)$.

Homomorfismo $\text{SU}(2)$ - $\text{SO}(3)$.

El grupo *unitario especial* $\text{SU}(2)$ de matrices unitarias de 2×2 con determinante $+1$ tiene las tres matrices de Pauli σ_i como generadores (mientras que las rotaciones de la ecuación (7.3) forman un subgrupo abeliano unidimensional). Por lo tanto $\text{SU}(2)$ es de orden 3 y depende de tres parámetros continuos reales ξ, η y ζ los cuales a menudo son llamados los parámetros de Caley-Klein. Sus elementos generales tienen la forma

$$U_2(\xi, \eta, \zeta) = \begin{pmatrix} e^{i\xi} \cos \eta & e^{i\zeta} \sin \eta \\ -e^{-i\zeta} \sin \eta & e^{-i\xi} \cos \eta \end{pmatrix} = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} . \quad (7.38)$$

Es fácil chequear que el determinante $\det(\mathbf{U}_2) = 1$ y que $\mathbf{U}_2^\dagger \mathbf{U}_2 = \mathbf{1} = \mathbf{U}_2 \mathbf{U}_2^\dagger$ se mantiene.

Para obtener los generadores diferenciamos

$$\begin{aligned} -i \frac{\partial \mathbf{U}_2}{\partial \xi} \Big|_{\xi=0, \eta=0} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \sigma_3, \\ \frac{-i}{\text{sen } \eta} \frac{\partial \mathbf{U}_2}{\partial \zeta} \Big|_{\zeta=0} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sigma_1, \\ -i \frac{\partial \mathbf{U}_2}{\partial \eta} \Big|_{\zeta=0, \eta=0} &= \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} = \sigma_2. \end{aligned} \quad (7.39)$$

Por supuesto, las matrices de Pauli son todas de traza nula y hermíticas.

Con las matrices de Pauli como generadores de los elementos (U_1, U_2, U_3) de $SU(2)$ pueden ser generados por

$$U_1 = \exp(ia_1 \sigma_1/2), \quad U_2 = \exp(ia_2 \sigma_2/2), \quad U_3 = \exp(ia_3 \sigma_3/2). \quad (7.40)$$

Los tres parámetros a_i son reales. El factor extra $1/2$ está presente en los exponentes ya que $s_i = \sigma_i/2$ satisface las mismas relaciones de conmutación ⁴

$$[s_i, s_j] = i \varepsilon_{ijk} s_k \quad (7.41)$$

como el momento angular en la ecuación (7.37).

La ecuación (7.3) da un operador de rotación para rotar las coordenadas cartesianas en el espacio tridimensional. Usando la matriz de momento angular s_3 , tenemos el correspondiente operador de rotación en el espacio de dos dimensiones (complejo) $R_z(\varphi) = \exp(i\varphi s_3/2)$.

Para rotar una función de onda vectorial de dos componentes (spinor) o una partícula de spin $1/2$ relativa a coordenadas fijas, el operador de rotación es $R_z(\varphi) = \exp(-i\varphi s_3/2)$ de acuerdo a la ecuación (7.35).

Usando la ecuación (7.40) la identidad de Euler, la ecuación (6.161), obtenemos

$$\mathbf{U}_j = \cos\left(\frac{a_j}{2}\right) + i \sigma_j \text{sen}\left(\frac{a_j}{2}\right).$$

Aquí el parámetro a_j aparece como un ángulo, el coeficiente de una matriz tipo momento angular φ en la ecuación (7.26). Con esta identificación de los exponentes, la forma general de la matriz $SU(2)$ (para rotar funciones más que las coordenadas) podría ser escrita como

$$\mathbf{U}(\alpha, \beta, \gamma) = \exp(-i\gamma s_3/2) \exp(-i\beta s_2/2) \exp(-i\alpha s_1/2).$$

Como vimos, los elementos de $SU(2)$ describen rotaciones en un espacio complejo bidimensional que deja invariante a $|z_1|^2 + |z_2|^2$. El determinante es $+1$. Hay tres parámetros independientes. Nuestro grupo ortogonal real $SO(3)$ de determinante $+1$, claramente describe rotaciones comunes en el espacio tridimensional con la importante característica de dejar invariante a $x^2 + y^2 + z^2$. También hay tres parámetros independientes. Las interpretaciones de rotación y la igualdad de números de parámetros sugiere la existencia de alguna clase de correspondencia entre los grupos $SU(2)$ y $SO(3)$. Aquí desarrollamos esta correspondencia.

⁴Las constantes de estructuras $(i\varepsilon_{ijk})$ conducen a las representaciones de $SU(2)$ de dimensión $2J = 1$ para generadores de dimensión $2j + 1$, con $J = 0, 1/2, 1, \dots$. Los casos con J entero también conducen a las representaciones de $SO(3)$.

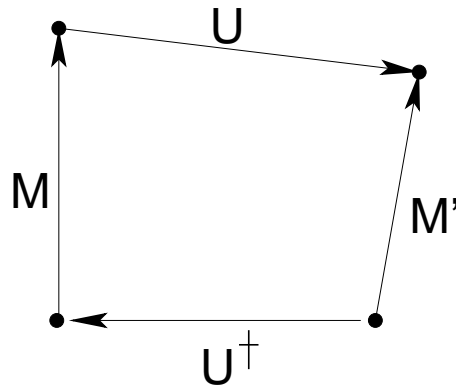


Figura 7.2: Ilustración de $M' = U M U^\dagger$ ecuación (7.42).

La operación $SU(2)$ sobre una matriz está dada por una transformación unitaria, la ecuación (7.5), con $R = U$ y la figura (7.2)

$$M' = U M U^\dagger . \quad (7.42)$$

Tomando M como una matriz de 2×2 , notemos que cualquier matriz de 2×2 puede ser escrita como una combinación lineal de la matriz unidad y las tres matrices de Pauli. Sea M la matriz de traza cero,

$$M = x\sigma_1 + y\sigma_2 + z\sigma_3 = \begin{pmatrix} z & x - iy \\ x + iy & -z \end{pmatrix} , \quad (7.43)$$

la matriz unidad no entra. Ya que la traza es invariante bajo transformaciones unitarias, M' debería tener la misma forma,

$$M = x'\sigma_1 + y'\sigma_2 + z'\sigma_3 = \begin{pmatrix} z' & x' - iy' \\ x' + iy' & -z' \end{pmatrix} . \quad (7.44)$$

El determinante también es invariante bajo una transformación unitaria. Por lo tanto

$$-(x^2 + y^2 + z^2) = -(x'^2 + y'^2 + z'^2) , \quad (7.45)$$

o $(x^2 + y^2 + z^2)$ es invariante bajo esta operación de $SU(2)$, como con $SO(3)$. $SU(2)$ debe, por lo tanto, describir una rotación. Esto sugiere que $SU(2)$ y $SO(3)$ pueden ser isomórficos o homomórficos.

Aproximemos el problema de qué rotación describe $SU(2)$ considerando casos especiales. Retomando la ecuación (7.38) sea $a = e^{i\xi}$ y $b = 0$, o

$$U_z = \begin{pmatrix} e^{i\xi} & 0 \\ 0 & e^{-i\xi} \end{pmatrix} . \quad (7.46)$$

En anticipación de la ecuación (7.50), esta U le está dado un subíndice z .

Realizando una transformación unitaria sobre cada una de las tres matrices de Pauli, tenemos

$$\begin{aligned} U_z \sigma_1 U_z^\dagger &= \begin{pmatrix} e^{i\xi} & 0 \\ 0 & e^{-i\xi} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} e^{-i\xi} & 0 \\ 0 & e^{i\xi} \end{pmatrix} \\ &= \begin{pmatrix} 0 & e^{2i\xi} \\ e^{-2i\xi} & 0 \end{pmatrix} . \end{aligned} \quad (7.47)$$

Reexpresamos este resultado en términos de las matrices de Pauli para obtener

$$U_z x \sigma_1 U_z^\dagger = x \cos 2\xi \sigma_1 - x \sin 2\xi \sigma_2 . \quad (7.48)$$

Similarmente,

$$\begin{aligned} U_z y \sigma_2 U_z^\dagger &= y \sin 2\xi \sigma_1 - y \cos 2\xi \sigma_2 , \\ U_z z \sigma_3 U_z^\dagger &= z \sigma_3 . \end{aligned} \quad (7.49)$$

A partir de esta expresión de doble ángulo vemos que podríamos comenzar con el ángulo medio: $\xi = \alpha/2$. Entonces, de las ecuaciones (7.42)–(7.44), (7.48) y (7.49),

$$\begin{aligned} x' &= x \cos \alpha + y \sin \alpha \\ y' &= -x \sin \alpha + y \cos \alpha \\ z' &= z . \end{aligned} \quad (7.50)$$

La transformación unitaria de 2×2 usando $U_z(\alpha/2)$ es equivalente al operador de rotación $R(\alpha)$ de la ecuación (7.3).

El establecimiento de la correspondencia de

$$U_y(\beta/2) = \begin{pmatrix} \cos \beta/2 & \sin \beta/2 \\ -\sin \beta/2 & \cos \beta/2 \end{pmatrix} \quad (7.51)$$

y $R_y(\beta)$ y de

$$U_x(\varphi/2) = \begin{pmatrix} \cos \varphi/2 & i \sin \varphi/2 \\ i \sin \varphi/2 & \cos \varphi/2 \end{pmatrix} \quad (7.52)$$

y $R_x(\varphi)$ pueden calcularse como ejercicio. Podemos notar que $U_k(\psi/2)$ tiene la forma general

$$U_k(\psi/2) = 1 \cos \psi/2 + i \sigma_k \sin \psi/2 , \quad (7.53)$$

donde $k = x, y, z$.

La correspondencia

$$U_z\left(\frac{\alpha}{2}\right) = \begin{pmatrix} e^{i\alpha/2} & 0 \\ 0 & e^{-i\alpha/2} \end{pmatrix} \leftrightarrow \begin{pmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} = R_z(\alpha) , \quad (7.54)$$

no es una simple correspondencia uno a uno. Específicamente, como α en R_z recorre desde 0 a 2π , el parámetro U_z , $\alpha/2$, recorre desde 0 a π . Encontramos que

$$\begin{aligned} R_z(\alpha + 2\pi) &= R_z(\alpha) \\ U_z(\alpha/2 + \pi) &= \begin{pmatrix} -e^{i\alpha/2} & 0 \\ 0 & -e^{-i\alpha/2} \end{pmatrix} = -U_z(\alpha/2) . \end{aligned} \quad (7.55)$$

Por lo tanto *ambos* $U_z(\alpha/2)$ y $U_z(\alpha/2+\pi) = -U_z(\alpha/2)$ corresponde a $R_z(\alpha)$. La correspondencia es de 2 a 1, o $SU(2)$ y $SO(3)$ son homomórficos. Este establecimiento de la correspondencia entre las representaciones de $SU(2)$ y de aquella $SO(3)$ significa que las representaciones coincidas de $SU(2)$ automáticamente nos proporciona de las representaciones de $SO(3)$.

Combinando las rotaciones, encontramos que una transformación unitaria usada

$$U(\alpha, \beta, \gamma) = U_z(\gamma/2)U_y(\beta/2)U_z(\alpha/2) , \quad (7.56)$$

corresponde a la rotación general de Euler $R_z(\gamma)R_y(\beta)R_z(\alpha)$. Por multiplicación directa,

$$\begin{aligned} U(\alpha, \beta, \gamma) &= \begin{pmatrix} e^{i\gamma/2} & 0 \\ 0 & e^{-i\gamma/2} \end{pmatrix} \begin{pmatrix} \cos \beta/2 & \text{sen } \beta/2 \\ -\text{sen } \beta/2 & \cos \beta/2 \end{pmatrix} \begin{pmatrix} e^{i\alpha/2} & 0 \\ 0 & e^{-i\alpha/2} \end{pmatrix} \\ &= \begin{pmatrix} e^{i(\gamma+\alpha)/2} \cos \beta/2 & e^{i(\gamma-\alpha)/2} \text{sen } \beta/2 \\ -e^{-i(\gamma-\alpha)/2} \text{sen } \beta/2 & e^{-i(\gamma+\alpha)/2} \cos \beta/2 \end{pmatrix} . \end{aligned} \quad (7.57)$$

Esta es nuestra forma general alternativa, la ecuación (7.38), con

$$\varepsilon = \frac{(\gamma + \alpha)}{2} , \quad \eta = \frac{\beta}{2} , \quad \zeta = \frac{(\gamma - \alpha)}{2} . \quad (7.58)$$

De la ecuación (7.57) podemos identificar los parámetros de la ecuación (7.38) como

$$\begin{aligned} a &= e^{i(\gamma+\alpha)/2} \cos \beta/2 \\ b &= e^{i(\gamma-\alpha)/2} \text{sen } \beta/2 \end{aligned} \quad (7.59)$$

SU(2) isospin y el octeto SU(3).

En el tratamiento de las partículas con interacciones fuertes de Física nuclear y de altas energías que conducen al grupo de $SU(2)$ de isospin y la simetría de sabor $SU(3)$, podríamos mirar el momento angular y el grupo de rotación $SO(3)$ por una analogía. Supongamos que tenemos un electrón en un potencial atractivo esféricamente simétrico de algún núcleo atómico. La función de onda para el electrón puede ser caracterizada por tres números cuánticos n, l, m , que están relacionados con los autovalores de los operadores conservados H, L^2, L_z . La energía,⁵ sin embargo, es $2l + 1$ veces degenerada, dependiendo solamente de n y l . La razón para esta degeneración puede ser expresado de dos maneras equivalentes:

1. El potencial es simétricamente esférico, independiente de θ, φ .
2. El hamiltoniano de Schrodinger $-(\hbar^2/2m_e)\nabla^2 + V(r)$ es invariante bajo rotaciones espaciales ordinarias $SO(3)$.

Como una consecuencia de la simetría esférica del potencial $V(r)$, el momento angular orbital \vec{L} es conservado. En la sección 7.2 las componentes cartesianas de \vec{L} están indentificadas como los generadores del grupo de rotación $SO(3)$. En vez de representar L_x, L_y, L_z por operadores, usaremos matrices. Las matrices L_i son matrices $(2l + 1) \times (2l + 1)$ con la

⁵Para un potencial de Coulomb puro la energía depende sólo de n .

misma dimensión del número de estados degenerados. La dimensión $2l + 1$ está identificada con los estados degenerados $2l + 1$.

Esta degeneración es removida por un campo magnético \vec{B} , hecho conocido como el efecto Zeeman. Esta interacción magnética añade un término al Hamiltoniano que *no* es invariante bajo $SO(3)$. Este es un término que quiebra la simetría.

En el caso de partículas con interacción fuerte (protones, neutrones, etc.) no podemos seguir la analogía directamente, ya que todavía no entendemos completamente las interacciones nucleares. La fuerza fuerte está descrita por la teoría gauge de Yang-Mills basada sobre la simetría de color $SU(3)$ llamada cromodinámica cuántica o abreviada QCD. Sin embargo, QCD es una teoría no lineal y por lo tanto complicada a grandes distancias y baja energía que permanece no resuelta. Por lo tanto, no conocemos el Hamiltoniano, en vez de esto, volveremos a la analogía.

En los años 1930, después del descubrimiento del neutrón, Heisenberg propuso que las fuerzas nucleares eran cargas independientes. Los neutrones difieren en masa de los protones solamente en un 1.6%. Si esta pequeña diferencia es ignorada, el neutrón y el protón podrían ser consideradas como dos estados de cargas (o isospin) de un doblete, llamado nucleón. El isospin l tiene proyección en el eje z $I_3 = 1/2$ para el protón y $I_3 = -1/2$ para el neutrón. El isospin no tiene nada que ver con el spin (el momento angular intrínseco de una partícula) pero las dos componentes del estado de isospin obedece las mismas relaciones matemáticas que el estado de spin $1/2$. Para el nucleón, $l = \tau/2$, son las matrices usuales de Pauli y los estados de isospin ($\pm 1/2$) son autovectores de la matriz de Pauli $\tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Similarmente, los tres estados de carga del pión π^+ , π^0 , π^- forman un triplete. El pión es la partícula más liviana con interacción fuerte y es la mediadora de la fuerza nuclear a distancia, como el fotón es partícula que media la fuerza electromagnética. La interacción fuerte trata igualmente a miembros de esa familia de partículas, o multipletes y conserva el isospin. La simetría es el grupo isospin $SU(2)$.

El octuplete mostrado en la tabla 7.1 llama la atención⁶. Los números cuánticos conservados que son análogos y generalizaciones de L_z y L^2 de $SO(3)$ son I_3 e I^2 para el isospin, e Y para *hipercarga*. Las partículas pueden ser agrupadas dentro de multipletes de carga o de isospin. Entonces la hipercarga puede ser tomada como dos veces el promedio de carga del multiplete. Para el nucleón, *i.e.*, el doblete neutrón-protón, $Y = 2 \cdot \frac{1}{2}(0 + 1) = 1$. Los valores de la hipercarga y los del isospin son listados en la tabla 7.1 para bariones como el nucleón y sus compañeros (aproximadamente degenerados). Ellos forman un octeto como lo muestra la figura 7.3. En 1961 Gell-Mann, e independientemente Ne'eman, sugirieron que la interacción fuerte debe ser (aproximadamente) invariante bajo un grupo espacial tridimensional unitario, $SU(3)$, esto es, tienen *simetría de sabor* $SU(3)$.

La elección de $SU(3)$ estuvo basada primero sobre los dos números cuánticos conservados e independientes $H_1 = I_3$ y $H_2 = Y$ (*i.e.*, generados con $[I_3, Y] = 0$), que llaman para un grupo de rango 2. Segundo, el grupo ha tenido una representación de ocho dimensiones para tener en cuenta los cercanamente degenerados bariones y cuatro octetos similares para los mesones. En un sentido $SU(3)$ es la generalización más simple del isospin $SU(2)$. Tres de sus generadores son matrices hermíticas de 3×3 de traza nula que contienen las matrices de

⁶Todas las masas están dadas en unidades de energía.

		Masa [MeV]	Y	I	I_3
Ξ	Ξ^-	1321.32	-1	$\frac{1}{2}$	$-\frac{1}{2}$
	Ξ^0	1314.9			$+\frac{1}{2}$
Σ	Σ^-	1197.43	0	1	-1
	Σ^0	1192.55			0
	Σ^+	1189.37			+1
Λ	Λ	1115.63	0	0	0
N	n	939.566	1	$\frac{1}{2}$	$-\frac{1}{2}$
	p	938.272			$+\frac{1}{2}$

Cuadro 7.1: Bariones con spin $\frac{1}{2}$ y paridad par

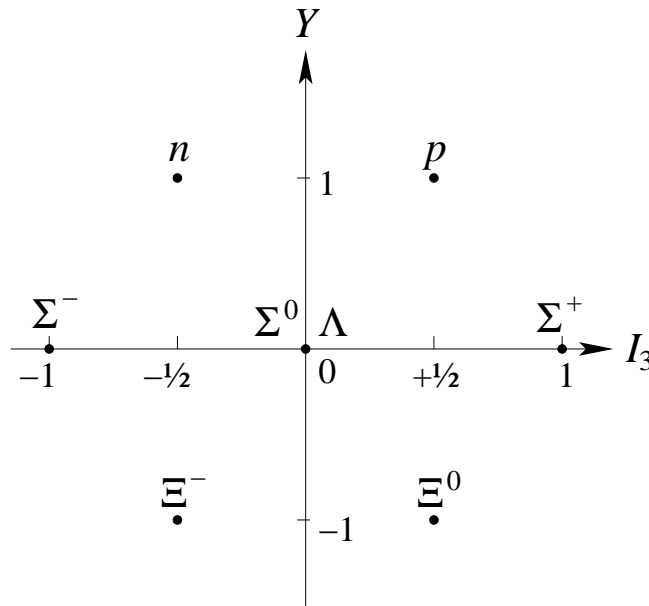


Figura 7.3: Octeto bariónico diagrama de peso para SU(3).

Pauli de 2×2 para los isospin τ_i en la esquina superior izquierda.

$$\lambda_i = \begin{pmatrix} \tau_i & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad i = 1, 2, 3. \tag{7.60}$$

De este modo, el grupo del isospin $SU(2)$ es un subgrupo de $SU(3)$ con $I_3 = \lambda_3/2$. Otros cuatro generadores tienen los no diagonales 1 de τ_1 e i , $-i$ de τ_2 en todas las otras posibles ubicaciones para formar las matrices hermíticas 3×3 de traza nula.

$$\begin{aligned} \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \\ \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, & \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}. \end{aligned} \quad (7.61)$$

El segundo generador diagonal tiene la matriz unidad bidimensional 1_2 en la esquina superior izquierda, la cual la hace claramente independiente del subgrupo $SU(2)$ isospin ya que su traza no nula en el subespacio, y -2 en el lugar de la tercera diagonal la hace traza nula,

$$\lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \quad (7.62)$$

Generalmente hay $3^2 - 1 = 8$ generadores para $SU(3)$ el cual tiene orden 8. De los conmutadores de esos generadores pueden obtenerse fácilmente las constantes de estructura de $SU(3)$.

Volviendo a la simetría de sabor $SU(3)$ imaginemos que el Hamiltoniano para nuestro octeto de bariones están compuesto de tres partes

$$H = H_{\text{fuerte}} + H_{\text{medio}} + H_{\text{electromagnético}}. \quad (7.63)$$

La primera parte, H_{fuerte} , tiene la simetría $SU(3)$ y conduce a la degeneración ocho. La introducción del término de quiebre de simetría, H_{medio} , remueve parte de la degeneración dando los cuatro multipletes del isospin (Ξ^-, Ξ^0) , $(\Sigma^-, \Sigma^0, \Sigma^+)$, Λ , y $N = (p, n)$ con diferentes masas. Estos aún son multipletes ya que H_{medio} tiene la simetría del isospin $SU(2)$. Finalmente, la presencia de fuerzas dependientes de la carga separan los multipletes de isospin y remueve la última degeneración. Esta secuencia se muestra en la figura 7.4

Aplicando teoría de perturbación de primer orden de Mecánica Cuántica, relaciones simples de masas de bariónicas pueden ser calculadas. Quizás el suceso más espectacular de este modelo $SU(3)$ ha sido su predicción de nuevas partículas. En 1961 cuatro mesones K y tres π (todos pseudoescalares; spin 0, paridad impar) sugieren otro octeto, similar al del octeto bariónico. $SU(3)$ predice un octavo mesón η , de masa 563 MeV. El mesón η con una masa determinada experimentalmente de 548 MeV fue encontrado poco después. Agrupamientos de nueve de los bariones más pesados (todos con spin $3/2$, con paridad par) sugirió un multiplete de 10 miembros o un decaplete de $SU(3)$. El décimo barión faltante fue predicho con una masa cercana a 1680 MeV y una carga negativa. En 1964 Ω^- cargada negativamente con masa (1675 ± 12) MeV fue descubierta.

La representación de octeto no es la más simple para $SU(3)$. La representación más simple son triangulares como se muestra en la figura 7.5 a partir de las cuales todas las otras pueden ser generadas por acoplamiento del momento angular generalizado. La representación

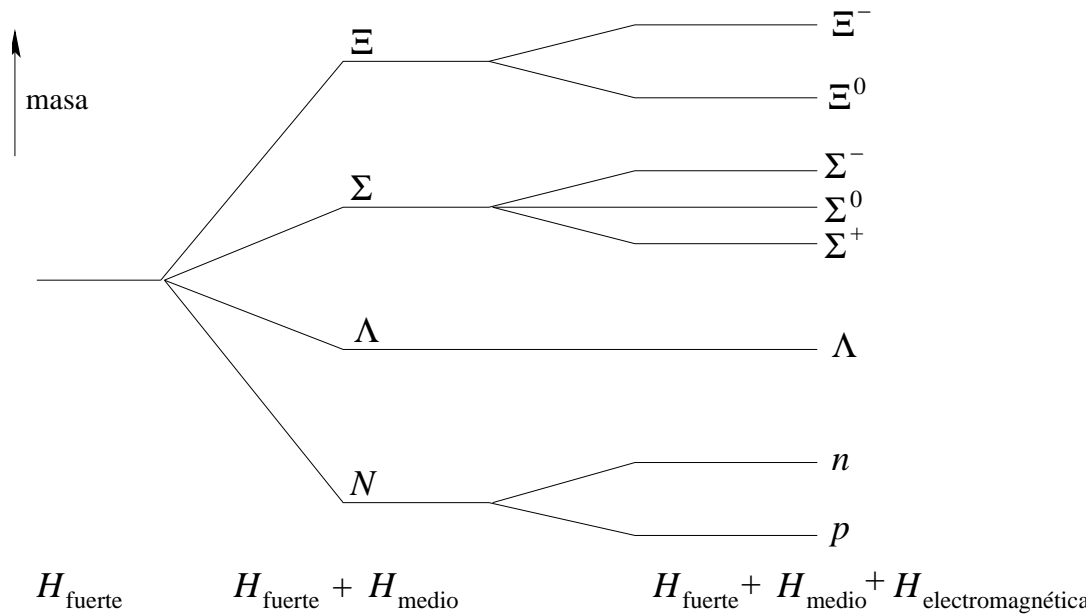


Figura 7.4: Separación de masa bariónica.

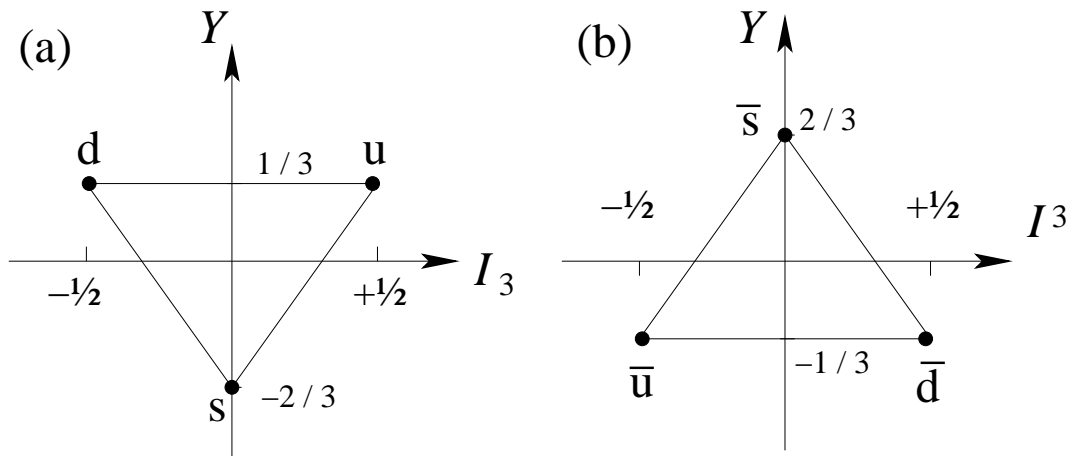


Figura 7.5: Separación de masa bariónica.

fundamental en la figura 7.5 (a) contiene los quark u (arriba) y d (abajo) y el s (extrañeza), y figura 7.5 (b) los correspondientes antiquarks. Ya que los octetos de mesones pueden ser obtenidos a partir de la representación de quark como pares $q\bar{q}$, $3^2 = 8 + 1$, esto sugiere que los mesones contienen quarks (y antiquarks) como sus constituyentes. El modelo de quarks resultante dan una exitosa descripción de la espectroscopía hadrónica. La solución de sus problemas con el principio de exclusión de Pauli eventualmente conduce a la teoría de gauge de SU(3)-color de las interacciones fuertes llamada cromodinámica cuántica o QCD.

Para mantener la teoría de grupo en su real perspectiva, podríamos enfatizar que la teoría de grupo identifica y formaliza las simetrías. Ella clasifica partículas (y algunas veces predice). Pero a parte de decir que una parte del hamiltoniano tiene simetría SU(2) y otra parte tiene

simetría $SU(3)$, la teoría de grupo no dice nada a cerca de la interacción de las partículas. Recuerde que la afirmación de que el potencial atómico es esféricamente simétrico no nos dice nada a cerca de la dependencia radial del potencial o de su función de onda. En contraste, en una teoría de gauge la interacción es mediada por bosones vectoriales (como el fotón media en la electrodinámica cuántica) y determinado únicamente por la derivada covariante de gauge.

7.3. Momento angular orbital.

El concepto clásico de momento angular $\vec{L}_{\text{clásico}} = \vec{r} \times \vec{p}$ es mostrado en el capítulo de vectores para presentar el producto cruz. Siguiendo con la representación usual de Schrödinger de la Mecánica Cuántica, el momento lineal clásico \vec{p} es reemplazado por el operador $-i\vec{\nabla}$. El operador de momento angular en la mecánica cuántica se convierte en ⁷

$$\vec{L}_{QM} = -i\vec{r} \times \vec{\nabla} . \quad (7.64)$$

Las componentes del momento angular satisfacen las relaciones de conmutación

$$[L_i, L_j] = i\varepsilon_{ijk}L_k . \quad (7.65)$$

El ε_{ijk} es el símbolo de Levi-Civita. Una suma sobre el índice k es sobreentendida.

El operador diferencial correspondiente al cuadrado del momento angular

$$\vec{L}^2 = \vec{L} \cdot \vec{L} = L_x^2 + L_y^2 + L_z^2 , \quad (7.66)$$

puede ser determinado a partir de

$$\vec{L} \cdot \vec{L} = (\vec{r} \times \vec{p}) \cdot (\vec{r} \times \vec{p}) , \quad (7.67)$$

la cual puede verificarse como ejercicio. Ya que \vec{L}^2 es un escalar rotacional, $[\vec{L}^2, L_i] = 0$, el cual también puede ser verificado directamente.

La ecuación (7.65) presenta las relaciones de conmutación básicas de los componentes del momento angular en Mecánica Cuántica. Por cierto, dentro del marco de la Mecánica Cuántica y la teoría de grupo, estas relaciones de conmutación definen un operador de momento angular.

Acercamiento a los operadores de subida y bajada.

Comencemos con una aproximación más general, donde el momento angular \vec{J} lo consideramos que puede representar un momento angular orbital \vec{L} , un spin $\sigma/2$, o un momento angular total $\vec{L} + \sigma/2$, etc. Supongamos que

1. J es un operador hermítico cuyas componentes satisfacen las relaciones de conmutación

$$[J_i, J_j] = i\varepsilon_{ijk}J_k , \quad [\vec{J}^2, J_i] = 0 . \quad (7.68)$$

⁷Por simplicidad, $\hbar = 1$. Esto significa que el momento angular es medido en unidades de \hbar .

2. $|\lambda M\rangle$ es simultaneamente una autofunción normalizada (o autovector) de J_z con autovalor M y una autofunción de \vec{J}^2 ,

$$J_z|\lambda M\rangle = M|\lambda M\rangle, \quad \vec{J}^2|\lambda M\rangle = \lambda|\lambda M\rangle. \quad (7.69)$$

Mostraremos ahora que $\lambda = J(J + 1)$. Este tratamiento ilustrará la generalidad y potencia de las técnicas de operadores particularmente el uso de operadores de subida y bajada.

Un operador de subida o bajada se defina como

$$J_+ = J_x + iJ_y, \quad J_- = J_x - iJ_y. \quad (7.70)$$

En términos de ese operador \vec{J}^2 puede ser reescrito como

$$\vec{J}^2 = \frac{1}{2}(J_+J_- + J_-J_+) + J_z^2. \quad (7.71)$$

A partir de las relaciones de conmutación, encontramos que

$$[J_z, J_+] = +J_+, \quad [J_z, J_-] = -J_-, \quad [J_+, J_-] = 2J_z. \quad (7.72)$$

Ya que J_+ conmuta con \vec{J}^2 , hagalo como ejercicio

$$\vec{J}^2(J_+|\lambda M\rangle) = J_+(\vec{J}^2|\lambda M\rangle) = \lambda(J_+|\lambda M\rangle). \quad (7.73)$$

Por lo tanto, $J_+|\lambda M\rangle$ todavía es una autofunción de \vec{J}^2 con autovalores λ , y similarmente para $J_-|\lambda M\rangle$. Pero de la ecuación (7.72)

$$J_zJ_+ = J_+(J_z + 1), \quad (7.74)$$

o

$$J_z(J_+|\lambda M\rangle) = J_+(J_z + 1)|\lambda M\rangle = (M + 1)(J_+|\lambda M\rangle). \quad (7.75)$$

Por lo tanto, $J_+|\lambda M\rangle$ todavía es una autofunción de J_z con autovalores $M + 1$. J_+ ha elevado el autovalor en 1 y por eso es llamado *operador de subida*. Similarmente, J_- baja los autovalores en 1 y a menudo es llamado *operador de bajada*.

Tomando los valores esperados y usando $J_x^\dagger = J_x$, $J_y^\dagger = J_y$,

$$\langle \lambda M | \vec{J}^2 - J_z^2 | \lambda M \rangle = \langle \lambda M | J_x^2 + J_y^2 | \lambda M \rangle = | \langle J_x | \lambda M \rangle |^2 + | \langle J_y | \lambda M \rangle |^2,$$

vemos que $\lambda - M^2 \geq 0$, tal que M es ligado. Sea J el más grande valor de M . Luego $J_+|\lambda J\rangle = 0$, lo cual implica que $J_-J_+|\lambda J\rangle = 0$. Luego combinando las ecuaciones (7.71) y (7.72) obtenemos

$$\vec{J}^2 = J_-J_+ + J_z(J_z + 1), \quad (7.76)$$

encontramos que a partir de la ecuación (7.76)

$$0 = J_-J_+|\lambda M = J\rangle = (\vec{J}^2 - J_z^2 - J_z)|\lambda M = J\rangle = (\lambda - J^2 - J)|\lambda M = J\rangle.$$

Por lo tanto

$$\lambda = J(J + 1) \geq 0; \quad (7.77)$$

con un J no negativo. Ahora reetiquetaremos los estados $|\lambda M\rangle = |JM\rangle$. Similarmente, sea J' el más pequeño de los M . Entonces $J_-|JJ'\rangle = 0$. A partir de

$$\vec{J}^2 = J_+J_- - J_z(J_z + 1) , \quad (7.78)$$

vemos que

$$0 = J_+J_-|JJ'\rangle = (\vec{J}^2 + J_z - J_z^2)|JJ'\rangle = (\lambda + J' - J'^2)|JJ'\rangle . \quad (7.79)$$

De manera que

$$\lambda = J(J + 1) = J'(J' - 1) = (-J)(-J - 1) .$$

Así $J' = -J$, y M corre en pasos enteros desde $-J$ a $+J$,

$$-J \leq M \leq +J . \quad (7.80)$$

Comenzando desde $|JJ\rangle$ y aplicando J_- repetidas veces, alcanzaremos todos los otros estados $|JM\rangle$. De manera que $|JM\rangle$ forma una representación irreducible; M varía y J está fijo.

Entonces usando las ecuaciones (7.68), (7.76) y (7.78) obtenemos

$$\begin{aligned} J_-J_+|JM\rangle &= [J(J + 1) - M(M + 1)]|JM\rangle = (J - M)(J + M + 1)|JM\rangle , \\ J_+J_-|JM\rangle &= [J(J + 1) - M(M - 1)]|JM\rangle = (J + M)(J - M + 1)|JM\rangle . \end{aligned} \quad (7.81)$$

Como J_+ y J_- son hermíticos conjugados,

$$J_+^\dagger = J_- , \quad J_-^\dagger = J_+ , \quad (7.82)$$

los autovalores o valores esperados en la ecuación (7.81) deberían ser positivos o cero.

Ya que J_+ aumenta el autovalor de M a $M + 1$, reetiquetaremos la autofunción resultante $|JM + 1\rangle$. La normalización está dada por la ecuación (7.81) como

$$J_+|JM\rangle = \sqrt{(J - M)(J + M + 1)}|JM + 1\rangle , \quad (7.83)$$

tomando la raíz cuadrada positiva y no introduciendo ningún factor de fase. Por los mismos argumentos

$$J_-|JM\rangle = \sqrt{(J + M)(J - M + 1)}|JM - 1\rangle . \quad (7.84)$$

Finalmente, ya que M va desde $-J$ a $+J$ en pasos unitarios, $2J$ debería ser un número entero. J es por lo tanto un entero o la mitad de un entero impar. Como hemos visto, el momento angular orbital está descrito con J entero. A partir de los spin de algunas partículas fundamentales y de algunos núcleos, tenemos que $J = 1/2, 3/2, 5/2, \dots$. Nuestro momento angular está cuantizado esencialmente como un resultado de relaciones de conmutaciones. En coordenadas polares esféricas θ, φ las funciones $\langle \theta, \varphi | lm \rangle = Y_l^m(\theta, \varphi)$ son armónicos esféricos.

Resumen de grupos y álgebras de Lie.

Las relaciones de conmutaciones generales, ecuación (7.14) en la sección 7.2, para un grupo de Lie clásico [SO(n) y SU(n) en particular] pueden ser simplificadas para verse más como la ecuación (7.72) para SO(3) y SU(2) en la sección 7.3.

Álgebra de Lie	A_l	B_l	D_l
Grupo de Lie	SU(1+1)	SO(2l+1)	SO(2l)
rango	1	1	1
orden	l(1+2)	l(2l+1)	l(2l-1)

Cuadro 7.2: Rango y orden de grupos rotacionales y unitarios.

Primero escogemos generadores H_i que sean linealmente independientes y que conmuten entre sí, estos son generalizaciones de J_z de SO(3) y SU(2). Sea l el número máximo de tales H_i con

$$[H_i, H_k] = 0 . \quad (7.85)$$

Entonces l es llamado el *rango* del grupo Lie G o de su álgebra. El rango y la dimensión u orden de algunos grupos Lie son dados en la tabla 7.2. Todos los otros generadores E_α puede mostrarse que son operadores de subida y bajada con respecto a todos los H_i , tal que

$$[H_i, E_\alpha] = \alpha_i E_\alpha . \quad (7.86)$$

El conjunto de los $(\alpha_1, \alpha_2, \dots, \alpha_l)$ son llamados los *vectores raíces*.

Ya que los H_i conmutan entre sí, ellos pueden ser diagonalizados simultáneamente. Ellos nos dan un conjunto de autovalores m_1, m_2, \dots, m_l . Al conjunto (m_1, m_2, \dots, m_l) se les llama vectores de peso de una representación irreductible. Hay l operadores invariantes C_i , llamados *operadores de Casimir*, los cuales conmutan con todos los generadores y son generalizaciones de J^2 ,

$$[C_i, H_j] = 0 , \quad [C_i, E_\alpha] = 0 . \quad (7.87)$$

El primero, C_1 , es una función cuadrática de los generadores, los otros son más complicados. Ya que C_i conmuta con todos los H_j , ellos pueden ser diagonalizados simultáneamente con los H_j . Sus autovalores c_1, c_2, \dots, c_l caracterizan las representaciones irreductibles y permanecen constantes mientras los vectores de peso varían sobre una representación irreductible particular. Por lo tanto, la autofunción general puede ser escrita como

$$|(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle , \quad (7.88)$$

generalizando $|JM\rangle$ de SO(3) y SU(2). Sus ecuaciones de autovalores son

$$H_i |(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle = m_i |(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle , \quad (7.89a)$$

$$C_i |(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle = c_i |(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle . \quad (7.89b)$$

Ahora podemos mostrar que $E_\alpha |(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle$ tiene los vector peso $(m_1 + \alpha_1, m_2 + \alpha_2, \dots, m_l + \alpha_l)$ usando las relaciones de conmutación, la ecuación (7.86), en conjunto con las ecuaciones (7.89a) y (7.89b),

$$\begin{aligned} H_i E_\alpha |(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle &= (E_\alpha H_i + [H_i, E_\alpha]) |(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle \\ &= (m_i + \alpha_i) E_\alpha |(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle . \end{aligned} \quad (7.90)$$

Por lo tanto

$$E_\alpha|(c_1, c_2, \dots, c_l)m_1, m_2, \dots, m_l\rangle \sim |(c_1, c_2, \dots, c_l)m_1 + \alpha_1, m_2 + \alpha_2, \dots, m_l + \alpha_l\rangle$$

estas son las generalizaciones de las ecuaciones (7.83) y (7.84) a partir de $SO(3)$. Esos cambios de autovalores por el operador E_α son llamados sus reglas de selección en mecánica cuántica.

7.4. Grupo homogéneo de Lorentz.

En relatividad especial requerimos que nuestras leyes físicas sean covariantes⁸ bajo

- traslaciones en el tiempo y en el espacio,
- rotaciones en el espacio real tridimensional, y
- transformaciones de Lorentz.

El requerimiento para la covarianza bajo traslaciones está basada en la homogeneidad del espacio y el tiempo. Covarianza bajo rotaciones es una afirmación de la isotropía del espacio. El requerimiento de la covarianza de Lorentz viene de la relatividad especial. Todas estas tres transformaciones en conjunto forman el grupo inhomogeneo de Lorentz o el grupo de Poincaré. Aquí excluimos las traslaciones. Las rotaciones espaciales y las transformaciones de Lorentz forman un grupo, el grupo homogéneo de Lorentz.

Primero generamos un subgrupo, las transformaciones de Lorentz en el cual la velocidad relativa \vec{v} está a lo largo del eje $x = x_1$. El generador puede ser determinado considerando un marco de referencia espacio-temporal moviéndose con una velocidad relativa infinitesimal δv . Las relaciones son similares a aquellas para rotaciones en el espacio real, excepto que aquí el ángulo de rotación es imaginario puro.

Las transformaciones de Lorentz son lineales no sólo en el espacio de coordenadas x_i si no que también en el tiempo t . Ellas se originan a partir de las ecuaciones de Maxwell de la electrodinámica las cuales son invariantes bajo la transformaciones de Lorentz, como veremos luego. Las transformaciones de Lorentz dejan invariante la forma cuadrática siguiente $c^2t^2 - x_1^2 - x_2^2 - x_3^2 = x_0^2 - x_1^2 - x_2^2 - x_3^2$ donde $x_0 = ct$. Vemos esto si encendemos una fuente de luz en el origen del sistema de coordenadas. En tiempo t la luz ha viajado una distancia $ct = \sqrt{\sum x_i^2}$, tal que $c^2t^2 - x_1^2 - x_2^2 - x_3^2 = 0$. La relatividad especial requiere esto en todos los sistemas (inercial) que se mueven con velocidad $v \leq c$ en cualquier dirección relativa al sistema x_i y que tengan el mismo origen a tiempo $t = 0$, se mantenga también que $c^2t'^2 - x_1'^2 - x_2'^2 - x_3'^2 = 0$. El espacio cuatridimensional con la métrica $x_0^2 - x_1^2 - x_2^2 - x_3^2$ es llamado espacio de Minkowski con el producto escalar de dos cuatrivectores definido como $a \cdot b = a_0b_0 - \vec{a} \cdot \vec{b}$. Usando el tensor métrico

$$(g_{\mu\nu}) = (g^{\mu\nu}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (7.91)$$

⁸Ser covariante significa que tienen la misma forma en diferentes sistemas de coordenadas tal que no hay un sistema de referencia privilegiado.

podemos subir y bajar índices de un cuadrivector tal como de las coordenadas $x^\mu = (x_0, \vec{x})$ es decir $x_\mu = g_{\mu\nu}x^\nu = (x_0, -\vec{x})$ y $x^\mu g_{\mu\nu}x^\nu = x_0^2 - \vec{x}^2$, la convección de suma de Einstein se dan por entendida. Para el gradiente $\partial^\mu = (\partial/\partial x_0, -\vec{\nabla}) = \partial/\partial x_\mu$ y $\partial_\mu = (\partial/\partial x_0, \vec{\nabla})$ tal que $\partial^2 = \partial^2/\partial x_0^2 - \nabla^2$ es un escalar de Lorentz, al igual que la métrica $x_0^2 - \vec{x}^2$.

Para $v \ll c$, en el límite no relativista, las transformaciones de Lorentz deben ser transformaciones de Galileo. Por lo tanto, para derivar la forma de una transformación de Lorentz a lo largo del eje x_1 , partimos con una transformación Galileana para una velocidad relativa infinitesimal δv :

$$x'_1 = x_1 - \delta vt = x_1 - x_0 \delta\beta . \quad (7.92)$$

Como es usual $\beta = \frac{v}{c}$. Por simetría también podemos escribir

$$x'_0 = x_0 - ax_1 \delta\beta , \quad (7.93)$$

donde a es un parámetro a fijar una vez que se imponga que $x_0^2 - x_1^2$ deba ser invariante,

$$x_0'^2 - x_1'^2 = x_0^2 - x_1^2 . \quad (7.94)$$

Recordemos que $x = (x_0; x_1, x_2, x_3)$ es el prototipo de cuadrivector en el espacio de Minkowski. Así la ecuación (7.94) es simplemente una afirmación de la invariancia del cuadrado de la magnitud del vector distancia bajo rotaciones en el espacio de Minkowski. Aquí es donde la relatividad especial compromete nuestra transformación. Elevando al cuadrado y restando las ecuaciones (7.92) y (7.93) y descartando términos del orden de $\delta\beta^2$, encontramos $a = -1$. Las ecuaciones (7.92) y (7.93) pueden ser combinadas como una ecuación matricial

$$\begin{pmatrix} x'_0 \\ x'_1 \end{pmatrix} = (1 - \delta\beta\sigma_1) \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} , \quad (7.95)$$

σ_1 es la matriz de Pauli, y el parámetro $\delta\beta$ representa un cambio infinitesimal. Repetimos la transformación N veces para desarrollar una transformación finita con el parámetro velocidad $\rho = N\delta\beta$. entonces

$$\begin{pmatrix} x'_0 \\ x'_1 \end{pmatrix} = \left(1 - \frac{\rho\sigma_1}{N}\right)^N \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} . \quad (7.96)$$

En el límite $N \rightarrow \infty$

$$\lim_{N \rightarrow \infty} \left(1 - \frac{\rho\sigma_1}{N}\right)^N = \exp(-\rho\sigma_1) . \quad (7.97)$$

Interpretamos la exponencial como una serie de Maclaurin

$$\exp(-\rho\sigma_1) = 1 - \rho\sigma_1 + \frac{(\rho\sigma_1)^2}{2!} - \frac{(\rho\sigma_1)^3}{3!} + \dots . \quad (7.98)$$

Notando que $\sigma^2 = 1$,

$$\exp(-\rho\sigma_1) = 1 \cosh \rho + \sigma_1 \sinh \rho . \quad (7.99)$$

Por lo tanto nuestra transformación de Lorentz finita es

$$\begin{pmatrix} x'_0 \\ x'_1 \end{pmatrix} = \begin{pmatrix} \cosh \rho & -\sinh \rho \\ -\sinh \rho & \cosh \rho \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} . \quad (7.100)$$

σ_1 ha generado las representaciones de esta especial transformación de Lorentz.

El $\cosh \rho$ y el $\sinh \rho$ pueden ser identificados considerando el origen del sistema de coordenadas primas, $x'_1 = 0$, o $x_1 = vt$. Sustituyendo en la ecuación (7.100), tenemos

$$0 = x_1 \cosh \rho - x_0 \sinh \rho . \quad (7.101)$$

Con $x_1 = vt$ y $x_0 = ct$.

$$\tanh \rho = \beta = \frac{v}{c} .$$

Note que la rapidez $\rho \neq \frac{v}{c}$ excepto en el límite $v \rightarrow 0$.

Usando $1 - \tanh^2 \rho = (\cosh^2 \rho)^{-1}$,

$$\cosh \rho = (1 - \beta^2)^{-1/2} \equiv \gamma , \quad \sinh \rho = \beta \gamma . \quad (7.102)$$

El anterior caso especial en que la velocidad es paralela a uno de los ejes espaciales es simple, pero ilustra la velocidad infinitesimal, la técnica de la exponenciación y el generador. Ahora esta técnica puede ser aplicada para derivar las transformaciones de Lorentz para una velocidad relativa \vec{v} no paralela a ningún eje. Las matrices dadas por la ecuación (7.100) para el caso $\vec{v} = \hat{x}v_x$ forman un subgrupo. Las matrices en el caso general no lo hacen. El producto de dos matrices de transformaciones de Lorentz, $L(\vec{v}_1)$ y $L(\vec{v}_2)$, producen una tercera matriz de transformación $L(\vec{v}_3)$, si las dos velocidades \vec{v}_1 y \vec{v}_2 son paralelas. La velocidad resultante v_3 está relacionada con v_1 y con v_2 mediante la regla de adición de velocidades de Einstein. Si \vec{v}_1 y \vec{v}_2 no son paralelas, no existe entonces una relación simple.

7.5. Covarianza de Lorentz de las ecuaciones de Maxwell.

Si una ley física se mantiene para todas las orientaciones de nuestro (real) espacial sistema de coordenadas (*i.e.* es invariante ante rotaciones), los términos de la ecuación deben ser covariantes bajo rotaciones. Esto significa que escribimos las leyes físicas en la forma matemática escalar=escalar, vector=vector, tensor de segundo rango=tensor de segundo rango, y así sucesivamente. Similarmente, si una ley física se mantiene para todos los sistemas inerciales, los términos de la ecuación deben ser covariantes bajo transformaciones de Lorentz.

Usando el espacio de Minkowski ($x = x_1, y = x_2, z = x_3, ct = x_0$) tenemos un espacio cuatridimensional cartesiano con métrica $g_{\mu\nu}$. Las transformaciones de Lorentz son lineales en el espacio y en el tiempo en este espacio real de cuatro dimensiones.

Consideremos las ecuaciones de Maxwell

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} , \quad (7.103a)$$

$$\vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \rho \vec{v} , \quad (7.103b)$$

$$\vec{\nabla} \cdot \vec{D} = \rho , \quad (7.103c)$$

$$\vec{\nabla} \cdot \vec{B} = 0 , \quad (7.103d)$$

y las relaciones

$$\vec{D} = \varepsilon_0 \vec{E}, \quad \vec{B} = \mu_0 \vec{H}. \quad (7.104)$$

Todos los símbolos tienen sus significados usuales y hemos supuesto el vacío por simplicidad.

Supongamos que las ecuaciones de Maxwell se mantienen en todos los sistemas inerciales; esto es, las ecuaciones de Maxwell son consistentes con la relatividad especial. (La covarianza de las ecuaciones de Maxwell bajo transformaciones de Lorentz fue realmente mostrada por Lorentz y Poincaré antes de que Einstein propusiera su teoría de la relatividad especial). Nuestro objetivo inmediato es reescribir las ecuaciones de Maxwell como ecuaciones tensoriales en el espacio de Minkowski. Esto hará la covarianza de Lorentz explícita.

En términos de los potenciales escalar y vectorial, podemos escribir

$$\begin{aligned} \vec{B} &= \vec{\nabla} \times \vec{A}, \\ \vec{E} &= -\frac{\partial \vec{A}}{\partial t} - \vec{\nabla} \varphi. \end{aligned} \quad (7.105)$$

La ecuación anterior especifica el rotor de \vec{A} ; la divergencia de \vec{A} no está definida. Podemos, y por futuras conveniencias lo hacemos, imponer la siguiente relación sobre el vector potencial

$$\vec{\nabla} \cdot \vec{A} + \varepsilon_0 \mu_0 \frac{\partial \varphi}{\partial t} = 0. \quad (7.106)$$

Este es conocido como el gauge de Lorentz. Servirá a nuestros propósitos de desacoplar las ecuaciones diferenciales para \vec{A} y para φ .

Ahora reescribimos las ecuaciones de Maxwell en términos de los potenciales. A partir de la ecuación (7.103c) para $\vec{\nabla} \cdot \vec{D}$ y (7.105)

$$\nabla^2 \varphi + \vec{\nabla} \cdot \frac{\partial \vec{A}}{\partial t} = -\frac{\rho}{\varepsilon_0}, \quad (7.107)$$

considerando que la ecuación (7.103b) para $\vec{\nabla} \times \vec{H}$ y (7.105) y la identidad vectorial para el rotor del rotor produce

$$\frac{\partial^2 \vec{A}}{\partial t^2} + \vec{\nabla} \frac{\partial \varphi}{\partial t} + \frac{1}{\varepsilon_0 \mu_0} \left[\vec{\nabla} \vec{\nabla} \cdot \vec{A} - \nabla^2 \vec{A} \right] = \frac{\rho \vec{v}}{\varepsilon_0}. \quad (7.108)$$

Usando el gauge de Lorentz, la ecuación (7.106), y la relación $\varepsilon_0 \mu_0 = 1/c^2$, obtenemos

$$\begin{aligned} \left[\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right] \vec{A} &= -\mu_0 \rho \vec{v}, \\ \left[\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right] \varphi &= -\frac{\rho}{\varepsilon_0}. \end{aligned} \quad (7.109)$$

Ahora el operador diferencial

$$\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} = \partial^2 = -\partial^\mu \partial_\mu,$$

es un Laplaciano cuadridimensional. Usualmente este operador es llamado el d'Alembertiano y denotado por \square^2 . Puede probarse que es un escalar.

Por conveniencia definimos

$$\begin{aligned} A^1 &\equiv \frac{A_x}{\mu_0 c} = c\varepsilon_0 A_x, & A^3 &\equiv \frac{A_z}{\mu_0 c} = c\varepsilon_0 A_z, \\ A^2 &\equiv \frac{A_y}{\mu_0 c} = c\varepsilon_0 A_y, & A_0 &\equiv \varepsilon_0 \varphi = A^0. \end{aligned} \quad (7.110)$$

Si ponemos además

$$\frac{\rho v_x}{c} \equiv i^1, \quad \frac{\rho v_y}{c} \equiv i^2, \quad \frac{\rho v_z}{c} \equiv i^3, \quad \rho \equiv i_0 = i^0, \quad (7.111)$$

entonces la ecuación (7.109) puede ser escrita de la forma

$$\partial^2 A^\mu = i^\mu. \quad (7.112)$$

La ecuación anterior parece una ecuación tensorial, pero eso no basta. Para probar que es una ecuación tensorial, partimos investigando las propiedades de transformación de la corriente generalizada i^μ .

Ya que un elemento de carga de es una cantidad invariante, tenemos

$$de = \rho dx_1 dx_2 dx_3, \quad \text{invariante.} \quad (7.113)$$

Vimos que el elemento de volumen cuadridimensional es también un invariante, $dx_1 dx_2 dx_3 dx_0$, comparando estos resultados vemos que la densidad de carga ρ debe transformar de la misma manera que x_0 . Ponemos $\rho = i^0$ con i^0 establecida como la componente cero de un cuadvector. Las otras partes de la ecuación (7.111) pueden ser expandidas como

$$\begin{aligned} i^1 &= \frac{\rho v_x}{c} = \frac{\rho}{c} \frac{dx_1}{dt} \\ &= i^0 \frac{dx_1}{dt}. \end{aligned} \quad (7.114)$$

Ya que justo mostramos que i_0 transforma como dx_0 , esto significa que i_1 transforma como dx_1 . Con resultados similares para i_2 e i_3 . tenemos que i^λ transforma como dx^λ , probando de esta manera que i^λ es un vector, un vector del espacio cuadridimensional de Minkowski.

La ecuación (7.112), la cual deriva directamente de las ecuaciones de Maxwell, suponemos que se mantiene en todos los sistemas cartesianos. Entonces, por la regla del cociente A_μ es también un vector y (7.112) es una legítima ecuación tensorial.

Ahora, devolviendonos, la ecuación (7.105) puede ser escrita

$$\begin{aligned} \varepsilon_0 E_j &= -\frac{\partial A^j}{\partial x_0} + \frac{\partial A^0}{\partial x_j}, \quad j = 1, 2, 3, \\ \frac{1}{\mu c} B_i &= \frac{\partial A^k}{\partial x_j} - \frac{\partial A^j}{\partial x_k}, \quad (i, j, k) = (1, 2, 3), \end{aligned} \quad (7.115)$$

y permutaciones cíclicas.

Definimos un nuevo tensor

$$\partial^\mu A^\nu - \partial^\nu A^\mu = \frac{\partial A^\nu}{\partial x_\mu} - \frac{\partial A^\mu}{\partial x_\nu} \equiv F^{\mu\nu} = -F^{\nu\mu} \quad (\mu, \nu = 0, 1, 2, 3)$$

un tensor antisimétrico de segundo rango, ya que A^μ es un vector. Escribamoslo explícitamente

$$F_{\mu\nu} = \varepsilon_0 \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -cB_z & cB_y \\ -E_y & cB_z & 0 & -cB_x \\ -E_z & -cB_y & cB_x & 0 \end{pmatrix}, \quad F^{\mu\nu} = \varepsilon_0 \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -cB_z & cB_y \\ E_y & cB_z & 0 & -cB_x \\ E_z & -cB_y & cB_x & 0 \end{pmatrix}. \quad (7.116)$$

Notemos que en nuestro espacio de Minkowski \vec{E} y \vec{B} no son más vectores sino que juntos forman un tensor de segundo rango. Con este tensor podemos escribir las dos ecuaciones de Maxwell nohomogéneas (7.103b) y (7.103c) y combinándolas como una ecuación tensorial

$$\frac{\partial F_{\mu\nu}}{\partial x_\nu} = i_\mu. \quad (7.117)$$

El lado izquierdo es una divergencia cuatridimensional de un tensor y por lo tanto un vector. Esto es, por supuesto, equivalente a contraer un tensor de tercer rango $\frac{\partial F^{\mu\nu}}{\partial x_\lambda}$. Las ecuaciones de Maxwell (7.103a) para $\vec{\nabla} \times \vec{E}$ y la ecuación (7.103d) para $\vec{\nabla} \cdot \vec{B}$ pueden ser expresadas en forma tensorial

$$\frac{\partial F_{23}}{\partial x_1} + \frac{\partial F_{31}}{\partial x_2} + \frac{\partial F_{12}}{\partial x_3} = 0, \quad (7.118)$$

para (7.103d) y tres ecuaciones de la forma

$$-\frac{\partial F_{30}}{\partial x_2} - \frac{\partial F_{02}}{\partial x_3} - \frac{\partial F_{23}}{\partial x_0} = 0, \quad (7.119)$$

para (7.103a). Una segunda ecuación permutando 120 y una tercera permutando 130.

Ya que

$$\partial^\lambda F^{\mu\nu} = \frac{\partial F^{\mu\nu}}{\partial x_\lambda} \equiv t^{\lambda\mu\nu},$$

es un tensor de tercer rango, las ecuaciones (7.117) y (7.119) pueden ser expresadas por la ecuación tensorial

$$t^{\lambda\mu\nu} + t^{\nu\lambda\mu} + t^{\mu\nu\lambda} = 0. \quad (7.120)$$

En todos los casos anteriores los índices μ , ν y λ se suponen diferentes.

Transformaciones de Lorentz de \vec{E} y \vec{B} .

La construcción de las ecuaciones tensoriales (7.118) y (7.120) completan nuestro objetivo inicial de reescribir las ecuaciones de Maxwell en forma tensorial. Ahora explotamos las propiedades tensoriales de nuestros cuatrivectores y del tensor $F^{\mu\nu}$.

Para las transformaciones de Lorentz que corresponden a movimientos a lo largo del eje $z(x_3)$ con velocidad v , los “cosenos directores” están dados por

$$\begin{aligned}x'_0 &= \gamma(x_0 - \beta x_3) \\x'_3 &= \gamma(x_3 - \beta x_0) ,\end{aligned}\tag{7.121}$$

donde

$$\beta = \frac{v}{c}$$

y

$$\gamma = (1 - \beta^2)^{-1/2} .\tag{7.122}$$

Usando las propiedades de transformación tensorial, podemos calcular los campos eléctrico y magnético en el sistema en movimiento en términos de los valores en el marco de referencias original. A partir de las ecuaciones (7.116) y (7.121) obtenemos

$$\begin{aligned}E'_x &= \frac{1}{\sqrt{1 - \beta^2}} \left(E_x - \frac{v}{c^2} B_y \right) , \\E'_y &= \frac{1}{\sqrt{1 - \beta^2}} \left(E_y + \frac{v}{c^2} B_x \right) , \\E'_z &= E_z ,\end{aligned}\tag{7.123}$$

y

$$\begin{aligned}B'_x &= \frac{1}{\sqrt{1 - \beta^2}} \left(B_x + \frac{v}{c^2} E_y \right) , \\B'_y &= \frac{1}{\sqrt{1 - \beta^2}} \left(B_y - \frac{v}{c^2} E_x \right) , \\B'_z &= B_z .\end{aligned}\tag{7.124}$$

Este acoplamiento de \vec{E} y \vec{B} es esperado. Consideremos, por ejemplo, el caso de campo eléctrico nulo en el sistema sin prima

$$E_x = E_y = E_z = 0 .$$

Claramente, no habrá fuerza sobre una partícula carga estacionaria. Cuando la partícula está en movimiento con una velocidad pequeña \vec{v} a lo largo del eje z un observador sobre la partícula ve campos (ejerciendo una fuerza sobre la partícula cargada) dados por

$$\begin{aligned}E'_x &= -v B_y , \\E'_y &= v B_x ,\end{aligned}$$

donde \vec{B} es un campo magnético en el sistema sin primas. Estas ecuaciones pueden ser puestas en forma vectorial

$$\vec{E}' = \vec{v} \times \vec{B} , \quad \text{o bien,} \quad \vec{F} = q\vec{v} \times \vec{B} ,\tag{7.125}$$

la cual es usualmente tomada como la definición operacional del campo magnético \vec{B} .

Invariantes electromagnéticas.

Finalmente, las propiedades tensoriales (o vectoriales) nos permiten construir una multitud de cantidades invariantes. Una de las importantes es el producto escalar de los cuadvectores A_λ y i_λ . Tenemos

$$\begin{aligned} A^\lambda i_\lambda &= -c\varepsilon_0 A_x \frac{\rho v_x}{c} - c\varepsilon_0 A_y \frac{\rho v_y}{c} - c\varepsilon_0 A_z \frac{\rho v_z}{c} + \varepsilon_0 \varphi \rho \\ &= \varepsilon_0 (\rho \varphi - \vec{A} \cdot \vec{J}), \quad \text{invariante,} \end{aligned} \tag{7.126}$$

con \vec{A} el usual potencial vector y \vec{J} la densidad de corriente ordinaria. El primer término $\rho\varphi$ es el ordinario acoplamiento electroestático con dimensiones de energía per unidad de volumen. En consecuencia nuestro recién construido invariante escalar es un densidad de energía. La interacción dinámica del campo y corriente es dado por el producto $\vec{A} \cdot \vec{J}$. Este invariante $A^\lambda i_\lambda$ aparece en los Lagrangianos electromagnéticos.

Capítulo 8

Series infinitas.

versión final corregida 2.31, 6 de Mayo del 2003¹

8.1. Conceptos fundamentales

Las series infinitas, literalmente sumas de un número infinito de términos, ocurre frecuentemente tanto en matemáticas pura como aplicada. Ellas podrían ser usadas por los matemáticos puros para definir funciones como una aproximación fundamental a la teoría de funciones, tanto como para calcular valores precisos de constantes y funciones trascendentales. En matemática, en ciencias y en ingeniería las series infinitas son ubicuas, es por ello que aparecen en la evaluación de integrales, en la solución de ecuaciones diferenciales, en series de Fourier y compite con las representaciones integral para la descripción de funciones especiales.

Encaramos el problema que significa la suma de un número infinito de términos. La aproximación usual es por sumas parciales. Si tenemos una sucesión de términos infinitos $u_1, u_2, u_3, u_4, u_5, \dots$, definimos la suma parcial i -ésima como

$$s_i = \sum_{n=1}^i u_n , \quad (8.1)$$

Esta es una suma finita y no ofrece dificultades. Si las sumas parciales s_i convergen a un límite (finito) cuando $i \rightarrow \infty$,

$$\lim_{i \rightarrow \infty} s_i = S , \quad (8.2)$$

La serie infinita $\sum_{n=1}^{\infty} u_n$ se dice que es convergente y tiene el valor S . Note cuidadosamente que nosotros razonablemente y plausiblemente, pero aún arbitrariamente *definimos* que la serie infinita es igual a S . Podemos notar que una condición necesaria para esta convergencia a un límite es que el $\lim_{n \rightarrow \infty} u_n = 0$. Esta condición, sin embargo, no es suficiente para garantizar la convergencia. La ecuación (8.2) usualmente está escrita en notación matemática formal:

La condición para la existencia de un límite S es que para cada $\varepsilon > 0$, haya un N fijo tal que

$$|S - s_i| < \varepsilon , \quad \text{para todo } i > N.$$

¹Este capítulo está basado en el quinto capítulo del libro: *Mathematical Methods for Physicists, fourth edition* de George B. Arfken & Hans J. Weber, editorial ACADEMIC PRESS.

Esta condición a menudo derivada del *criterio de Cauchy* aplicado a las sumas parciales s_i . El criterio de Cauchy es:

Una condición necesaria y suficiente para que una sucesión (s_i) converja es que para cada $\varepsilon > 0$ exista un número fijo N tal que

$$|s_j - s_i| < \varepsilon \quad \text{para todos los } i, j > N.$$

Esto significa que la sumas parciales individuales deben mantenerse cercanas cuando nos movemos lejos en la secuencia.

El criterio de Cauchy puede fácilmente extenderse a sucesiones de funciones. La vemos en esta forma en la sección 8.5 en la definición de convergencia uniforme y más adelante en el desarrollo del espacio de Hilbert.

Nuestras sumas parciales s_i pueden no converger a un límite simple sino que podría oscilar, como en el caso

$$\sum_{n=0}^{\infty} u_n = 1 - 1 + 1 - 1 + 1 - \dots + (-1)^n - \dots .$$

Claramente, $s_i = 1$ para i impar pero 0 para i par. No hay convergencia a un límite, y series tal como estas son llamadas oscilantes.

Para las series

$$1 + 2 + 3 + \dots + n + \dots$$

tenemos

$$s_n = \frac{n(n+1)}{2}$$

Cuando $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} s_n = \infty .$$

Cada vez que las sumas parciales diverjan (tienden a $\pm\infty$), la serie infinita se dice que *diverge*. A menudo el término divergente es extendido para incluir series oscilatorias.

Ya que evaluamos las sumas parciales por aritmética ordinaria, la serie convergente, definida en términos del límite de las sumas parciales, asume una posición de importancia suprema. Dos ejemplos pueden clarificar la naturaleza de convergencia o divergencia de una serie y servirá como una base para una investigación más detallada en la próxima sección.

Ejemplo Series geométricas.

La sucesión geométrica, comenzando con a y con una razón r ($r \geq 0$), está dado por

$$a + ar + ar^2 + ar^3 + \dots + ar^{n-1} + \dots .$$

La suma parcial n -ésima está dada por

$$s_n = a \frac{1 - r^n}{1 - r} \tag{8.3}$$

Tomando el límite cuando $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} s_n = \frac{a}{1-r}, \quad \text{para } r < 1. \quad (8.4)$$

De modo que, por definición, la serie geométrica infinita converge para $r < 1$ y está dada por

$$\sum_{n=1}^{\infty} ar^{n-1} = \frac{a}{1-r}. \quad (8.5)$$

Por otra parte, si $r \geq 1$, la condición necesaria $u_n \rightarrow 0$ no se satisface y la serie infinita diverge.

Ejemplo Series armónicas.

Consideremos la serie armónica

$$\sum_{n=1}^{\infty} n^{-1} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots. \quad (8.6)$$

Tenemos que el $\lim_{n \rightarrow \infty} u_n = \lim_{n \rightarrow \infty} 1/n = 0$, pero esto no es suficiente para garantizar la convergencia. Si agrupamos los términos (no cambiando el orden) como

$$1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \left(\frac{1}{9} + \cdots + \frac{1}{16}\right) + \cdots, \quad (8.7)$$

se verá que cada par de paréntesis encierra p términos de la forma

$$\frac{1}{p+1} + \frac{1}{p+2} + \cdots + \frac{1}{p+p} > \frac{p}{2p} = \frac{1}{2}. \quad (8.8)$$

Formando sumas parciales sumando un grupo entre paréntesis por vez, obtenemos

$$\begin{aligned} s_1 &= 1, & s_4 &> \frac{5}{2}, \\ s_2 &= \frac{3}{2}, & s_5 &> \frac{6}{2}, \\ s_3 &> \frac{4}{2}, & s_n &> \frac{n+1}{2}. \end{aligned} \quad (8.9)$$

Las series armónicas consideradas de esta manera ciertamente son divergentes. Una demostración independiente y alternativa de su divergencia aparece en la sección 8.2.

Usando el teorema del binomio, podríamos expandir la función $(1+x)^{-1}$:

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots + (-x)^{n-1} + \cdots. \quad (8.10)$$

Si tomamos $x \rightarrow 1$, la serie se convierte

$$1 - 1 + 1 - 1 + 1 - 1 + \cdots, \quad (8.11)$$

una serie que etiquetamos como oscilatoria anteriormente. Aunque no converge en el sentido usual, significa que puede ser ligada a su serie. Euler, por ejemplo, asignado un valor de $1/2$ a esta sucesión oscilatoria sobre la base de la correspondencia entre esta serie y la bien definida función $(1+x)^{-1}$. Desafortunadamente, tal correspondencia entre la serie y la función no es única y esta aproximación deberá ser redefinida. Otros métodos de asignar un significado a una serie oscilatoria o divergente, métodos de definir una suma, han sido desarrollados. Otro ejemplo de generalizar la convergencia lo vemos en las serie asintótica o semi-convergente, consideradas más adelante.

8.2. Pruebas de Convergencia

Aunque las series no convergentes pueden ser útiles en ciertos casos especiales, usualmente insistimos, como una materia de conveniencia si no de necesidad, que nuestras series sean convergentes. Por lo tanto esto llega a ser una materia de extrema importancia para ser capaz de decir si una serie dada es o no convergente. Desarrollaremos un número de posibles pruebas, comenzando con una prueba simple pero poco sensible y posteriormente trabajar con una más complicada pero muy sensible.

Por ahora consideremos una serie de términos positivos, $a_n > 0$, posponiendo los términos negativos hasta la próxima sección.

8.2.1. Pruebas de comparación.

Si término a término una serie de términos $u_n \leq a_n$, en el cual los a_n forman una serie convergente, las series $\sum_n u_n$ también es convergente. Simbólicamente, tenemos

$$\begin{aligned} \sum_n a_n &= a_1 + a_2 + a_3 + \cdots, & \text{convergente,} \\ \sum_n u_n &= u_1 + u_2 + u_3 + \cdots. \end{aligned}$$

Si $u_n \leq a_n$ para todo n , luego $\sum_n u_n \leq \sum_n a_n$ y $\sum_n u_n$ por lo tanto es convergente.

Si término a término es una serie de términos $v_n \geq b_n$, en el cual b_n forma una serie divergente, las series $\sum_n v_n$ también es divergente. Note que las comparaciones de u_n con b_n o v_n con a_n no dan información. Aquí tenemos

$$\begin{aligned} \sum_n b_n &= b_1 + b_2 + b_3 + \cdots, & \text{divergente,} \\ \sum_n v_n &= v_1 + v_2 + v_3 + \cdots. \end{aligned}$$

Si $v_n \geq b_n$ para todo n , luego $\sum_n v_n \geq \sum_n b_n$ y $\sum_n v_n$ por lo tanto es divergente.

Para las series convergente a_n tenemos las series geométricas, mientras las series armónicas servirán como las series divergentes b_n . En tanto otras series son identificadas como convergentes o divergentes, ellas pueden ser usadas como las series conocidas en estas pruebas de comparación.

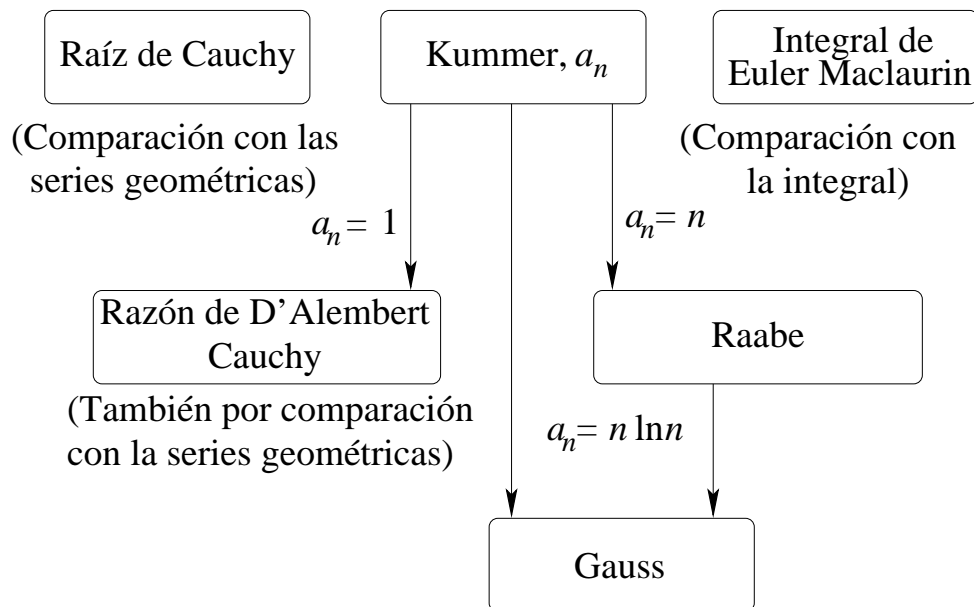


Figura 8.1: Prueba de comparación.

Todos las pruebas desarrolladas en esta sección son esencialmente pruebas de comparación. La figura 8.1 muestra estas pruebas y sus relaciones.

Ejemplo Las series p .

Probamos $\sum_n n^{-p}$, $p = 0.999$, por convergencia. Ya que $n^{-0.999} > n^{-1}$, y $b_n = n^{-1}$ forman la serie armónica divergente, la prueba de comparación muestra que $\sum_n n^{-0.999}$ es divergente. Generalizando, $\sum_n n^{-p}$ se ve como divergente para todo $p \leq 1$.

8.2.2. Prueba de la raíz de Cauchy.

Si $(a_n)^{1/n} \leq r < 1$ para todo n suficientemente grande, con r independiente de n , entonces $\sum_n a_n$ es convergente. Si $(a_n)^{1/n} \geq 1$ para todo n suficientemente grande, entonces $\sum_n a_n$ es divergente.

La primera parte de esta prueba se verifica fácilmente elevando $(a_n)^{1/n} \leq r$ a la n -ésima potencia. Obtenemos

$$a_n \leq r^n < 1 .$$

Ya que r^n es sólo el término n -ésimo en una serie geométrica convergente, $\sum_n a_n$ es convergente por la prueba de comparación. Conversamente, si $(a_n)^{1/n} \geq 1$, entonces $a_n \geq 1$ y la serie debería diverger. La prueba de la raíz es particularmente útil en establecer las propiedades de la serie de potencias.

8.2.3. Prueba de la razón de D' Alembert o Cauchy.

Si $a_{n+1}/a_n \leq r < 1$ para todo n suficientemente grande, y r independiente de n , entonces $\sum_n a_n$ es convergente. Si $a_{n+1}/a_n \geq 1$ de un n en adelante, entonces $\sum_n a_n$ es divergente.

La convergencia está dada por la comparación directa con las series geométricas ($1 + r + r^2 + \dots$). En la segunda parte $a_{n+1} \geq a_n$ y la divergencia debe ser razonablemente obvia. Aunque la prueba no es tan sensible como la prueba de la raíz de Cauchy, esta prueba de la razón e D' Alembert es una de las más fáciles de aplicar y es ampliamente usada. Una afirmación alternativa de la prueba de la razón está en la forma de un límite: si

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} < 1, & \quad \text{convergencia} \\ & > 1, \quad \text{divergencia} \\ & = 1, \quad \text{indeterminado.} \end{aligned} \tag{8.12}$$

A causa de la posibilidad de ser indeterminado, la prueba de la razón es probable que falle en puntos cruciales, y se hace necesario una prueba más delicada y sensible.

Podríamos preguntarnos cómo podría levantarse esta indeterminación. Realmente fue disimulado en el primera afirmación $a_{n+1}/a_n \leq r < 1$. Podríamos encontrar $a_{n+1}/a_n < 1$ para todo n finito pero ser inapropiado escoger un $r < 1$ e *independiente* de n tal que $a_{n+1}/a_n \leq r$ para todo n suficientemente grande. Un ejemplo está dado por las series armónicas

$$\frac{a_{n+1}}{a_n} = \frac{n}{n+1} < 1, \tag{8.13}$$

Ya que

$$\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = 1, \tag{8.14}$$

no existe una razón fija $r < 1$ y la prueba de la razón falla.

Ejemplo Prueba de la razón de D' Alembert.

Probar la convergencia de $\sum_n \frac{n}{2^n}$

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)/2^{n+1}}{n/2^n} = \frac{1}{2} \frac{n+1}{n}. \tag{8.15}$$

Ya que

$$\frac{a_{n+1}}{a_n} \leq \frac{3}{4} \quad \text{para } n \geq 2, \tag{8.16}$$

tenemos convergencia. Alternativamente,

$$\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = \frac{1}{2}, \tag{8.17}$$

y de nuevo converge.

8.2.4. Prueba integral de Cauchy o Maclaurin.

Esta es otra clase de prueba de comparación en la cual comparamos una serie con una integral. Geométricamente, comparamos el área de una serie de un rectángulo de ancho unitario con el área bajo la curva.

Sea $f(x)$ una función continua, monótonamente decreciente en la cual $f(n) = a_n$. Luego $\sum_n a_n$ converge si $\int_0^\infty f(x) dx$ es finita y diverge si la integral es infinita. Para la i -ésima suma parcial

$$s_i = \sum_{n=1}^i a_n = \sum_{n=1}^i f(n) . \quad (8.18)$$

Pero

$$s_i > \int_1^{i+1} f(x) dx , \quad (8.19)$$

por la figura 8.2a, $f(x)$ es monótonamente decreciente. Por otra parte, de la figura 8.2b,

$$s_i - a_1 < \int_1^i f(x) dx , \quad (8.20)$$

en la cual la serie está representada por los rectángulos inscritos. Tomando el límite como $i \rightarrow \infty$, tenemos

$$\int_1^\infty f(x) dx < \sum_{n=1}^\infty a_n < \int_1^\infty f(x) dx + a_1 . \quad (8.21)$$

De modo que la serie infinita converge o diverge cuando la integral correspondiente converge o diverge respectivamente.

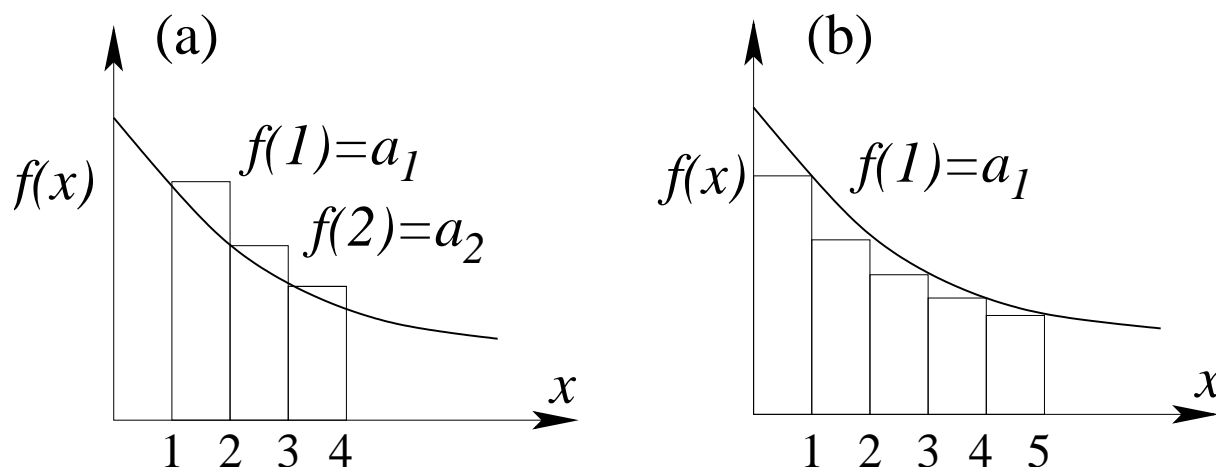


Figura 8.2: (a) Comparación de la integral y la suma de bloques sobresalientes. (b) Comparación de la integral y la suma de bloques envueltos.

La prueba de la integral es particularmente útil para acotar superior e inferiormente el resto de una serie, después de que algunos números de términos iniciales hayan sido sumados.

Esto es,

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^N a_n + \sum_{n=N+1}^{\infty} a_n ,$$

donde

$$\int_{N+1}^{\infty} f(x) dx < \sum_{n=N+1}^{\infty} a_n < \int_{N+1}^{\infty} f(x) dx + a_{N+1} .$$

Podemos liberar la prueba de la integral de los requerimientos muy restrictivos de que la función de interpolación $f(x)$ sea positiva y monótonamente decreciente, basta que la función $f(x)$ tenga una derivada continua que satisfaga

$$\sum_{n=N_i+1}^{N_f} f(n) = \int_{N_i}^{N_f} f(x) dx + \int_{N_i}^{N_f} (x - [x])f'(x) dx . \quad (8.22)$$

Aquí $[x]$ denota el entero mayor por debajo de x , tal que $x - [x]$ varía como diente de sierra entre 0 y 1.

Ejemplo Función Zeta de Riemann.

La función zeta de Riemann está definida por

$$\zeta(p) = \sum_{n=1}^{\infty} n^{-p} . \quad (8.23)$$

Podemos tomar $f(x) = x^{-p}$ y entonces

$$\int_1^{\infty} x^{-p} dx = \begin{cases} \left. \frac{x^{-p+1}}{-p+1} \right|_1^{\infty} , & p \neq 1 \\ \ln x \Big|_1^{\infty} , & p = 1 \end{cases} \quad (8.24)$$

La integral y por lo tanto la serie son divergentes para $p \leq 1$ y convergente para $p > 1$. De modo que la ecuación (8.23) lleva la condición de $p > 1$. Esto, incidentalmente, es una prueba independiente de que la serie armónica ($p = 1$) diverge y lo hace en forma logarítmica. La suma del primer millón de términos $\sum_{n=1}^{1,000,000} n^{-1}$, es solamente 14.392726...

Esta comparación con la integral también puede ser usada para dar una cota superior a la constante Euler-Mascheroni definida por

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{m=1}^n \frac{1}{m} - \ln n \right) . \quad (8.25)$$

Volviendo a las sumas parciales,

$$s_n = \sum_{m=1}^n m^{-1} - \ln n < \int_1^n \frac{dx}{x} - \ln n + 1 . \quad (8.26)$$

Evaluando la integral del lado derecho, $s_n < 1$ para todo n y por lo tanto $\gamma < 1$. Realmente la constante de Euler-Mascheroni es 0.57721566...

8.2.5. Prueba de Kummer.

Esta es la primera de tres pruebas que son algo más difíciles para aplicar que las anteriores. Su importancia radica en su poder y sensibilidad. Frecuentemente, al menos una de las tres funcionará cuando las pruebas más fáciles sean indeterminadas. Debe recordarse, sin embargo, que estas pruebas, como aquellas previamente discutidas, están finalmente basadas en comparaciones. Esto significa que todas las pruebas de convergencia dadas aquí, incluyendo la de Kummer, puedan fallar algunas veces.

Consideremos una serie de términos positivos u_i y una sucesión de constantes positivas finitas a_i . Si

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} \geq C > 0, \quad (8.27)$$

para todo $n \geq N$, algún número fijo, entonces $\sum_{i=1}^{\infty} u_i$ converge. Si

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} \leq 0 \quad (8.28)$$

y $\sum_{i=1}^{\infty} a_i^{-1}$ diverge, luego $\sum_{i=1}^{\infty} u_i$ diverge.

La prueba de este poderoso test es simple y queda como ejercicio.

Si las constantes positivas a_n de la prueba de Kummer son elegidas como $a_n = n$, tenemos la prueba de Raabe.

8.2.6. Prueba de Raabe.

Si $u_n > 0$ y si

$$n \left(\frac{u_n}{u_{n+1}} - 1 \right) \geq P > 1, \quad (8.29)$$

para todo $n \geq N$, donde N es un entero positivo independiente de n , entonces $\sum_i u_i$ converge. Si

$$n \left(\frac{u_n}{u_{n+1}} - 1 \right) \leq 1, \quad (8.30)$$

entonces $\sum_i u_i$ diverge ($\sum n^{-1}$ diverge).

La forma en límite en el test de Raabe es

$$\lim_{n \rightarrow \infty} n \left(\frac{u_n}{u_{n+1}} - 1 \right) = P. \quad (8.31)$$

Tenemos convergencia para $P > 1$, y divergencia para $P < 1$, y no hay prueba para $P = 1$ exactamente como con el test de Kummer. Esta indeterminancia está expresada en que podemos encontrar ejemplos de una serie convergente y una divergente en que ambas series tienden a $P = 1$ en la ecuación (8.31).

El test de Raabe es más sensible que la prueba de la razón de D'Alembert ya que $\sum_{n=1}^{\infty} n^{-1}$ diverge más lentamente que $\sum_{n=1}^{\infty} 1$. Obtenemos una prueba aún más sensible (y una relativamente fácil de aplicar) si escogemos $a_n = n \ln n$. Esto es la prueba de Gauss.

8.2.7. Prueba de Gauss.

Si $u_n > 0$ para todo n finito y

$$\frac{u_n}{u_{n+1}} = 1 + \frac{h}{n} + \frac{B(n)}{n^2}, \quad (8.32)$$

en el cual $B(n)$ es una función acotada de n para $n \rightarrow \infty$, luego $\sum_i u_i$ converge para $h > 1$ y diverge para $h \leq 1$.

La razón u_n/u_{n+1} de la ecuación (8.32) a menudo llega a ser como la razón de dos formas cuadráticas:

$$\frac{u_n}{u_{n+1}} = \frac{n^2 + a_1 n + a_0}{n^2 + b_1 n + b_0}. \quad (8.33)$$

Se puede mostrar que tenemos convergencia para $a_1 > b_1 + 1$ y divergencia para $a_1 \leq b_1 + 1$.

El test de Gauss es un test extremadamente sensible para la convergencia de series. Esto funcionará para prácticamente todas las series que encontraremos en Física. Para $h > 1$ o $h < 1$ la prueba se deduce directamente del test de Raabe

$$\lim_{n \rightarrow \infty} n \left[1 + \frac{h}{n} + \frac{B(n)}{n^2} - 1 \right] = \lim_{n \rightarrow \infty} \left[h + \frac{B(n)}{n} \right] = h. \quad (8.34)$$

Si $h = 1$, falla el test de Raabe. Sin embargo, si volvemos al test de Kummer y usamos $a_n = n \ln n$, tenemos

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left\{ n \ln n \left[1 + \frac{1}{n} + \frac{B(n)}{n^2} \right] - (n+1) \ln(n+1) \right\} \\ &= \lim_{n \rightarrow \infty} \left[n \ln n \cdot \frac{(n+1)}{n} - (n+1) \ln(n+1) \right] \\ &= \lim_{n \rightarrow \infty} (n+1) \left[\ln n - \ln n - \ln \left(1 + \frac{1}{n} \right) \right]. \end{aligned} \quad (8.35)$$

Pidiendo prestado un resultado de la sección 8.6 (el cual no es dependiente de la prueba de Gauss) tenemos

$$\lim_{n \rightarrow \infty} -(n+1) \ln \left(1 + \frac{1}{n} \right) = \lim_{n \rightarrow \infty} -(n+1) \left(\frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3} \dots \right) = -1 < 0. \quad (8.36)$$

De modo que tenemos divergencia para $h = 1$. Esto es un ejemplo de una aplicación exitosa del test de Kummer en el cual el test de Raabe falla.

Ejemplo Series de Legendre.

La relación de recurrencia para la solución en serie de la ecuación de Legendre pueden ser colocadas en la forma

$$\frac{a_{2j+2}}{a_{2j}} = \frac{2j(2j+1) - l(l+1)}{(2j+1)(2j+2)}. \quad (8.37)$$

Esto es equivalente a u_{2j+2}/u_{2j} para $x = +1$. Para $j \gg l$

$$\frac{a_{2j}}{a_{2j+2}} \rightarrow \frac{(2j+1)(2j+2)}{2j(2j+1)} = \frac{2j+2}{2j} = 1 + \frac{1}{j}. \quad (8.38)$$

Por la ecuación (8.33) la serie es divergente. Más adelante exigiremos que las series de Legendre sean finitas (se corten) para $x = 1$. Eliminaremos la divergencia ajustando los parámetros $n = 2j_0$, un entero par. Esto truncará la serie, convirtiendo la serie infinita en un polinomio.

8.2.8. Mejoramiento de convergencia.

En esta sección no nos preocupará establecer la convergencia como una propiedad matemática abstracta. En la práctica, la razón de convergencia puede ser de considerable importancia. Aquí presentamos un método que mejora la razón de la convergencia de una serie ya convergente.

El principio básico de este método, debido a Kummer, es formar una combinación lineal de nuestra serie lentamente convergente y una o más series cuya suma es conocida. Entre las series conocidas, la colección

$$\begin{aligned}\alpha_1 &= \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = 1, \\ \alpha_2 &= \sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)} = \frac{1}{4}, \\ \alpha_3 &= \sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)(n+3)} = \frac{1}{18}, \\ &\vdots \\ \alpha_p &= \sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)\cdots(n+p)} = \frac{1}{p \cdot p!},\end{aligned}$$

es particularmente útil. Las series están combinadas término a término y los coeficientes en combinación lineal son escogidos para cancelar los términos que convergen lentamente.

Ejemplo Función zeta de Riemann, $\zeta(3)$.

Sea la serie a ser sumada $\sum_{n=1}^{\infty} n^{-3}$. En la sección 8.10 está identificada como una función zeta de Riemann, $\zeta(3)$. Formamos una combinación lineal

$$\sum_{n=1}^{\infty} n^{-3} + a_2 \alpha_2 = \sum_{n=1}^{\infty} n^{-3} + \frac{a_2}{4}.$$

α_1 no está incluida ya que converge más lentamente que $\zeta(3)$. Combinando términos, obtenemos sobre la mano izquierda

$$\sum_{n=1}^{\infty} \left\{ \frac{1}{n^3} + \frac{a_2}{n(n+1)(n+2)} \right\} = \sum_{n=1}^{\infty} \frac{n^2(1+a_2) + 3n + 2}{n^3(n+1)(n+2)}.$$

Si escogemos $a_2 = -1$, la ecuación precedente tiende a

$$\zeta(3) = \sum_{n=1}^{\infty} n^{-3} = \frac{1}{4} + \sum_{n=1}^{\infty} \frac{3n+2}{n^3(n+1)(n+2)}. \quad (8.39)$$

La serie resultante no es muy bonita pero converge como n^{-4} , apreciablemente más rápido que n^{-3} .

El método puede ser extendido incluyendo $a_3\alpha_3$ para obtener la convergencia como n^{-5} , $a_4\alpha_4$ para obtener la convergencia como n^{-6} , etc. Eventualmente, usted tiene que alcanzar un compromiso entre cuánta álgebra usted hace y cuánta aritmética la computadora hace. Como las computadoras lo hacen más rápido, el balance está seguramente sustituyendo menos álgebra hecha por usted, por más aritmética realizada por el computador.

8.3. Series alternadas.

En la sección 8.2 nos limitamos a series de términos positivos. Ahora, en contraste, consideraremos series infinitas en las cuales los signos se alternan. La cancelación parcial debida a la alternancia de los signos hace la convergencia más rápida y mucho más fácil de identificar. Probaremos que el criterio de Leibniz es una condición general para la convergencia de una serie alternada.

8.3.1. Criterio de Leibniz.

Consideremos la serie $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$ con $a_n > 0$. Si a_n es monótonamente decreciente (para N suficientemente grande) y el $\lim_{n \rightarrow \infty} a_n = 0$, entonces la serie converge.

Para probar esto, examinemos las sumas parciales pares

$$\begin{aligned} s_{2n} &= a_1 - a_2 + a_3 - \dots - a_{2n} , \\ s_{2n+2} &= s_{2n} + (a_{2n+1} - a_{2n+2}) . \end{aligned} \tag{8.40}$$

Ya que $a_{2n+1} > a_{2n+2}$, tenemos

$$s_{2n+2} > s_{2n} . \tag{8.41}$$

Por otra parte,

$$s_{2n+2} = a_1 - (a_2 - a_3) - (a_4 - a_5) - \dots - a_{2n+2} . \tag{8.42}$$

De modo que, con cada par de términos $a_{2p} - a_{2p+1} > 0$,

$$s_{2n+2} < a_1 . \tag{8.43}$$

Con las sumas parciales pares acotamos $s_{2n} < s_{2n+2} < a_1$ y los términos a_n decrecen monótonamente aproximándose a cero, esta serie alternada converge.

Un resultado más importante puede ser extraído de las sumas parciales. A partir de las diferencias entre el límite de la serie S y las sumas parciales s_n

$$\begin{aligned} S - s_n &= a_{n+1} - a_{n+2} + a_{n+3} - a_{n+4} + \dots \\ &= a_{n+1} - (a_{n+2} - a_{n+3}) - (a_{n+4} - a_{n+5}) - \dots \end{aligned} \tag{8.44}$$

o

$$S - s_n < a_{n+1} . \tag{8.45}$$

La ecuación (8.45) dice que el error en el corte de una serie alternada después de n términos es menor que a_{n+1} , el primer término excluido. Un conocimiento del error obtenido de esta manera puede ser de gran importancia práctica.

8.3.2. Convergencia absoluta.

Dada una serie en términos de u_n en la cual u_n puede variar en signo, si $\sum |u_n|$ converge, entonces $\sum u_n$ se dice que es *absolutamente convergente*. Si $\sum u_n$ converge pero $\sum |u_n|$ diverge, la convergencia recibe el nombre de *condicional*.

La serie alternada armónica es un ejemplo simple de esta convergencia condicionada. Tenemos

$$\sum_{n=1}^{\infty} (-1)^{n-1} n^{-1} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + \frac{1}{n} - \cdots \quad (8.46)$$

convergente por el criterio de Leibniz, pero

$$\sum_{n=1}^{\infty} n^{-1} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} + \cdots$$

se ha demostrado que es divergente en la sección 8.1 y 8.2.

Podemos notar que todas las pruebas desarrolladas en la sección 8.2 supone una serie de términos positivos. Por lo tanto, todas las pruebas en esa sección garantizan la convergencia absoluta.

Ejemplo

Para $0 < x < \pi$ la serie de Fourier

$$\sum_{n=1}^{\infty} \frac{\cos(nx)}{n} = -\ln \left(2 \operatorname{sen} \frac{x}{2} \right), \quad (8.47)$$

converge teniendo coeficientes que cambian de signo frecuentemente, pero no tanto para que el criterio de convergencia de Leibniz se aplique fácilmente. Apliquemos el test de la integral de la ecuación (8.22). Usando integración por partes vemos de inmediato que

$$\int_1^{\infty} \frac{\cos(nx)}{n} dn = \left[\frac{\operatorname{sen}(nx)}{nx} \right]_1^{\infty} - \frac{1}{x} \int_1^{\infty} \frac{\operatorname{sen}(nx)}{n^2} dn$$

converge para $n \rightarrow \infty$, y la integral del lado derecho incluso converge absolutamente. El término derivado en la ecuación (8.22) tiene la forma

$$\int_1^{\infty} (n - [n]) \left\{ -\frac{x}{n} \operatorname{sen}(nx) - \frac{\cos(nx)}{n^2} \right\} dn,$$

donde el segundo término converge absolutamente y no necesita ser considerado. Lo próximo es observar que $g(N) = \int_1^N (n - [n]) \operatorname{sen}(nx) dn$ es acotado para $N \rightarrow \infty$, tal como $\int_1^N \operatorname{sen}(nx) dn$ es acotado debido a la naturaleza periódica de $\operatorname{sen}(nx)$ y a su regular cambio de signo. Usando integración por partes nuevamente

$$\int_1^{\infty} \frac{g'(n)}{n} dn = \left[\frac{g(n)}{n} \right]_1^{\infty} + \int_1^{\infty} \frac{g(n)}{n^2} dn,$$

vemos que el segundo término es absolutamente convergente, y el primero va a cero en el límite superior. Por lo tanto la serie en la ecuación (8.47) converge, lo cual es duro de ver usando otro test de convergencia.

8.4. Álgebra de series.

Establecer la convergencia absoluta es importante porque puede probarse que las series absolutamente convergentes pueden ser manipuladas de acuerdo a las reglas familiares del álgebra o aritmética.

1. Si una serie infinita es absolutamente convergente, la suma de la serie es independiente del orden en el cual los términos son añadidos.
2. La serie puede ser multiplicada por otra serie absolutamente convergente. El límite del producto será el producto de los límites de las series individuales. El producto de las series, una doble serie, también será absolutamente convergente.

No hay tales garantías en series condicionalmente convergentes. Nuevamente consideremos la serie armónica alternada. Si escribimos

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots = 1 - \left(\frac{1}{2} - \frac{1}{3}\right) - \left(\frac{1}{4} - \frac{1}{5}\right) - \dots, \quad (8.48)$$

es claro que la suma

$$\sum_{n=1}^{\infty} (-1)^{n-1} n^{-1} < 1. \quad (8.49)$$

Sin embargo, si rearrreglamos los términos sutilmente, podemos hacer que la serie armónica alternada converja a $3/2$. Reagrupamos los términos de la ecuación (8.48), tomando

$$\begin{aligned} & \left(1 + \frac{1}{3} + \frac{1}{5}\right) - \frac{1}{2} + \left(\frac{1}{7} + \frac{1}{9} + \frac{1}{11} + \frac{1}{13} + \frac{1}{15}\right) - \frac{1}{4} \\ & + \left(\frac{1}{17} + \dots + \frac{1}{25}\right) - \frac{1}{6} + \left(\frac{1}{27} + \dots + \frac{1}{35}\right) - \frac{1}{8} + \dots. \end{aligned} \quad (8.50)$$

Tratando los términos agrupados en paréntesis como términos simples por conveniencia, obtenemos las sumas parciales

$$\begin{array}{ll} s_1 = 1.5333 & s_2 = 1.0333 \\ s_3 = 1.5218 & s_4 = 1.2718 \\ s_5 = 1.5143 & s_6 = 1.3476 \\ s_7 = 1.5103 & s_8 = 1.3853 \\ s_9 = 1.5078 & s_{10} = 1.4078 \end{array}$$

A partir de esta tabulación de los s_n y el gráfico de s_n versus n en la figura 8.3 es clara la convergencia a $3/2$. Hemos rearrreglado los términos, tomando términos positivos hasta que la suma parcial sea igual o mayor que $3/2$, luego sumando los términos negativos hasta que la suma parcial caiga bajo $3/2$, etc. Como las series se extienden hasta infinito, todos los términos originales eventualmente aparecerán, pero las sumas parciales de este reordenamiento de esta serie armónica alternada converge a $3/2$. Por un reordenamiento de términos una serie condicionalmente convergente podría ser hecha para converger a algún valor deseado o para que diverja. Esta afirmación es dada como el teorema de Riemann. Obviamente, series condicionalmente convergentes deberían ser tratadas con precaución.

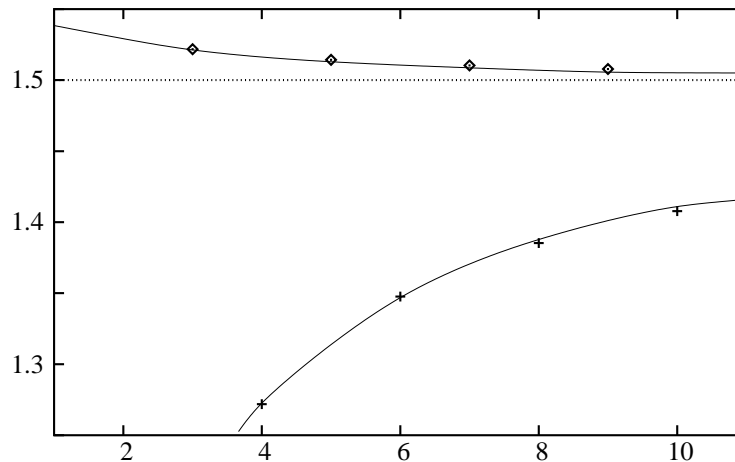


Figura 8.3: Serie armónica alternada, rearrreglo de términos para dar convergencia a 1.5.

8.4.1. Mejoramiento de la convergencia, aproximaciones racionales.

La serie

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}, \quad -1 < x \leq 1, \quad (8.51)$$

converge muy suavemente cuando x se aproxima a $+1$. La razón de convergencia podría ser mejorada sustancialmente multiplicando ambos lados de la ecuación (8.51) por un polinomio y ajustando los coeficientes del polinomio para cancelar las porciones que convergen más lentamente en la serie. Consideremos la posibilidad más simple: Multiplicar $\ln(1+x)$ por $1+a_1x$.

$$(1+a_1x)\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} + a_1 \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^{n+1}}{n}.$$

Combinando las dos series sobre la derecha término a término, obtenemos

$$\begin{aligned} (1+a_1x)\ln(1+x) &= x + \sum_{n=2}^{\infty} (-1)^{n-1} \left(\frac{1}{n} - \frac{a_1}{n-1} \right) x^n \\ &= x + \sum_{n=2}^{\infty} (-1)^{n-1} \frac{n(1-a_1) - 1}{n(n-1)} x^n. \end{aligned}$$

Claramente, si tomamos $a_1 = 1$, el n en el numerador desaparece y nuestra serie combinada converge como n^{-2} .

Continuando este proceso, encontramos que $(1+2x+x^2)\ln(1+x)$ se anula como n^{-3} , $(1+3x+3x^2+x^3)\ln(1+x)$ se anula cuando n^{-4} . En efecto estamos desplazándonos desde una expansión de serie simple de la ecuación (8.51) a una representación racional en la cual

la función $\ln(1 + x)$ está representada por la razón de una serie y un polinomio:

$$\ln(1 + x) = \frac{x + \sum_{n=1}^{\infty} \frac{(-1)^n x^n}{n(n-1)}}{1 + x} .$$

Tales aproximaciones racionales pueden ser ambas compactas y precisas. Los programas computacionales hacen extensivo el uso de ellas.

8.4.2. Reordenamiento de series dobles.

Otro aspecto del reordenamiento de series aparece en el tratamiento de series dobles (figura 8.4):

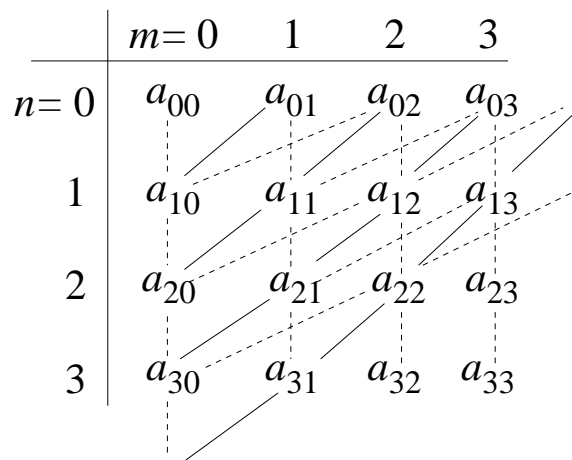


Figura 8.4: Series dobles, la suma sobre n es indicada por líneas segmentadas verticales.

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{n,m} .$$

sustituyamos

$$\begin{aligned} n &= q \geq 0 , \\ m &= p - q \geq 0 , \\ (q &\leq p) . \end{aligned}$$

Esto resulta en la identidad

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{n,m} = \sum_{p=0}^{\infty} \sum_{q=0}^p a_{q,p-q} . \tag{8.52}$$

La suma sobre p y q de la ecuación (8.52) está ilustrada en la figura 8.5. La sustitución

$$n = s \geq 0 , \quad m = r - 2s \geq 0 , \quad \left(s \leq \frac{r}{2} \right)$$

	$p= 0$	1	2	3
$q= 0$	a_{00}	a_{01}	a_{02}	a_{03}
1		a_{10}	a_{11}	a_{12}
2			a_{20}	a_{21}
3				a_{30}

Figura 8.5: Series dobles nuevamente, la primera suma es representada por líneas segmentadas verticales pero estas líneas verticales corresponden a las diagonales en la figura 8.4.

tiende a

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{n,m} = \sum_{r=0}^{\infty} \sum_{s=0}^{[r/2]} a_{s,r-2s} . \tag{8.53}$$

con $[r/2] = r/2$ para r par, $(r - 1)/2$ para r impar. La suma sobre r y s de la ecuación (8.53) está mostrada en la figura 8.6. Las ecuaciones (8.52) y (8.53) son claramente reordenamientos del arreglo de coeficientes $a_{n,m}$, reordenamientos que son válidos en tanto tengamos convergencia absoluta. La combinación de las ecuaciones (8.52) y (8.53),

	$r= 0$	1	2	3	4
$s= 0$	a_{00}	a_{01}	a_{02}	a_{03}	a_{04}
1			a_{10}	a_{11}	a_{12}
2					a_{20}

Figura 8.6: Series dobles. La suma sobre s corresponde a la suma a lo largo de la líneas segmentadas inclinadas, en la figura 8.4.

$$\sum_{p=0}^{\infty} \sum_{q=0}^p a_{q,p-q} = \sum_{r=0}^{\infty} \sum_{s=0}^{[r/2]} a_{s,r-2s} . \tag{8.54}$$

es usada en la determinación de la forma en serie de los polinomios de Legendre.

8.5. Series de funciones.

Extendemos nuestro concepto de series infinitas para incluir la posibilidad que cada término u_n pueda ser una función de alguna variable, $u_n = u_n(x)$. Numerosas ilustraciones de tales series de funciones aparecerán más adelante. Las sumas parciales llegan a ser funciones de la variable x

$$s_n(x) = u_1(x) + u_2(x) + \cdots + u_n(x) , \quad (8.55)$$

tal como lo hacemos para la suma de serie, definimos el límite como el límite de las sumas parciales

$$\sum_{n=1}^{\infty} u_n(x) = S(x) = \lim_{n \rightarrow \infty} s_n(x) . \quad (8.56)$$

Hasta ahora nos hemos ocupado del comportamiento de las sumas parciales como una función de n . Ahora consideremos cómo las cantidades anteriores dependen de x . Aquí el concepto clave es la convergencia uniforme.

8.5.1. Convergencia uniforme.

Si para cualquier $\varepsilon > 0$ pequeño, existe un número N , independiente de x en el intervalo $[a, b]$ con $(a \leq x \leq b)$ tal que

$$|S(x) - s_n(x)| < \varepsilon , \forall n \geq N , \quad (8.57)$$

se dice que la serie converge uniformemente en el intervalo $[a, b]$. Esto dice que para que nuestra serie sea uniformemente convergente, debe ser posible encontrar un N finito tal que la cola de la serie infinita, $|\sum_{i=N+1}^{\infty} u_i(x)|$, sea menor que un ε arbitrariamente pequeño para todo x en el intervalo dado.

Esta condición, ecuación (8.57), la cual define la convergencia uniforme, es ilustrada en la figura 8.7. El punto es que no importa cuan pequeño sea ε podemos siempre tomar un n suficientemente grande tal que la magnitud absoluta de la diferencia entre $S(x)$ y $s_n(x)$ sea menor que ε para todo x , $a \leq x \leq b$. Si esto no puede ser hecho, entonces $\sum u_n(x)$ no es uniformemente convergente en el intervalo $[a, b]$.

Ejemplo

$$\sum_{n=1}^{\infty} u_n(x) = \sum_{n=1}^{\infty} \frac{x}{[(n-1)x+1][nx+1]} . \quad (8.58)$$

La suma parcial $s_n(x) = nx(nx+1)^{-1}$ puede ser verificada por inducción matemática. Por inspección esta expresión para $s_n(x)$ es válida para $n = 1, 2$. Suponemos que se mantiene

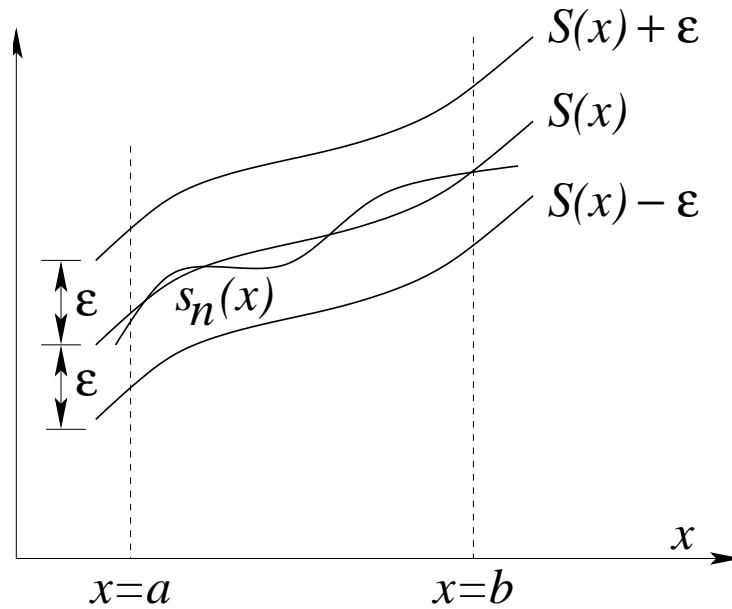


Figura 8.7: Convergencia uniforme.

para el término n y probamos para $n + 1$.

$$\begin{aligned} s_{n+1} &= s_n + \frac{x}{[nx + 1][(n + 1)x + 1]} \\ &= \frac{nx}{[nx + 1]} + \frac{x}{[nx + 1][(n + 1)x + 1]} \\ &= \frac{(n + 1)x}{(n + 1)x + 1}, \end{aligned}$$

completando la prueba.

Tomando $n \rightarrow \infty$ tenemos

$$\begin{aligned} S(0) &= \lim_{n \rightarrow \infty} s_n(0) = 0, \\ S(x \neq 0) &= \lim_{n \rightarrow \infty} s_n(x \neq 0) = 1. \end{aligned}$$

Tenemos una discontinuidad en el límite de la serie en $x = 0$. Sin embargo, $s_n(x)$ es una función continua de x , en el intervalo $0 \leq x < 1$, para todo n finito. La ecuación (8.57) con ε suficientemente pequeño, será violado para todo n finito. Nuestra serie no converge uniformemente.

8.5.2. Prueba M de Weierstrass.

La prueba más comúnmente usada para la convergencia uniforme es la prueba M de Weierstrass. Si podemos construir una serie de números $\sum_1^\infty M_i$, en la cual $M_i \geq |u_i(x)|$ para todo x en el intervalo $[a, b]$ y $\sum_1^\infty M_i$ es convergente, nuestra serie $\sum_1^\infty u_i(x)$ será *uniformemente* convergente en $[a, b]$.

La prueba de este test M de Weierstrass es directa y simple. Ya que $\sum_i M_i$ converge, existen algunos números N tal que $n + 1 \geq N$,

$$\sum_{i=n+1}^{\infty} M_i < \varepsilon . \quad (8.59)$$

Esto a partir de nuestra definición de convergencia. Entonces, con $|u_i(x)| \leq M_i$ para todo x en el intervalo $a \leq x \leq b$,

$$\sum_{i=n+1}^{\infty} |u_i(x)| < \varepsilon . \quad (8.60)$$

De modo que

$$|S(x) - s_n(x)| = \left| \sum_{i=n+1}^{\infty} u_i(x) \right| < \varepsilon , \quad (8.61)$$

y por definición $\sum_1^{\infty} u_i(x)$ es uniformemente convergente en $[a, b]$. Ya que tenemos especificados valores absolutos en el planteamiento de la prueba M de Weierstrass, la serie $\sum_1^{\infty} u_i(x)$ también es vista como serie *absolutamente* convergente.

Podemos notar que la convergencia uniforme y convergencia absoluta son propiedades independientes. Una no implica la otra. Para ejemplos específicos,

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n + x^2} , \quad -\infty < x < \infty \quad (8.62)$$

y

$$\sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n} = \ln(1 + x) , \quad 0 \leq x \leq 1 , \quad (8.63)$$

converge uniformemente en los intervalos indicados pero no converge absolutamente. Por otra parte,

$$\sum_{n=1}^{\infty} (1 - x)x^n = \begin{cases} 1 , & 0 \leq x < 1 \\ 0 , & x = 1 \end{cases} , \quad (8.64)$$

converge absolutamente pero no uniformemente en $[0, 1]$.

A partir de la definición de convergencia uniforme podríamos mostrar que cualquier serie

$$f(x) = \sum_{n=1}^{\infty} u_n(x) , \quad (8.65)$$

no puede converger uniformemente en ningún intervalo que incluya una discontinuidad de $f(x)$.

Ya que la prueba M de Weierstrass establece tanto la convergencia uniforme como absoluta, necesariamente falla para series que son uniformes pero condicionalmente convergentes.

8.5.3. Prueba de Abel.

Una prueba algo más delicada para la convergencia uniforme ha sido dada por Abel. Si

$$u_n(x) = a_n f_n(x) ,$$

$$\sum a_n = A , \quad \text{convergente,}$$

y las funciones $f(x)$ son monótonas [$f_{n+1}(x) \leq f_n(x)$] y acotadas, $0 \leq f_n(x) \leq M$, para todo x en $[a, b]$, entonces $\sum u_n(x)$ converge uniformemente en $[a, b]$.

Las series uniformemente convergentes tienen tres propiedades particularmente útiles.

1. Si los términos individuales $u_n(x)$ son continuos, la suma de la serie

$$f(x) = \sum_{n=1}^{\infty} u_n(x) , \quad (8.66)$$

es también continua.

2. Si los términos individuales $u_n(x)$ son continuos, las series pueden ser integradas término a término. La suma de las integrales es igual a la integral de la suma.

$$\int_a^b f(x) dx = \sum_{n=1}^{\infty} \int_a^b u_n(x) dx . \quad (8.67)$$

3. Las derivadas de la suma de la serie $f(x)$ es igual a la suma de los términos individuales derivados,

$$\frac{df(x)}{dx} = \sum_{n=1}^{\infty} \frac{du_n(x)}{dx} , \quad (8.68)$$

siempre que las siguientes condiciones sean satisfechas:

$$u_n(x) \text{ y } \frac{du_n(x)}{dx} \text{ son continuas en } [a, b].$$

$$\sum_{n=1}^{\infty} \frac{du_n(x)}{dx} \text{ es uniformemente convergente en } [a, b].$$

La integración término a término de una serie uniformemente convergente² requiere sólo continuidad de los términos individuales. Esta condición casi siempre es satisfecha en las aplicaciones físicas. La diferenciación término a término de una serie a menudo no es válida porque deben satisfacer condiciones más restrictivas. Por cierto, encontraremos casos en series de Fourier, en la cual la diferenciación término a término de una serie uniformemente convergente tiende a una serie divergente.

²La integración término a término también puede ser válida en ausencia de convergencia uniforme.

8.6. Expansión de Taylor.

Esta es una expansión de una función en una serie infinita o en una serie finita más un término remanente. Los coeficientes de los términos sucesivos de la serie involucra las derivadas sucesivas de la función. Este tipo de expansiones de son ampliamente usadas. Ahora derivaremos la expansión de Taylor.

Supongamos que nuestra función $f(x)$ tiene una derivada n -ésima continua en el intervalo $a \leq x \leq b$. Entonces, integrando esta n -ésima derivada n veces,

$$\begin{aligned} \int_a^x f^{(n)}(x) dx &= f^{(n-1)}(x) \Big|_a^x = f^{(n-1)}(x) - f^{(n-1)}(a) \\ \int_a^x \left(\int_a^x f^{(n)}(x) dx \right) dx &= \int_a^x [f^{(n-1)}(x) - f^{(n-1)}(a)] dx \\ &= f^{(n-2)}(x) - f^{(n-2)}(a) - (x-a)f^{(n-1)}(a) . \end{aligned} \quad (8.69)$$

Continuando, obtenemos

$$\int \int \int_a^x f^{(n)}(x)(dx)^3 = f^{(n-3)}(x) - f^{(n-3)}(a) - (x-a)f^{(n-2)}(a) - \frac{(x-a)^2}{2} f^{(n-1)}(a) . \quad (8.70)$$

Finalmente, integrando por n -ésima vez,

$$\begin{aligned} \int_a^x \cdots \int_a^x f^{(n)}(x)(dx)^n &= f(x) - f(a) - (x-a)f'(a) + \\ &\quad - \frac{(x-a)^2}{2!} f''(a) - \cdots - \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) . \end{aligned} \quad (8.71)$$

Note que esta expresión es exacta. No hay términos que hayan sido excluidos, ni aproximaciones hechas. Ahora, resolviendo para $f(x)$, tenemos

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!} f''(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) + R_n . \quad (8.72)$$

El remanente, R_n , está dado por la integral n -dimensional

$$\int_a^x \cdots \int_a^x f^{(n)}(x)(dx)^n . \quad (8.73)$$

Este remanente, ecuación (8.73), puede ser puesto en una forma más inteligible usando la forma integral del teorema del valor medio

$$\int_a^x g(x) dx = (x-a)g(\xi) , \quad (8.74)$$

con $a \leq \xi \leq x$. Integrando n veces obtenemos la forma Lagrangiana del remanente:

$$R_n = \frac{(x-a)^n}{n!} f^{(n)}(\xi) . \quad (8.75)$$

Con la expansión de Taylor en esta forma no estamos interesados en cualquier pregunta de convergencia de series infinitas. Esta serie es finita, la sola pregunta que nos importa es la magnitud del remanente.

Cuando la función $f(x)$ es tal que

$$\lim_{n \rightarrow \infty} R_n = 0, \quad (8.76)$$

la ecuación (8.72) se convierte en la serie de Taylor

$$\begin{aligned} f(x) &= f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \dots \\ &= \sum_{n=0}^{\infty} \frac{(x-a)^n}{n!} f^{(n)}(a). \end{aligned} \quad (8.77)$$

Nuestra serie de Taylor especifica el valor de una función en un punto, x , en términos del valor de la función y sus derivadas en un punto de referencia, a . Esta es una expansión en potencias de un cambio en la variable, $\Delta x = x - a$ en este caso. La notación puede ser variada según la conveniencia del usuario. Con la sustitución $x \rightarrow x + h$ y $a \rightarrow x$ tenemos una forma alterna

$$f(x+h) = \sum_{n=0}^{\infty} \frac{h^n}{n!} f^{(n)}(x).$$

Cuando usamos el operador $D = d/dx$ la expansión de Taylor se convierte en

$$f(x+h) = \sum_{n=0}^{\infty} \frac{h^n D^n}{n!} f(x) = e^{hD} f(x).$$

Un forma en operadores equivalente de la expansión e Taylor. Una derivación de la expansión de Taylor en el contexto de la teoría de variable compleja aparece en el próximo capítulo.

8.6.1. Teorema de Maclaurin.

Si expandimos alrededor del origen ($a = 0$), la ecuación (8.77) es conocida como la serie de Maclaurin

$$\begin{aligned} f(x) &= f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \dots \\ &= \sum_{n=0}^{\infty} \frac{x^n}{n!} f^{(n)}(0). \end{aligned} \quad (8.78)$$

Una aplicación inmediata de la serie de Maclaurin (o serie de Taylor) está en la expansión de varias funciones transcendentales en una serie infinita.

Ejemplo

Sea $f(x) = e^x$. Diferenciando, tenemos

$$f^{(n)}(0) = 1, \quad (8.79)$$

para todo n , $n = 1, 2, 3, \dots$. Entonces, para la ecuación (8.78), tenemos

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}. \quad (8.80)$$

Esta es la expansión en serie de la función exponencial. Algunos autores usan esta serie para definir la función exponencial.

Aunque esta serie es claramente convergente para todo x , podríamos chequear el término remanente, R_n . Por la ecuación (8.75) tenemos

$$\begin{aligned} R_n &= \frac{x^n}{n!} f^{(n)}(\xi) \\ &= \frac{x^n}{n!} e^{\xi}, \quad 0 \leq |\xi| \leq x. \end{aligned} \quad (8.81)$$

Por lo tanto

$$|R_n| \leq \frac{x^n}{n!} e^x \quad (8.82)$$

y

$$\lim_{n \rightarrow \infty} R_n = 0 \quad (8.83)$$

para todo los valores finitos de x , el cual indica que esta expansión de Maclaurin de e^x es válida sobre el intervalo $-\infty < x < \infty$.

Ejemplo

Sea $f(x) = \ln(1+x)$. Diferenciando, obtenemos

$$\begin{aligned} f'(x) &= \frac{1}{(1+x)}, \\ f^{(n)}(x) &= (-1)^{n-1} (n-1)! \frac{1}{(1+x)^n}. \end{aligned} \quad (8.84)$$

La expansión de Maclaurin produce

$$\begin{aligned} \ln(1+x) &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots + R_n \\ &= \sum_{p=1}^n (-1)^{p-1} \frac{x^p}{p} + R_n. \end{aligned} \quad (8.85)$$

En este caso el remanente está dado por

$$\begin{aligned} R_n &= \frac{x^n}{n!} f^{(n)}(\xi), \quad 0 \leq \xi \leq x \\ &\leq \frac{x^n}{n}, \quad 0 \leq \xi \leq x \leq 1. \end{aligned} \quad (8.86)$$

Ahora el remanente se aproxima a cero cuando n crece indefinidamente, dado $0 \leq x \leq 1$ ³. Como una serie infinita

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}, \quad (8.87)$$

la cual converge para $-1 < x \leq 1$. El intervalo $-1 < x < 1$ es fácilmente establecido por la prueba de la razón de D' Alembert. La convergencia en $x = 1$ se deduce a partir del criterio de Leibniz. En particular, en $x = 1$, tenemos

$$\begin{aligned} \ln 2 &= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots \\ &= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}, \end{aligned} \quad (8.88)$$

la serie armónica alterna condicionalmente convergente.

8.6.2. Teorema Binomial.

Una segunda, aplicación extremadamente importante de las expansiones de Taylor y Maclaurin es la derivación del teorema binomial para potencias negativas y/o no enteras.

Sea $f(x) = (1+x)^m$, en la cual m puede ser negativo y no está limitado a valores enteros. La aplicación directa de la ecuación (8.78) da

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \dots + R_n. \quad (8.89)$$

Para esta función el remanente es

$$R_n = \frac{x^n}{n!} (1+\xi)^{m-n} \times m(m-1) \cdots (m-n+1) \quad (8.90)$$

y ξ con $0 \leq \xi \leq x$. Ahora, para $n > m$, $(1+\xi)^{m-n}$ es un máximo para $\xi = 0$. Por lo tanto

$$R_n \leq \frac{x^n}{n!} \times m(m-1) \cdots (m-n+1). \quad (8.91)$$

Note que los factores dependientes de m no dan un cero a menos que m sea entero no negativo; R_n tiende a cero cuando $n \rightarrow \infty$ si x está restringido al intervalo $0 \leq x \leq 1$. La expansión binomial resulta

$$(1+x)^m = 1 + mx + \frac{m(m-1)}{2!}x^2 + \frac{m(m-1)(m-2)}{3!}x^3 + \dots. \quad (8.92)$$

En otra notación equivalente

$$\begin{aligned} (1+x)^m &= \sum_{n=0}^{\infty} \frac{m!}{n!(m-n)!} x^n \\ &= \sum_{n=0}^{\infty} \binom{m}{n} x^n. \end{aligned} \quad (8.93)$$

³Este intervalo puede ser fácilmente extendido a $-1 < x \leq 1$ pero no a $x = -1$.

Cuando la cantidad $\binom{m}{n}$ es igual a $m!/(n!(m-n)!)$, es llamado el *coeficiente binomial*. Aunque hemos mostrado solamente que el remanente se anula,

$$\lim_{n \rightarrow \infty} R_n = 0 ,$$

para $0 \leq x < 1$, realmente puede mostrarse que la serie en la ecuación (8.92) converge en el intervalo extendido $-1 < x < 1$. Para m un entero, $(m-n)! = \pm\infty$ si $n > m$ y las series automáticamente terminan en $n = m$.

Ejemplo Energía relativista.

La energía total relativista de una partícula es

$$E = mc^2 \left(1 - \frac{v^2}{c^2}\right)^{-1/2} . \quad (8.94)$$

Comparemos esta ecuación con la energía cinética clásica, $\frac{1}{2}mv^2$.

Por la ecuación (8.92) con $x = -\frac{v^2}{c^2}$ y $m = -\frac{1}{2}$ tenemos

$$E = mc^2 \left[1 - \frac{1}{2} \left(-\frac{v^2}{c^2}\right) + \frac{(-1/2)(-3/2)}{2!} \left(-\frac{v^2}{c^2}\right)^2 + \frac{(-1/2)(-3/2)(-5/2)}{3!} \left(-\frac{v^2}{c^2}\right)^3 + \dots \right] .$$

o

$$E = mc^2 + \frac{1}{2}mv^2 + \frac{3}{8}mv^2 \frac{v^2}{c^2} + \frac{5}{16}mv^2 \left(\frac{v^2}{c^2}\right)^2 + \dots . \quad (8.95)$$

El primer término, mc^2 , lo identificamos como la masa en reposo. Entonces

$$E_{\text{cinética}} = \frac{1}{2}mv^2 \left[1 + \frac{3}{4} \frac{v^2}{c^2} + \frac{5}{8} \left(\frac{v^2}{c^2}\right)^2 + \dots \right] . \quad (8.96)$$

Para la velocidad de la partícula $v \ll c$, donde c es la velocidad de la luz, la expresión en los paréntesis cuadrados se reduce a la unidad y vemos que la porción cinética de la energía relativista total concuerda con el resultado clásico.

Para polinomios podemos generalizar la expansión binomial a

$$(a_1 + a_2 + \dots + a_m)^n = \sum \frac{n!}{n_1!n_2! \dots n_m!} a_1^{n_1} a_2^{n_2} \dots a_m^{n_m} ,$$

donde la suma anterior incluye todas las combinaciones diferentes de los n_1, n_2, \dots, n_m tal que $\sum_{i=1}^m n_i = n$. Aquí n_i y n son enteros. Esta generalización encuentra considerables usos en Mecánica Estadística.

Las series de Maclaurin pueden aparecer algunas veces indirectamente más que el uso directo de la ecuación (8.78). Por ejemplo, la manera más conveniente para obtener la expansión en serie

$$\operatorname{sen}^{-1} x = \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} \frac{x^{2n+1}}{2n+1} = x + \frac{x^3}{6} + \frac{3x^5}{40} + \dots, \quad (8.97)$$

es hacer uso de la relación

$$\operatorname{sen}^{-1} x = \int_0^x \frac{dt}{(1-t^2)^{1/2}}.$$

Expandimos $(1-t^2)^{-1/2}$ (teorema binomial) y luego integramos término a término. Esta integración término a término es discutida en la sección 8.7. El resultado es la ecuación (8.97). Finalmente, podemos tomar el límite cuando $x \rightarrow 1$. La serie converge por la prueba de Gauss.

8.6.3. Expansión de Taylor de más de una variable.

La función f tiene más de una variable independiente, es decir, $f = f(x, y)$, la expansión de Taylor se convierte en

$$\begin{aligned} f(x, y) = & f(a, b) + (x-a) \frac{\partial f}{\partial x} + (y-b) \frac{\partial f}{\partial y} + \\ & + \frac{1}{2!} \left[(x-a)^2 \frac{\partial^2 f}{\partial x^2} + 2(x-a)(y-b) \frac{\partial^2 f}{\partial x \partial y} + (y-b)^2 \frac{\partial^2 f}{\partial y^2} \right] + \\ & + \frac{1}{3!} \left[(x-a)^3 \frac{\partial^3 f}{\partial x^3} + 3(x-a)^2(y-b) \frac{\partial^3 f}{\partial x^2 \partial y} + \right. \\ & \left. + 3(x-a)(y-b)^2 \frac{\partial^3 f}{\partial x \partial y^2} + (y-b)^3 \frac{\partial^3 f}{\partial y^3} \right] + \dots, \end{aligned} \quad (8.98)$$

con todas las derivadas evaluadas en el punto (a, b) . Usando $\alpha_j t = x_j - x_{j0}$, podemos escribir la expansión de Taylor para m variables independientes en la forma simbólica

$$f(x_j) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \left(\sum_{i=1}^m \alpha_i \frac{\partial}{\partial x_i} \right)^n f(x_k) \Big|_{x_k=x_{k0}}. \quad (8.99)$$

Una forma vectorial conveniente es

$$\psi(\vec{r} + \vec{a}) = \sum_{n=0}^{\infty} \frac{1}{n!} (\vec{a} \cdot \vec{\nabla})^n \psi(\vec{r}). \quad (8.100)$$

8.7. Series de potencias.

Las series de potencias son un tipo especial y extremadamente útil de series infinitas de la forma

$$\begin{aligned} f(x) &= a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots \\ &= \sum_{n=0}^{\infty} a_n x^n, \end{aligned} \quad (8.101)$$

donde los coeficientes a_i son constantes e independientes de x .⁴

8.7.1. Convergencia.

La ecuación (8.101) puede testearse rápidamente para la convergencia ya sea por la prueba de la raíz de Cauchy o por la prueba de la razón de D' Alembert. Si

$$\lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = R^{-1}, \quad (8.102)$$

la serie converge para $-R < x < R$. Este es el intervalo o radio de convergencia. Ya que las pruebas de la raíz y la razón fallan cuando el límite es la unidad, el punto final del intervalo requiere atención especial.

Por ejemplo, si $a_n = n^{-1}$, entonces $R = 1$ y, la serie converge para $x = -1$, pero diverge para $x = +1$. Si $a_n = n!$, entonces $R = 0$ y la serie diverge para todo $x \neq 0$.

8.8. Convergencia uniforme y absoluta.

Supongamos que nuestra serie de potencia es convergente para $-R < x < R$; entonces será uniforme y absolutamente convergente en cualquier intervalo interior, $-S \leq x \leq S$, donde $0 < S < R$. Esto podría ser probado directamente por la prueba M de Weierstrass usando $M_i = |a_i|S^i$.

8.8.1. Continuidad.

Ya que cada término $u_n(x) = a_n x^n$ es una función continua de x y $f(x) = \sum a_n x^n$ converge uniformemente para $-S \leq x \leq S$, $f(x)$ debería ser una función continua en el intervalo de convergencia uniforme. Este comportamiento es contradictorio con el comportamiento impresionantemente diferente de las series de Fourier. Las series de Fourier son usadas frecuentemente para representar funciones discontinuas tales como ondas cuadradas y ondas dientes de sierra.

8.8.2. Diferenciación e integración.

Con $u_n(x)$ continua y $\sum a_n x^n$ uniformemente convergente, encontramos que la serie diferenciada es una serie de potencia con funciones continuas y del mismo radio de convergencia que la serie original. Los nuevos factores introducidos por diferenciación (o integración) no afecta ni a la prueba de la raíz ni a la de la razón. Por lo tanto nuestra serie podría ser diferenciada o integrada tan a menudo como deseemos dentro del intervalo de convergencia uniforme.

En vista de las restricciones algo severas puestas en la diferenciación, esto es un resultado valioso y notable.

⁴La ecuación (8.101) puede ser reescrita con $z = x + iy$, reemplazando a x . Luego todos los resultados de esta sección se aplican a series complejas

8.8.3. Teorema de unicidad.

En la sección precedente, usando las series de Maclaurin, expandimos e^x y $\ln(1+x)$ en series infinitas. En los capítulos venideros las funciones son frecuentemente representadas e incluso definidas por series infinitas. Ahora estableceremos que la representación de la serie de potencias es única.

Si

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} a_n x^n, & -R_a < x < R_a \\ &= \sum_{n=0}^{\infty} b_n x^n, & -R_b < x < R_b, \end{aligned} \quad (8.103)$$

con intervalos de convergencia sobrepuestos, incluyendo el origen, luego

$$a_n = b_n, \quad (8.104)$$

para todo n ; esto es, supongamos dos representaciones de serie de potencias (diferentes) y luego procedamos a demostrar que las dos son idénticas.

De la ecuación (8.103)

$$\sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} b_n x^n, \quad -R < x < R \quad (8.105)$$

donde R es el más pequeño entre R_a, R_b . Haciendo $x = 0$ para eliminar todo salvo el término constante, obtenemos

$$a_0 = b_0. \quad (8.106)$$

Ahora, aprovechándose de la diferenciabilidad de nuestra serie de potencia, diferenciamos la ecuación (8.105), obteniendo

$$\sum_{n=1}^{\infty} n a_n x^{n-1} = \sum_{n=1}^{\infty} n b_n x^{n-1}. \quad (8.107)$$

De nuevo ajustamos $x = 0$ para aislar el nuevo término constante y encontramos

$$a_1 = b_1. \quad (8.108)$$

Repitiendo este proceso n veces, obtenemos

$$a_n = b_n, \quad (8.109)$$

lo cual muestra que las dos series coinciden. Por lo tanto nuestra representación en serie de potencia es única.

Esto será un punto crucial cuando usamos una serie de potencia para desarrollar soluciones de ecuaciones diferenciales. Esta unicidad de las series de potencia aparece frecuentemente en física teórica. La teoría de perturbaciones en Mecánica Cuántica es un ejemplo de esto. La representación en serie de potencia de funciones es a menudo útil en formas de evaluación indeterminadas, particularmente cuando la regla de l'Hospital puede ser inconveniente de aplicar.

Ejemplo

Evaluemos

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} . \quad (8.110)$$

Remplazando $\cos x$ por su expansión en serie de Maclaurin, obtenemos

$$\begin{aligned} \frac{1 - \cos x}{x^2} &= \frac{1 - (1 - x^2/2! + x^4/4! - \dots)}{x^2} \\ &= \frac{x^2/2! - x^4/4! + \dots}{x^2} \\ &= \frac{1}{2!} - \frac{x^2}{4!} + \dots . \end{aligned}$$

Tomando $x \rightarrow 0$, tenemos

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} = \frac{1}{2} . \quad (8.111)$$

La unicidad de las series de potencia significa que los coeficientes a_n pueden ser identificadas con las derivadas en una serie de Maclaurin. A partir de

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(0) x^n$$

tenemos

$$a_n = \frac{1}{n!} f^{(n)}(0) .$$

8.8.4. Inversión de series de potencia.

Supongamos que tenemos una serie

$$\begin{aligned} y - y_0 &= a_1(x - x_0) + a_2(x - x_0)^2 + \dots \\ &= \sum_{n=1}^{\infty} a_n(x - x_0)^n , \end{aligned} \quad (8.112)$$

en la cual está dada $(y - y_0)$ en términos de $(x - x_0)$. Sin embargo, podría ser deseable tener una expresión explícita para $(x - x_0)$ en términos de $(y - y_0)$. Necesitamos resolver la ecuación (8.112) para $(x - x_0)$ por inversión de nuestra serie. Supongamos que

$$x - x_0 = \sum_{n=0}^{\infty} b_n(y - y_0)^n , \quad (8.113)$$

con b_n determinado en términos de los supuestamente conocidos a_n . Una aproximación a fuerza bruta, la cual es perfectamente adecuada para los primeros coeficientes, ya que es simplemente sustituir la ecuación (8.112) en la ecuación (8.113). Igualando los coeficientes

de $(x - x_0)^n$ en ambos lados de la ecuación (8.113), ya que la serie de potencia es única, obtenemos

$$\begin{aligned} b_1 &= \frac{1}{a_1}, \\ b_2 &= \frac{a_2}{a_1^3}, \\ b_3 &= \frac{1}{a_1^5}(2a_2^2 - a_1a_3), \\ b_4 &= \frac{1}{a_1^7}(5a_1a_2a_3 - a_1^2a_4 - 5a_2^3), \quad \text{y así sucesivamente.} \end{aligned} \tag{8.114}$$

Los coeficientes mayores son listados en tablas generalmente. Una aproximación más general y mucho más elegante es desarrollada usando variables complejas.

8.9. Integrales elípticas.

Las integrales elípticas son incluidas aquí parcialmente como una ilustración del uso de las series de potencias y por su propio interés intrínseco. Este interés incluye la ocurrencia de las integrales elípticas en una gran variedad de problemas físicos.

Ejemplo Período de un péndulo simple.

Para pequeñas oscilaciones en la amplitud nuestro péndulo, figura 8.8, tiene un movimiento armónico simple con un período $T = 2\pi(l/g)^{1/2}$. Para una amplitud grande θ_M tal que $\sin \theta_M \neq \theta_M$, la segunda ley de movimiento de Newton y las ecuaciones de Lagrange conducen a una ecuación diferencial no lineal ($\sin \theta$ es una función no lineal de θ), así que necesitamos un acercamiento diferente.

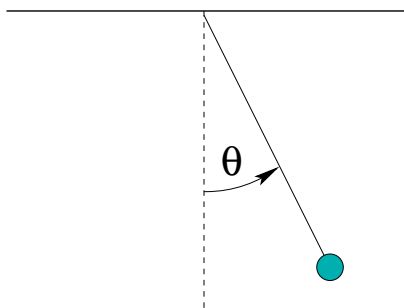


Figura 8.8: Péndulo simple.

La masa oscilante m tiene una energía cinética de $ml^2(d\theta/dt)^2/2$ y una energía potencial de $-mgl \cos \theta$ ($\theta = \pi/2$ como la elección del cero de la energía potencial). Ya que $d\theta/dt = 0$ en $\theta = \theta_M$, el principio de la conservación de la energía da

$$\frac{1}{2}ml^2 \left(\frac{d\theta}{dt} \right)^2 - mgl \cos \theta = -mgl \cos \theta_M. \tag{8.115}$$

Resolviendo para $d\theta/dt$ obtenemos

$$\frac{d\theta}{dt} = \pm \left(\frac{2g}{l}\right)^{1/2} (\cos \theta - \cos \theta_M)^{1/2} \quad (8.116)$$

con la cancelación de la masa m . Tomando t como cero cuando $\theta = 0$ y $d\theta/dt > 0$. Una integración desde $\theta = 0$ a $\theta = \theta_M$ produce

$$\int_0^{\theta_M} (\cos \theta - \cos \theta_M)^{-1/2} d\theta = \left(\frac{2g}{l}\right)^{1/2} \int_0^t dt = \left(\frac{2g}{l}\right)^{1/2} t. \quad (8.117)$$

Esto es 1/4 del ciclo, y por lo tanto el tiempo t es 1/4 del período, T . Notemos que $\theta \leq \theta_M$, trataremos la sustitución

$$\text{sen} \left(\frac{\theta}{2}\right) = \text{sen} \left(\frac{\theta_M}{2}\right) \text{sen} \varphi. \quad (8.118)$$

Con esto, la ecuación (8.117) se convierte en

$$T = 4 \left(\frac{l}{g}\right)^{1/2} \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - \text{sen}^2 \left(\frac{\theta_M}{2}\right) \text{sen}^2 \varphi}} \quad (8.119)$$

Aunque no hay un obvio mejoramiento en la ecuación (8.117), la integral ahora corresponde a la integral elíptica completa del primer tipo, $K(\text{sen} \theta_M/2)$. A partir de la expansión de serie, el período de nuestro péndulo puede ser desarrollado como una serie de potencia en $\text{sen} \theta_M/2$:

$$T = 2\pi \left(\frac{l}{g}\right)^{1/2} \left[1 + \frac{1}{4} \text{sen}^2 \frac{\theta_M}{2} + \frac{9}{64} \text{sen}^4 \frac{\theta_M}{2} + \dots \right] \quad (8.120)$$

8.9.1. Definiciones.

Generalizando el ejemplo anterior para incluir el límite superior como una variable, la *integral elíptica del primer tipo* está definida como

$$F(\varphi \setminus \alpha) = \int_0^{\varphi} \frac{d\theta}{\sqrt{1 - \text{sen}^2 \alpha \text{sen}^2 \theta}} \quad (8.121)$$

o

$$F(x \setminus m) = \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-mt^2)}}, \quad 0 \leq m < 1. \quad (8.122)$$

Para $\varphi = \pi/2$, $x = 1$, tenemos la *integral elíptica completa de primer tipo*,

$$\begin{aligned} K(m) &= \int_0^{\pi/2} \frac{d\theta}{\sqrt{1 - m \text{sen}^2 \theta}} \\ &= \int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-mt^2)}}, \end{aligned} \quad (8.123)$$

con $m = \text{sen}^2 \alpha$, $0 \leq m < 1$.

La *integral elíptica de segundo tipo* está definida por

$$E(\varphi \setminus \alpha) = \int_0^\varphi \sqrt{1 - \operatorname{sen}^2 \alpha \operatorname{sen}^2 \theta} d\theta \quad (8.124)$$

o

$$E(x \setminus m) = \int_0^x \sqrt{\frac{1 - mt^2}{1 - t^2}} dt, \quad 0 \leq m < 1 \quad (8.125)$$

Nuevamente, para el caso $\varphi = \pi/2$, $x = 1$, tenemos la *integral elíptica completa de segundo tipo*:

$$\begin{aligned} E(m) &= \int_0^{\pi/2} \sqrt{1 - m \operatorname{sen}^2 \theta} d\theta \\ &= \int_0^1 \sqrt{\frac{1 - mt^2}{1 - t^2}} dt, \quad 0 \leq m < 1. \end{aligned} \quad (8.126)$$

La figura 8.9 muestra el comportamiento de $K(m)$ y $E(m)$. Los valores de ambas funciones pueden encontrarse en tablas o evaluar en *software* como *Mathematica*.

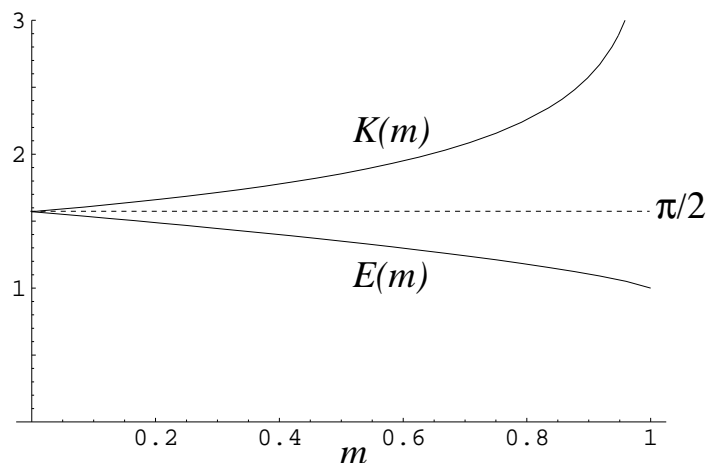


Figura 8.9: Integrales elípticas completas, $K(m)$, $E(m)$.

8.9.2. Expansión de series.

Para nuestro intervalo $0 \leq m < 1$, el denominador de $K(m)$ puede ser expandido en serie binomial

$$\begin{aligned} (1 - m \operatorname{sen}^2 \theta)^{-1/2} &= 1 + \frac{1}{2} m \operatorname{sen}^2 \theta + \frac{3}{8} m^2 \operatorname{sen}^4 \theta + \dots \\ &= \sum_{n=0}^{\infty} \frac{(2n-1)!!}{(2n)!!} m^n \operatorname{sen}^{2n} \theta. \end{aligned} \quad (8.127)$$

Para cualquier intervalo cerrado $[0, m_{\max}]$, con $m_{\max} < 1$, esta serie es uniformemente convergente y puede ser integrada término a término.

$$\int_0^{\pi/2} \text{sen}^{2n} \theta \, d\theta = \frac{(2n-1)!!}{(2n)!!} \frac{\pi}{2}. \quad (8.128)$$

De modo que

$$K(m) = \frac{\pi}{2} \left[1 + \left(\frac{1}{2}\right)^2 m + \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 m^2 + \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\right)^2 m^3 + \dots \right]. \quad (8.129)$$

Similarmente,

$$E(m) = \frac{\pi}{2} \left[1 - \left(\frac{1}{2}\right)^2 \frac{m}{1} - \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 \frac{m^2}{3} - \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}\right)^2 \frac{m^3}{5} - \dots \right]. \quad (8.130)$$

Más adelante estas series son identificadas como funciones hipergeométricas, y tenemos

$$K(m) = \frac{\pi}{2} {}_2F_1 \left(\frac{1}{2}, \frac{1}{2}, 1; m \right) \quad (8.131)$$

$$E(m) = \frac{\pi}{2} {}_2F_1 \left(-\frac{1}{2}, \frac{1}{2}, 1; m \right) \quad (8.132)$$

8.9.3. Valores límites.

De las series en las ecuaciones (8.129) y (8.130), o a partir de las integrales definidas, obtenemos

$$\lim_{m \rightarrow 0} K(m) = \frac{\pi}{2}, \quad (8.133)$$

$$\lim_{m \rightarrow 0} E(m) = \frac{\pi}{2}. \quad (8.134)$$

Para $m \rightarrow 1$, las expansiones en series no son muy útiles, A partir de la representación integral tenemos que

$$\lim_{m \rightarrow 1} K(m) = \infty, \quad (8.135)$$

diverge logarítmicamente, y por otra parte, la integral para $E(m)$ tiene un límite finito

$$\lim_{m \rightarrow 1} E(m) = 1. \quad (8.136)$$

Las integrales elípticas han sido usadas ampliamente en el pasado para evaluar integrales. Por ejemplo, integrales de la forma

$$I = \int_0^x R(t, \sqrt{a_4 t^4 + a_3 t^3 + a_2 t^2 + a_1 t + a_0}) \, dt,$$

donde R es una función racional de t y del radical, pueden ser expresadas en términos de integrales elípticas. Con los computadores actuales disponibles para una evaluación numérica rápida y directa, el interés en estas técnicas de integrales elípticas ha declinado. Sin embargo, las integrales elípticas mantienen su interés a causa de su apariencia en problemas en Física.

8.10. Números de Bernoulli.

Los números de Bernoulli fueron introducidos por Jacques Bernoulli. Hay muchas definiciones equivalentes, pero debe tenerse extremo cuidado, porque algunos autores introducen variaciones en la numeración o en signo. Un acercamiento relativamente simple para definir los números de Bernoulli es por la serie⁵

$$\frac{x}{e^x - 1} = \sum_{n=0}^{\infty} \frac{B_n x^n}{n!}, \quad (8.137)$$

la cual converge para $|x| < 2\pi$ usando el test del cociente. Diferenciando esta serie de potencia repetidamente y luego evaluando para $x = 0$, obtenemos

$$B_n = \left[\frac{d^n}{dx^n} \left(\frac{x}{e^x - 1} \right) \right]_{x=0}. \quad (8.138)$$

Específicamente,

$$B_1 = \left. \frac{d}{dx} \left(\frac{x}{e^x - 1} \right) \right|_{x=0} = \frac{1}{e^x - 1} - \frac{x e^x}{(e^x - 1)^2} \Big|_{x=0} = -\frac{1}{2}, \quad (8.139)$$

como puede ser visto por la expansión en series de los denominadores. Usando $B_0 = 1$ y $B_1 = -1/2$, es fácil verificar que la función

$$\frac{x}{e^x - 1} - 1 + \frac{x}{2} = \sum_{n=2}^{\infty} \frac{B_n x^n}{n!} = -\frac{x}{e^{-x} - 1} - 1 - \frac{x}{2}, \quad (8.140)$$

es par en x , tal que todos los $B_{2n+1} = 0$.

Para derivar una relación de recurrencia para los números de Bernoulli, multiplicamos

$$\begin{aligned} \frac{e^x - 1}{x} \frac{x}{e^x - 1} = 1 &= \left[\sum_{m=0}^{\infty} \frac{x^m}{(m+1)!} \right] \left[1 - \frac{x}{2} + \sum_{n=1}^{\infty} \frac{B_{2n} x^{2n}}{(2n)!} \right] \\ &= 1 + \sum_{m=1}^{\infty} x^m \left[\frac{1}{(m+1)!} - \frac{1}{2m!} \right] + \sum_{N=2}^{\infty} x^N \sum_{1 \leq n \leq N/2} \frac{B_{2n}}{[(2n)!(N-2n+1)!]}. \end{aligned} \quad (8.141)$$

La ecuación (8.141) produce

$$\frac{1}{2}(N+1) - 1 = \sum_{1 \leq n \leq N/2} B_{2n} \binom{N+1}{2n} = \frac{1}{2}(N-1), \quad (8.142)$$

la cual es equivalente a

$$\begin{aligned} N - \frac{1}{2} &= \sum_{n=1}^N B_{2n} \binom{2N+1}{2n}, \\ N - 1 &= \sum_{n=1}^{N-1} B_{2n} \binom{2N}{2n}. \end{aligned} \quad (8.143)$$

⁵La función $\frac{x}{e^x - 1}$ puede ser considerada una *función generatriz* ya que genera los números de Bernoulli.

n	B_n	B_n
0	1	1.0000 00000
1	$-\frac{1}{2}$	-0.5000 00000
2	$\frac{1}{6}$	0.1666 66667
3	$-\frac{1}{30}$	-0.0333 33333
4	$\frac{1}{42}$	0.0238 09524
5	$-\frac{1}{30}$	-0.0333 33333
6	$\frac{5}{66}$	0.0757 57576

Cuadro 8.1: Números de Bernoulli

A partir de la ecuación (8.143) los números de Bernoulli en la tabla 8.1 se obtienen rápidamente. Si la variable x en la ecuación (8.137) es remplazada por $2xi$ (y B_1 elegido igual a $-1/2$), obtenemos una definición alternativa (y equivalente) de B_{2n} , la expresión

$$x \cot x = \sum_{n=0}^{\infty} (-1)^n B_{2n} \frac{(2x)^{2n}}{(2n)!}, \quad -\pi < x < \pi. \quad (8.144)$$

Usando el método del residuo o trabajando a partir de la representación de producto infinito de $\sin(x)$, encontramos que

$$B_{2n} = \frac{(-1)^{n-1} 2(2n)!}{(2\pi)^{2n}} \sum_{p=1}^{\infty} \frac{1}{p^{2n}}, \quad n = 1, 2, 3, \dots \quad (8.145)$$

Esta representación de los números de Bernoulli fue descubierta por Euler. Es fácil ver a partir de la ecuación (8.145) que $|B_{2n}|$ aumenta sin límite cuando $n \rightarrow \infty$. Ilustrando el comportamiento divergente de los números de Bernoulli, tenemos

$$\begin{aligned} B_{20} &= -5.291 \times 10^2 \\ B_{200} &= -3.647 \times 10^{215}. \end{aligned}$$

Algunos autores prefieren definir los números de Bernoulli con una versión modificada de la ecuación (8.145) usando

$$B_{2n} = \frac{2(2n)!}{(2\pi)^{2n}} \sum_{p=1}^{\infty} \frac{1}{p^{2n}}, \quad (8.146)$$

el subíndice es justo la mitad de nuestro subíndice original y todos los signos son positivos. Nuevamente, se debe chequear cuidadosamente la definición que se está usando de los números de Bernoulli.

Los números de Bernoulli aparecen frecuentemente en teoría de números. El teorema de von Standt-Clausen establece que

$$B_{2n} = A_n - \frac{1}{p_1} - \frac{1}{p_2} - \frac{1}{p_3} - \dots - \frac{1}{p_k}, \quad (8.147)$$

en el cual A_n es un entero y p_1, p_2, \dots, p_k son números primos tal que $p_i - 1$ es un divisor de $2n$. Podemos fácilmente verificar que esto se satisface para

$$\begin{aligned} B_6(A_3 = 1, p = 2, 3, 7) , \\ B_8(A_4 = 1, p = 2, 3, 5) , \\ B_{10}(A_5 = 1, p = 2, 3, 11) , \end{aligned} \quad (8.148)$$

y otros casos especiales.

Los números de Bernoulli aparecen en la suma de potencias enteras de enteros,

$$\sum_{j=1}^N j^p , \quad p \text{ entero.}$$

y en numerosas expansiones de series de las funciones trascendentales, incluyendo $\tan x$, $\cot x$, $\operatorname{sen}^{-1} x$, $\ln |\operatorname{sen} x|$, $\ln |\cos x|$, $\ln |\tan x|$, $\tanh x$, $\operatorname{coth} x$ y $\operatorname{cosh}^{-1} x$. Por ejemplo,

$$\tan(x) = x + \frac{x^3}{3} + \frac{2}{15}x^5 + \dots + \frac{(-1)^{n-1}2^{2n}(2^{2n} - 1)B_{2n}}{(2n)!}x^{2n-1} + \dots . \quad (8.149)$$

Los números de Bernoulli probablemente vengan en tales expansiones en series a causa de las ecuaciones de definición (8.137) y (8.143) y de su relación con la función zeta de Riemann

$$\zeta(2n) = \sum_{p=1}^{\infty} \frac{1}{p^{2n}} . \quad (8.150)$$

8.10.1. Funciones de Bernoulli.

Si la ecuación (8.137) puede ser fácilmente generalizada, tenemos

$$\frac{xe^{xs}}{e^x - 1} = \sum_{n=0}^{\infty} B_n(s) \frac{x^n}{n!} . \quad (8.151)$$

definiendo las *funciones de Bernoulli*, $B_n(s)$. Las primeras siete funciones de Bernoulli están dadas en la tabla 8.2.

De la función generadora, ecuación (8.151),

$$B_n(0) = B_n , \quad n = 1, 2, \dots . \quad (8.152)$$

la función de Bernoulli evaluadas en cero es igual al correspondiente número de Bernoulli. Dos propiedades particularmente importantes de las funciones de Bernoulli se deducen a partir de la definición: una relación de diferenciación

$$B'_n(s) = nB_{n-1}(s) , \quad n = 1, 2, \dots . \quad (8.153)$$

y una relación de simetría

$$B_n(1) = (-1)^n B_n(0) , \quad n = 1, 2, \dots . \quad (8.154)$$

Estas relaciones son usadas en el desarrollo de la fórmula de integración de Euler-Maclaurin.

$$\begin{aligned}
B_0 &= 1 \\
B_1 &= x - \frac{1}{2} \\
B_2 &= x^2 - x + \frac{1}{6} \\
B_3 &= x^3 - \frac{3}{2}x^2 + \frac{1}{2}x \\
B_4 &= x^4 - 2x^3 + x^2 - \frac{1}{30} \\
B_5 &= x^5 - \frac{5}{2}x^4 + \frac{5}{3}x^2 - \frac{1}{6}x \\
B_6 &= x^6 - 3x^5 + \frac{5}{2}x^4 - \frac{1}{2}x^2 + \frac{1}{42}
\end{aligned}$$

Cuadro 8.2: Funciones de Bernoulli

8.10.2. Fórmula de integración de Euler-Maclaurin.

Uno de los usos de las funciones de Bernoulli es la derivación de la fórmula de integración de Euler-Maclaurin. Esta fórmula es usada en el desarrollo de una expresión asintótica para la función factorial, serie de Stirling. La técnica es integración por partes repetida, usando la ecuación (8.153) para crear nuevas derivadas. Comenzamos con

$$\int_0^1 f(x) dx = \int_0^1 f(x)B_0(x) dx . \quad (8.155)$$

A partir de la ecuación (8.153)

$$B_1'(x) = B_0(x) = 1 . \quad (8.156)$$

Sustituyendo $B_1'(x)$ en la ecuación (8.155) e integrando por partes, obtenemos

$$\begin{aligned}
\int_0^1 f(x) dx &= f(1)B_1(1) - f(0)B_1(0) - \int_0^1 f'(x)B_1(x) dx \\
&= \frac{1}{2}[f(1) - f(0)] - \int_0^1 f'(x)B_1(x) dx
\end{aligned} \quad (8.157)$$

Nuevamente, usando la ecuación (8.153), tenemos

$$B_1(x) = \frac{1}{2}B_2'(x) , \quad (8.158)$$

e integrando por partes

$$\begin{aligned}
\int_0^1 f(x) dx &= \frac{1}{2}[f(1) - f(0)] - \frac{1}{2!}[f'(1)B_2(1) - f'(0)B_2(0)] + \\
&\quad \frac{1}{2!} \int_0^1 f^{(2)}(x)B_2(x) dx .
\end{aligned} \quad (8.159)$$

Usando las relaciones,

$$\begin{aligned}
B_{2n}(1) &= B_{2n}(0) = B_{2n} , & n &= 0, 1, 2, \dots \\
B_{2n+1}(1) &= B_{2n+1}(0) = 0 , & n &= 1, 2, 3, \dots ,
\end{aligned} \quad (8.160)$$

y continuando este proceso, tenemos

$$\int_0^1 f(x) dx = \frac{1}{2}[f(1) - f(0)] - \sum_{p=1}^q \frac{1}{(2p)!} B_{2p} [f^{(2p-1)}(1) - f^{(2p-1)}(0)] + \frac{1}{(2q)!} \int_0^1 f^{(2q)}(x) B_{2q}(x) dx . \quad (8.161)$$

Esta es la fórmula de integración de Euler-Maclaurin. Supone que la función $f(x)$ tiene todas las derivadas requeridas.

El intervalo de integración en la ecuación (8.161) puede ser trasladado de $[0, 1]$ a $[1, 2]$ reemplazando $f(x)$ por $f(x + 1)$. Sumando tales resultados hasta $[n - 1, n]$,

$$\int_0^n f(x) dx = \frac{1}{2}f(0) + f(1) + f(2) + \cdots + f(n-1) + \frac{1}{2}f(n) + \sum_{p=1}^q \frac{1}{(2p)!} B_{2p} [f^{(2p-1)}(n) - f^{(2p-1)}(0)] + \frac{1}{(2q)!} \int_0^1 B_{2q}(x) \sum_{\nu=0}^{n-1} f^{(2q)}(x + \nu) dx . \quad (8.162)$$

Los términos $\frac{1}{2}f(0) + f(1) + \cdots + \frac{1}{2}f(n)$ aparecen exactamente como una integración o cuadratura trapezoidal. La suma sobre p puede ser interpretada como una corrección a la aproximación trapezoidal. La ecuación (8.162) es la forma usada en la derivación de la fórmula de Stirling.

La fórmula de Euler-Maclaurin es a menudo útil para sumar series al convertirlas en integrales.

8.11. Función zeta de Riemann.

Estas series $\sum_{p=1}^{\infty} p^{-2n}$ fueron usadas como series de comparación para probar la convergencia y en la ecuación (8.144) como una definición de los números de Bernoulli, B_{2n} . También sirve para definir la función zeta de Riemann por

$$\zeta(s) \equiv \sum_{n=1}^{\infty} \frac{1}{n^s} , \quad s > 1 . \quad (8.163)$$

La tabla 8.3 muestra los valores de $\zeta(s)$ para s entero, $s = 2, 3, \dots, 10$. La figura 8.10 es un gráfico de $\zeta(s) - 1$. Una expresión integral para esta función zeta de Riemann aparecerá como parte del desarrollo de la función gama.

Otra interesante expresión para la función zeta puede ser derivada como

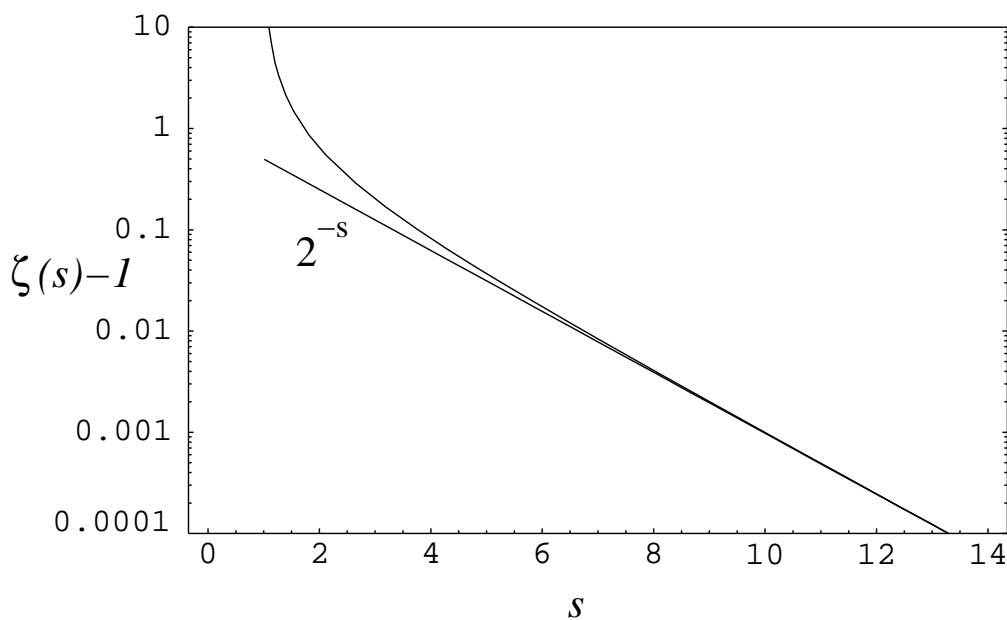
$$\zeta(s)(1 - 2^{-s}) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots - \left(\frac{1}{2^s} + \frac{1}{4^s} + \frac{1}{6^s} + \cdots \right) \quad (8.164)$$

eliminando todos los n^{-s} , donde n es un múltiplo de 2. Entonces

$$\zeta(s)(1 - 2^{-s})(1 - 3^{-s}) = 1 + \frac{1}{3^s} + \frac{1}{5^s} + \frac{1}{7^s} + \frac{1}{9^s} + \cdots - \left(\frac{1}{3^s} + \frac{1}{9^s} + \frac{1}{15^s} + \cdots \right) , \quad (8.165)$$

s	$\zeta(s)$
2	1.64493 40668
3	1.20205 69032
4	1.08232 32337
5	1.03692 77551
6	1.01734 30620
7	1.00834 92774
8	1.00407 73562
9	1.00200 83928
10	1.00099 45751

Cuadro 8.3: Función zeta de Riemann.

Figura 8.10: Función zeta de Riemann, $\zeta(s) - 1$, versus s .

eliminando todos los términos remanentes, donde n es un múltiplo de 3. Continuando, tenemos $\zeta(s)(1 - 2^{-s})(1 - 3^{-s})(1 - 5^{-s}) \dots (1 - P^{-s})$, donde P es un número primo, y todos los términos n^{-s} , en el cual n es un múltiplo entero por sobre P , son cancelados. Para $P \rightarrow \infty$,

$$\zeta(s)(1 - 2^{-s})(1 - 3^{-s}) \dots (1 - P^{-s}) = \zeta(s) \prod_{P(\text{primo})=2}^{\infty} (1 - P^{-s}) = 1. \quad (8.166)$$

Por lo tanto

$$\zeta(s) = \left[\prod_{P(\text{primo})=2}^{\infty} (1 - P^{-s}) \right]^{-1} \quad (8.167)$$

dando $\zeta(s)$ como un producto infinito.⁶

Este procedimiento de cancelación tiene una clara aplicación en el cálculo numérico. La ecuación (8.164) dará $\zeta(s)(1 - 2^{-s})$ con la misma precisión como la ecuación (8.163) da $\zeta(s)$, pero solamente con la mitad de términos. (En cuyo caso, podría hacerse una corrección para despreciar la cola de la serie por la técnica de Maclaurin reemplazando la serie por una integral).

Conjuntamente con la función zeta de Riemann, habitualmente se definen otras tres funciones de sumas de potencia recíprocas:

$$\eta(s) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^s} = (1 - 2^{1-s})\zeta(s) ,$$

$$\lambda(s) = \sum_{n=0}^{\infty} \frac{1}{(2n+1)^s} = \left(1 - \frac{1}{2^s}\right) \zeta(s) ,$$

y

$$\beta(s) = \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n+1)^s} .$$

A partir de los números de Bernoulli o de las series de Fourier podemos determinar algunos valores especiales

$$\zeta(2) = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots = \frac{\pi^2}{6}$$

$$\zeta(4) = 1 + \frac{1}{2^4} + \frac{1}{3^4} + \cdots = \frac{\pi^4}{90}$$

$$\eta(2) = 1 - \frac{1}{2^2} + \frac{1}{3^2} - \cdots = \frac{\pi^2}{12}$$

$$\eta(4) = 1 - \frac{1}{2^4} + \frac{1}{3^4} - \cdots = \frac{7\pi^4}{720}$$

$$\lambda(2) = 1 + \frac{1}{3^2} + \frac{1}{5^2} + \cdots = \frac{\pi^2}{8}$$

$$\lambda(4) = 1 + \frac{1}{3^4} + \frac{1}{5^4} + \cdots = \frac{\pi^4}{96}$$

$$\beta(1) = 1 - \frac{1}{3} + \frac{1}{5} - \cdots = \frac{\pi}{4}$$

$$\beta(3) = 1 - \frac{1}{3^3} + \frac{1}{5^3} - \cdots = \frac{\pi^3}{32}$$

La constante de Catalán

$$\beta(2) = 1 - \frac{1}{3^2} + \frac{1}{5^2} - \cdots = 0.9159\ 6559 \dots ,$$

⁶Este es el punto de partida para la vasta aplicación de la función zeta de Riemann a la teoría de números.

8.11.1. Mejoramiento de la convergencia.

Si requerimos sumar una serie convergente $\sum_{n=1}^{\infty} a_n$ cuyos términos son funciones racionales de n , la convergencia puede ser mejorada dramáticamente introduciendo la función zeta de Riemann.

Ejemplo Mejorando la convergencia.

El problema es evaluar la serie $\sum_{n=1}^{\infty} \frac{1}{(1+n^2)}$. Expandiendo $\frac{1}{(1+n^2)} = \frac{1}{n^2} \frac{1}{\left(1 + \frac{1}{n^2}\right)}$ por

división directa, tenemos

$$\begin{aligned} \frac{1}{1+n^2} &= \frac{1}{n^2} \left(1 - \frac{1}{n^2} + \frac{1}{n^4} - \frac{n^{-6}}{1+n^{-2}} \right) \\ &= \frac{1}{n^2} - \frac{1}{n^4} + \frac{1}{n^6} - \frac{1}{n^8 + n^6} . \end{aligned}$$

Por lo tanto

$$\sum_{n=1}^{\infty} \frac{1}{1+n^2} = \zeta(2) - \zeta(4) + \zeta(6) - \sum_{n=1}^{\infty} \frac{1}{n^8 + n^6} .$$

Las funciones ζ son conocidas y el remanente de la series converge como n^{-6} . Claramente, el proceso puede ser continuado hasta cuando uno desee. Usted puede hacer una elección entre cuánta álgebra hará y cuánta aritmética hará el computador.

Otros métodos para mejorar la efectividad computacional están dadas al final de la sección [8.2](#) y [8.4](#).

8.12. Series asintóticas o semi-convergentes.

Las series asintóticas aparecen frecuentemente en Física. En cálculo numérico ellas son empleadas para el cálculo de una variedad de funciones. Consideremos aquí dos tipos de integrales que conducen a series asintóticas: primero, una integral de la forma

$$I_1(x) = \int_x^{\infty} e^{-u} f(u) du ,$$

donde la variable x aparece como el límite inferior de una integral. Segundo, consideremos la forma

$$I_2(x) = \int_0^{\infty} e^{-u} f\left(\frac{u}{x}\right) du ,$$

con la función f expandible en serie de Taylor. Las series asintóticas a menudo ocurren como solución de ecuaciones diferenciales. Un ejemplo de este tipo de series aparece como una de las soluciones de la ecuación de Bessel.

8.12.1. Función gama incompleta.

La naturaleza de una serie asintótica es quizás mejor ilustrada por un ejemplo específico. Supongamos que tenemos una función integral exponencial⁷

$$Ei(x) = \int_{-\infty}^x \frac{e^u}{u} du, \quad (8.168)$$

o

$$-Ei(-x) = \int_x^{\infty} \frac{e^{-u}}{u} du = E_1(x), \quad (8.169)$$

para ser evaluada para grandes valores de x . Mejor todavía, tomemos una generalización de la función factorial incompleta (función gama incompleta),

$$I(x, p) = \int_x^{\infty} e^{-u} u^{-p} du = \Gamma(1 - p, x), \quad (8.170)$$

en la cual x y p son positivas. De nuevo, buscamos evaluarla para valores grandes de x .

Integrando por partes, obtenemos

$$I(x, p) = \frac{e^{-x}}{x^p} - p \int_x^{\infty} e^{-u} u^{-p-1} du = \frac{e^{-x}}{x^p} - \frac{pe^{-x}}{x^{p+1}} + p(p+1) \int_x^{\infty} e^{-u} u^{-p-2} du \quad (8.171)$$

Continuando para integrar por partes, desarrollamos la serie

$$I(x, p) = e^{-x} \left(\frac{1}{x^p} - \frac{p}{x^{p+1}} + \frac{p(p+1)}{x^{p+2}} - \dots (-1)^{n-1} \frac{(p+n-2)!}{(p-1)!x^{p+n-1}} \right) + (-1)^n \frac{(p+n-1)!}{(p-1)!} \int_x^{\infty} e^{-u} u^{-p-n} du. \quad (8.172)$$

Esta es una serie notable. Chequeando la convergencia por la prueba de D' Alembert, encontramos

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{|u_{n+1}|}{|u_n|} &= \lim_{n \rightarrow \infty} \frac{(p+n)!}{(p+n-1)!} \frac{1}{x} \\ &= \lim_{n \rightarrow \infty} \frac{(p+n)}{x} \\ &= \infty \end{aligned} \quad (8.173)$$

para todos los valores finitos de x . Por lo tanto nuestras series son series infinitas que divergen en todas partes!. Antes de descartar la ecuación (8.172) como inútil, veamos cuan bien una suma parcial dada se aproxima a la función factorial incompleta, $I(x, p)$.

$$= (-1)^{n+1} \frac{(p+n)!}{(p-1)!} \int_x^{\infty} e^{-u} u^{-p-n-1} du = R_n(x, p). \quad (8.174)$$

⁷Esta función ocurre con frecuencia en problemas astrofísicos que involucran gases con una distribución de energía de Maxwell-Boltzmann.

En valor absoluto

$$|I(x, p) - s_n(x, p)| \leq \frac{(p+n)!}{(p-1)!} \int_x^\infty e^{-u} u^{-p-n-1} du .$$

Luego sustituimos $u = v + x$ la integral se convierte en

$$\begin{aligned} \int_x^\infty e^{-u} u^{-p-n-1} du &= e^{-x} \int_0^\infty e^{-v} (v+x)^{-p-n-1} dv \\ &= \frac{e^{-x}}{x^{p+n+1}} \int_0^\infty e^{-v} \left(1 + \frac{v}{x}\right)^{-p-n-1} dv . \end{aligned}$$

Para x grande la integral final se aproxima a 1 y

$$|I(x, p) - s_n(x, p)| \approx \frac{(p+n)!}{(p-1)!} \frac{e^{-x}}{x^{p+n+1}} . \tag{8.175}$$

Esto significa que si tomamos un x suficientemente grande, nuestra suma parcial s_n es arbitrariamente una buena aproximación a la función deseada $I(x, p)$. Nuestra serie divergente, por lo tanto, es perfectamente buena para cálculos de sumas parciales. Por esta razón algunas veces es llamada *serie semi-convergente*. Notemos que la potencia de x en el denominador del remanente $(p+n+1)$ es más alto que la potencia de x en último término incluido en $s_n(x, p)$, $(p+n)$.

Ya que el remanente $R_n(x, p)$ alterna en signo, las sucesivas sumas parciales dan alternadamente cotas superiores e inferiores para $I(x, p)$. El comportamiento de la serie (con $p = 1$) como una función del número de términos incluidos es mostrado en la figura 8.11. Tenemos

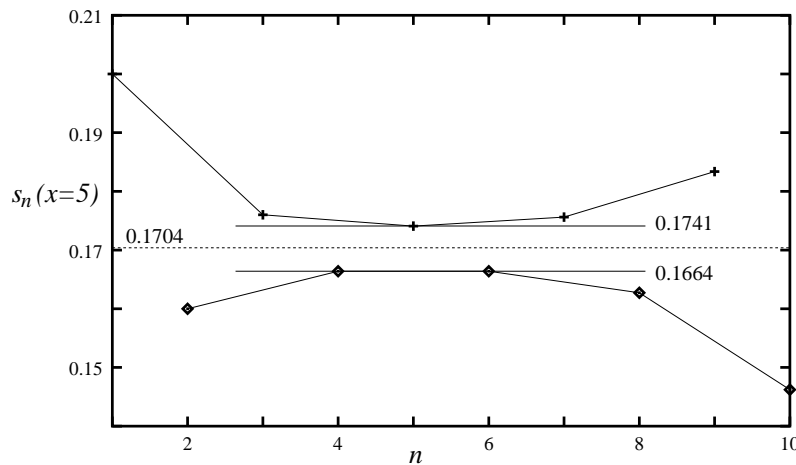


Figura 8.11: Sumas parciales de $e^x E_1(x) \Big|_{x=5}$.

$$\begin{aligned} e^x E_1(x) &= e^x \int_x^\infty \frac{e^{-u}}{u} du \\ &= s_n(x) \approx \frac{1}{x} - \frac{1!}{x^2} + \frac{2!}{x^3} - \frac{3!}{x^4} + \dots + (-1)^n \frac{n!}{x^{n+1}} , \end{aligned} \tag{8.176}$$

la cual es evaluada en $x = 5$. Para un valor dado de x las sucesivas cotas superiores e inferiores dadas por las sumas parciales primero convergen y luego divergen. La determinación óptima de $e^x E_1(x)$ está dada por la aproximación más cercana de las cotas superiores e inferiores, esto es, entre $s_4 = s_6 = 0.1664$ y $s_5 = 0.1741$ para $x = 5$. Por lo tanto

$$0.1664 \leq e^x E_1(x) \Big|_{x=5} \leq 0.1741 . \quad (8.177)$$

Realmente, a partir de las tablas,

$$e^x E_1(x) \Big|_{x=5} = 0.1704 , \quad (8.178)$$

dentro de los límites establecidos por nuestra expansión asintótica. Note cuidadosamente que la inclusión de términos adicionales en la serie de expansión más allá del punto óptimo, literalmente reduce la precisión de la representación.

Cuando aumentamos x , la diferencia entre la cota superior más baja y la cota inferior más alta disminuirá. Tomando x suficientemente grande, uno podría calcular $e^x E_1(x)$ para cualquier grado de precisión deseado.

8.12.2. Integrales coseno y seno.

Las series asintóticas también pueden ser desarrolladas a partir de integrales definidas si el integrando tiene el comportamiento requerido. Como un ejemplo, las integrales seno y coseno están definidas por

$$Ci(x) = - \int_x^\infty \frac{\cos t}{t} dt , \quad (8.179)$$

$$si(x) = - \int_x^\infty \frac{\sen t}{t} dt , \quad (8.180)$$

Combinando éstas con funciones trigonométricas regulares, podemos definir

$$\begin{aligned} f(x) &= Ci(x) \sen(x) - si(x) \cos(x) = \int_0^\infty \frac{\sen(x)}{y+x} dy \\ g(x) &= -Ci(x) \cos(x) - si(x) \sin(x) = \int_0^\infty \frac{\cos(x)}{y+x} dy \end{aligned} \quad (8.181)$$

con la nueva variable $y = t - x$. Llevando a variable compleja, tenemos

$$\begin{aligned} g(x) + if(x) &= \int_0^\infty \frac{e^{iy}}{y+x} dy \\ &= \int_0^\infty \frac{ie^{-xu}}{1+iu} du \end{aligned} \quad (8.182)$$

en el cual $u = -iy/x$. Los límites de integración, 0 a ∞ , a más que de 0 a $-i\infty$, puede ser justificado por el teorema de Cauchy. Racionalizando el denominador e igualando la parte

real y la parte imaginaria, obtenemos

$$\begin{aligned} g(x) &= \int_0^\infty \frac{ue^{-xu}}{1+u^2} du, \\ f(x) &= \int_0^\infty \frac{e^{-xu}}{1+u^2} du. \end{aligned} \quad (8.183)$$

La convergencia de las integrales requiere que $\operatorname{Re}(x) > 0$.⁸

Ahora, desarrollamos la expansión asintótica, consideremos el cambio de variable $v = xu$ y expandimos el factor $[1 + (v/x)^2]^{-1}$ por el teorema del binomio. Tenemos

$$\begin{aligned} f(x) &\approx \frac{1}{x} \int_0^\infty e^{-v} \sum_{0 \leq n \leq N} (-1)^n \frac{v^{2n}}{x^{2n}} dv = \frac{1}{x} \sum_{0 \leq n \leq N} (-1)^n \frac{(2n)!}{x^{2n}} \\ g(x) &\approx \frac{1}{x^2} \int_0^\infty e^{-v} \sum_{0 \leq n \leq N} (-1)^n \frac{v^{2n+1}}{x^{2n}} dv = \frac{1}{x^2} \sum_{0 \leq n \leq N} (-1)^n \frac{(2n+1)!}{x^{2n}}. \end{aligned} \quad (8.184)$$

De las ecuaciones (8.181) y (8.184)

$$\begin{aligned} Ci(x) &\approx \frac{\operatorname{sen}(x)}{x} \sum_{0 \leq n \leq N} (-1)^n \frac{(2n)!}{x^{2n}} - \frac{\cos(x)}{x^2} \sum_{0 \leq n \leq N} (-1)^n \frac{(2n+1)!}{x^{2n}} \\ si(x) &\approx -\frac{\cos(x)}{x} \sum_{0 \leq n \leq N} (-1)^n \frac{(2n)!}{x^{2n}} - \frac{\operatorname{sen}(x)}{x^2} \sum_{0 \leq n \leq N} (-1)^n \frac{(2n+1)!}{x^{2n}}, \end{aligned} \quad (8.185)$$

las expansiones asintóticas deseadas.

La técnica de expandir el integrando de una integral definida e integrar término a término lo volveremos a aplicar para desarrollar una expansión asintótica de la función de Bessel modificada K_ν y también para las expansiones de las dos funciones hipergeométricas confluentes $M(a, c; x)$ y $U(a, c; x)$.

8.12.3. Definición de series asintóticas.

El comportamiento de estas series (ecuaciones (8.172) y (8.185)) en consistencia con las propiedades definidas para una serie asintótica⁹. Siguiendo a Poincaré, tomamos

$$x^n R_n(x) = x^n [f(x) - s_n(x)], \quad (8.186)$$

donde

$$s_n(x) = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \cdots + \frac{a_n}{x^n}. \quad (8.187)$$

La expansión asintótica de $f(x)$ tiene las propiedades que

$$\lim_{x \rightarrow \infty} x^n R_n(x) = 0, \quad \text{para } n \text{ fijo}, \quad (8.188)$$

⁸La parte real.

⁹No es necesario que las series asintóticas sean series de potencia.

y

$$\lim_{n \rightarrow \infty} x^n R_n(x) = \infty, \quad \text{para } x \text{ fijo,} \quad (8.189)$$

Vemos la ecuaciones (8.172) y (8.173) como un ejemplo de estas propiedades. Para series de potencias, como las supuestas en la forma de $s_n(x)$, $R_n(x) \sim x^{-n-1}$. Con condiciones (8.188) y (8.189) satisfechas, escribimos

$$f(x) \approx \sum_{n=0}^{\infty} a_n \frac{1}{x^n}. \quad (8.190)$$

Notemos el uso de \approx en lugar de $=$. La función $f(x)$ es igual a la serie solamente en el límite cuando $x \rightarrow \infty$.

Las expansiones asintóticas de dos funciones pueden ser multiplicadas entre sí y el resultado será una expansión asintótica de un producto de dos funciones.

La expansión asintótica de una función dada $f(t)$ puede ser integrada término a término (justo como en una serie uniformemente convergente de una función continua) a partir de $x \leq t < \infty$ y el resultado será una expansión asintótica de $\int_x^{\infty} f(t) dt$. Una diferenciación término a término, sin embargo, es válida solamente bajo condiciones muy especiales.

Algunas funciones no poseen una expansión asintótica; e^x es un ejemplo de tales funciones. Sin embargo, si una función tiene una expansión asintótica, tiene solamente una. La correspondencia no es uno a uno; muchas funciones pueden tener la misma expansión asintótica.

Uno de los métodos más poderoso y útil de generar expansiones asintóticas, es el método de *steepest descents*, será desarrollado más adelante. Las aplicaciones incluyen la derivación de la fórmula de Stirling para la función factorial (completa) y las formas asintóticas de las varias funciones de Bessel.

8.12.4. Aplicaciones a cálculo numérico.

Las series asintóticas son usadas frecuentemente en el cálculo de funciones por los computadores. Este es el caso de las funciones de Neumann $N_0(x)$ y $N_1(x)$, y las funciones modificadas de Bessel $I_n(x)$ y $K_n(x)$. Las series asintóticas para integrales del tipo exponencial, ecuación (8.176), para las integrales de Fresnel, y para la función de error de Gauss, son usadas para la evaluación de estas integrales para valores grandes del argumento. Cuán grande debería ser el argumento depende de la precisión requerida.

8.13. Productos infinitos.

Consideremos una sucesión de factores positivos $f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdots f_n (f_i > 0)$. Usando π mayúscula para indicar el producto, tenemos

$$f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdots f_n = \prod_{i=1}^n f_i. \quad (8.191)$$

Definimos p_n , como el producto parcial, en analogía con s_n la suma parcial,

$$p_n = \prod_{i=1}^n f_i, \quad (8.192)$$

y entonces investigamos el límite

$$\lim_{n \rightarrow \infty} p_n = P . \quad (8.193)$$

Si P es finito (pero no cero), decimos que el producto infinito es convergente. Si P es infinito o cero, el producto infinito es etiquetado como divergente.

Ya que el producto divergerá a infinito si

$$\lim_{n \rightarrow \infty} f_n > 1 \quad (8.194)$$

o a cero para

$$0 < \lim_{n \rightarrow \infty} f_n < 1 , \quad (8.195)$$

es conveniente escribir nuestro producto como

$$\prod_{n=1}^{\infty} (1 + a_n) .$$

La condición $a_n \rightarrow 0$ es entonces una condición necesaria (pero no suficiente) para la convergencia.

El producto infinito puede ser relacionado a una serie infinita por el método obvio de tomar el logaritmo

$$\ln \prod_{n=1}^{\infty} (1 + a_n) = \sum_{n=1}^{\infty} \ln(1 + a_n) . \quad (8.196)$$

Una relación más útil es probada por el siguiente teorema.

8.13.1. Convergencia de un producto infinito.

Si $0 \leq a_n < 1$, el producto infinito $\prod_{n=1}^{\infty} (1 + a_n)$ y $\prod_{n=1}^{\infty} (1 - a_n)$ converge si $\sum_{n=1}^{\infty} a_n$ converge y diverge si $\sum_{n=1}^{\infty} a_n$ diverge.

Considerando el término $1 + a_n$, vemos que de la ecuación (8.80)

$$1 + a_n \leq e^{a_n} . \quad (8.197)$$

Por lo tanto el producto parcial p_n

$$p_n \leq e^{s_n} , \quad (8.198)$$

y haciendo $n \rightarrow \infty$,

$$\prod_{n=1}^{\infty} (1 + a_n) \leq \exp \sum_{n=1}^{\infty} a_n . \quad (8.199)$$

estableciendo una cota superior para el producto infinito.

Para desarrollar una cota más baja, notemos que

$$p_n = 1 + \sum_{i=1}^n a_i + \sum_{i=1}^n \sum_{j=1}^n a_i a_j + \cdots > s_n , \quad (8.200)$$

ya que $a_i \geq 0$. De modo que

$$\prod_{n=1}^{\infty} (1 + a_n) \geq \sum_{n=1}^{\infty} a_n . \quad (8.201)$$

Si la suma infinita permanece finita, el producto infinito también lo hará. Si la suma infinita diverge, también lo hará el producto infinito.

El caso de $\prod(1 - a_n)$ es complicado por el signo negativo, pero una prueba de que depende de la prueba anterior puede ser desarrollada notando que para $a_n < 1/2$ (recuerde que $a_n \rightarrow 0$ para convergencia)

$$(1 - a_n) \leq \frac{1}{1 + a_n}$$

y

$$(1 - a_n) \geq \frac{1}{1 + 2a_n} . \quad (8.202)$$

8.13.2. Funciones seno, coseno y gama.

El lector reconocerá que un polinomio de orden n $P_n(x)$ con n raíces reales puede ser escrito como un producto de n factores:

$$P_n(x) = (x - x_1)(x - x_2) \cdots (x - x_n) = \prod_{i=1}^n (x - x_i) . \quad (8.203)$$

De la misma manera podemos esperar que una función con un número infinito de raíces pueda ser escrito como un producto infinito, un factor para cada raíz. Esto es por cierto el caso de las funciones trigonométricas. Tenemos dos representaciones muy útiles en productos infinitos,

$$\text{sen}(x) = x \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{n^2\pi^2} \right) , \quad (8.204)$$

$$\text{cos}(x) = \prod_{n=1}^{\infty} \left[1 - \frac{4x^2}{(2n-1)^2\pi^2} \right] . \quad (8.205)$$

La más conveniente y quizás la más elegante derivación de estas dos expresiones es usando variable compleja. Por nuestro teorema de convergencia, las ecuaciones (8.204) y (8.205) son convergentes para todos los valores finitos de x . Específicamente, para el producto infinito para el $\text{sen}(x)$, $a_n = x^2/n^2\pi^2$,

$$\begin{aligned} \sum_{n=1}^{\infty} a_n &= \frac{x^2}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{x^2}{\pi^2} \zeta(2) \\ &= \frac{x^2}{6} . \end{aligned} \quad (8.206)$$

La serie correspondiente a la ecuación (8.205) se comporta en una manera similar.

La ecuación (8.204) conduce a dos resultados interesantes. Primero, si fijamos $x = \pi/2$, obtenemos

$$1 = \frac{\pi}{2} \prod_{n=1}^{\infty} \left[1 - \frac{1}{(2n)^2} \right] = \frac{\pi}{2} \prod_{n=1}^{\infty} \left[\frac{(2n)^2 - 1}{(2n)^2} \right] . \quad (8.207)$$

Resolviendo para $\pi/2$, obtenemos

$$\frac{\pi}{2} = \prod_{n=1}^{\infty} \left[\frac{(2n)^2}{(2n-1)(2n+1)} \right] = \frac{2 \cdot 2}{1 \cdot 3} \cdot \frac{4 \cdot 4}{3 \cdot 5} \cdot \frac{6 \cdot 6}{5 \cdot 7} \cdots, \quad (8.208)$$

la cual es la famosa fórmula de Wallis para $\pi/2$.

El segundo resultado involucra la función factorial o función gama. Una definición de la función gama es

$$\Gamma(x) = \left[x e^{\gamma x} \prod_{r=1}^{\infty} \left(1 + \frac{x}{r} \right) e^{-\frac{x}{r}} \right]^{-1}, \quad (8.209)$$

donde γ es la constante de Euler-Mascheroni, sección 8.2. Si tomamos el producto de $\Gamma(x)$ y $\Gamma(-x)$, la ecuación (8.209) tiende a

$$\begin{aligned} \Gamma(x)\Gamma(-x) &= - \left[x e^{\gamma x} \prod_{r=1}^{\infty} \left(1 + \frac{x}{r} \right) e^{-\frac{x}{r}} x e^{-\gamma x} \prod_{r=1}^{\infty} \left(1 - \frac{x}{r} \right) e^{\frac{x}{r}} \right]^{-1} \\ &= - \left[x^2 \prod_{r=1}^{\infty} \left(1 - \frac{x^2}{r^2} \right) \right]^{-1}. \end{aligned} \quad (8.210)$$

Usando la ecuación (8.204) con x reemplazado por πx , obtenemos

$$\Gamma(x)\Gamma(-x) = -\frac{\pi}{x \operatorname{sen}(\pi x)}. \quad (8.211)$$

Anticipando una relación de recurrencia desarrollada posteriormente, tenemos que usando $-x\Gamma(-x) = \Gamma(1-x)$, la ecuación (8.211) puede ser escrita como

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\operatorname{sen}(\pi x)}. \quad (8.212)$$

Esto será útil cuando tratamos la función gama.

Estrictamente hablando, podríamos chequear el intervalo en x para el cual la ecuación (8.209) es convergente. Claramente, para $x = 0, -1, -2, \dots$ los factores individuales se anulan. La prueba que el producto infinito converge para todos los otros valores (finitos) de x es dejado como ejercicio.

Estos productos infinitos tienen una variedad de usos en matemática analítica. Sin embargo, a causa de su lentitud de convergencia, ellas no son aptas para un trabajo numérico preciso.

Capítulo 9

Ecuaciones diferenciales.

versión final 2.1.7 de Julio del 2003¹

9.1. Ecuaciones diferenciales parciales, características y condiciones de borde.

En Física el conocimiento de la fuerza en una ecuación de movimiento usualmente conduce a una ecuación diferencial. Por lo tanto, casi todas las partes elementales y numerosas partes avanzadas de la Física teórica están formuladas en términos de ecuaciones diferenciales. Algunas veces son ecuaciones diferenciales ordinarias en una variable (ODE). Más a menudo las ecuaciones son ecuaciones diferenciales parciales (PDE) en dos o más variables.

Recordemos que la operación de tomar una derivada ordinaria o parcial, es una *operación lineal*² (\mathcal{L})

$$\frac{d(a\varphi(x) + b\psi(x))}{dx} = a \frac{d\varphi}{dx} + b \frac{d\psi}{dx} ,$$

para ODE que involucran derivadas en una variable x solamente y no cuadráticas, $(d\psi/dx)^2$, o potencias mayores. Similarmente, para derivadas parciales,

$$\frac{\partial(a\varphi(x, y) + b\psi(x, y))}{\partial x} = a \frac{\partial\varphi(x, y)}{\partial x} + b \frac{\partial\psi(x, y)}{\partial x} .$$

En general

$$\mathcal{L}(a\varphi + b\psi) = a\mathcal{L}(\varphi) + b\mathcal{L}(\psi) . \quad (9.1)$$

Así, las ODE y las PDE aparecen como ecuaciones de operadores lineales

$$\mathcal{L}(\psi) = F ,$$

donde F es una función conocida de una (para ODE) o más variables (para PDE), \mathcal{L} es una combinación lineal de derivadas, ψ es una función o solución desconocida. Cualquier combinación lineal de soluciones es de nuevo una solución; esto es el *principio de superposición*.

¹Este capítulo está basado en el octavo capítulo del libro: *Mathematical Methods for Physicists, fourth edition* de George B. Arfken & Hans J. Weber, editorial ACADEMIC PRESS.

²Estamos especialmente interesados en operadores lineales porque en mecánica cuántica las cantidades físicas están representadas por operadores lineales operando en un espacio complejo de Hilbert de dimensión infinita.

Ya que la dinámica de muchos sistemas físicos involucran sólo dos derivadas, *e.g.*, la aceleración en mecánica clásica y el operador de energía cinética, $\sim \nabla^2$, en mecánica cuántica, las ecuaciones diferenciales de segundo orden ocurren más frecuentemente en Física. [Las ecuaciones de Maxwell y de Dirac son de primer orden pero involucran dos funciones desconocidas. Eliminando una incógnita conducen a una ecuación diferencial de segundo orden por la otra.]

9.1.1. Ejemplos de PDE.

Entre las PDE más frecuentemente encontradas tenemos:

1. La ecuación de Laplace, $\nabla^2\psi = 0$. Esta ecuación muy común y muy importante aparece en el estudio de
 - a. Fenómenos electromagnéticos incluyendo electroestáticos, dieléctricos, corrientes estacionarias y magnetoestática.
 - b. Hidrodinámica (flujo irrotacional de líquidos perfectos y superficies de ondas).
 - c. Flujo de calor.
 - d. Gravitación.
2. La ecuación de Poisson, $\nabla^2\psi = -4\pi\rho$. En contraste a la ecuación homogénea de Laplace, la ecuación de Poisson es no homogénea con un término de fuente $-4\pi\rho$.
3. Las ecuaciones de onda (Helmholtz) y las ecuaciones de difusión tiempo independiente, $\nabla^2\psi \pm k^2\psi = 0$. Estas ecuaciones aparecen en fenómenos tan diversos como
 - a. Ondas elásticas en sólidos, incluyendo cuerdas vibrantes, barras y membranas.
 - b. En sonido o acústica.
 - c. En ondas electromagnéticas.
 - d. En reactores nucleares.
4. La ecuación de difusión tiempo dependiente

$$\nabla^2\psi = \frac{1}{a^2} \frac{\partial\psi}{\partial t}.$$

5. Las ecuaciones de onda tiempo dependiente,

$$\nabla^2\psi = \frac{1}{c^2} \frac{\partial^2\psi}{\partial t^2}.$$

La forma cuadrimensional que involucra el D'Alembertiano, un análogo cuadrimensional del Laplaciano en el espacio Minkowski,

$$\partial^\mu \partial_\mu = \partial^2 = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2.$$

Luego las ecuaciones de onda tiempo dependiente quedan $\partial^2\psi = 0$.

6. La ecuación del potencial escalar, $\partial^2\psi = 4\pi\rho$. Como la ecuación de Poisson esta ecuación es no homogénea con un término de fuente $4\pi\rho$.
7. La ecuación de Klein-Gordon, $\partial^2\psi = -\mu^2\psi$, y las correspondientes ecuaciones vectoriales en las cuales la función escalar ψ es reemplazada por una función vectorial. Otras formas complicadas son comunes.
8. La ecuación de onda de Schrödinger,

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = i\hbar\frac{\partial\psi}{\partial t}$$

y

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi ,$$

para el caso tiempo independiente.

9. Las ecuaciones para ondas elásticas y líquidos viscosos y la ecuación telegráfica.
10. Ecuaciones diferenciales parciales acopladas de Maxwell para los campos eléctricos y magnéticos son aquellas de Dirac para funciones de ondas relativistas del electrón.

Algunas técnicas generales para resolver PDE de segundo orden son discutidas en esta sección:

1. Separación de variables, donde el PDE es separada en ODEs que están relacionadas por constantes comunes las cuales aparecen como autovalores de operadores lineales, $\mathcal{L}\psi = l\psi$, usualmente en una variable. La ecuación de Helmholtz dada como ejemplo 3 anteriormente tiene esta forma, donde el autovalor k^2 puede surgir por la separación del tiempo t respecto de las variables espaciales. Como en el ejemplo 8, la energía E es el autovalor que surge en la separación de t respecto de \vec{r} en la ecuación de Schrödinger.
2. Conversión de una PDE en una ecuación integral usando funciones de Green que se aplica a PDE no homogéneas tales como los ejemplos 2 y 6 dados más arriba.
3. Otros métodos analíticos tales como el uso de transformadas integrales que serán desarrolladas en el próximo curso.
4. Cálculo numérico. El desarrollo de los computadores ha abierto una abundancia de posibilidades basadas en el cálculo de diferencias finitas. Aquí también tenemos los métodos de relajación. Métodos como Runge-Kutta y predictor-corrector son aplicados a ODEs.

Ocasionalmente, encontramos ecuaciones de orden mayor. En ambos la teoría del movimiento suave de un líquido viscoso y la teoría de un cuerpo elástico encontramos la ecuación

$$(\nabla^2)^2\psi = 0 .$$

Afortunadamente, estas ecuaciones diferenciales de orden más altos son relativamente raras y no son discutidas en una etapa introductoria como esta.

Aunque no son tan frecuentemente encontrados y quizás no son tan importantes como las ecuaciones diferenciales de segundo orden, las ecuaciones diferenciales de primer orden aparecen en Física teórica y algunas veces son pasos intermedios para ecuaciones diferenciales de segundo orden. Las soluciones de algunos de los tipos más importantes de ODE de primer orden son desarrollados en la sección 9.2. Las PDEs de primer orden siempre pueden ser reducidas a ODEs. Este es un proceso directo pero lento e involucra una búsqueda para las características que son presentadas brevemente más adelante.

9.1.2. Clases de PDE y característica.

Las PDEs de segundo orden forman tres clases:

- (i) Las PDEs elípticas que involucran ∇^2 o $c^{-2}\partial^2/\partial t^2 + \nabla^2$.
- (ii) Las PDEs parabólica, $a\partial/\partial t - \nabla^2$.
- (iii) Las PDEs hiperbólica, $c^{-2}\partial^2/\partial t^2 - \nabla^2$.

Estos operadores canónicos aparecen por un cambio de variables $\xi = \xi(x, y)$, $\eta = \eta(x, y)$ en un operador lineal (para dos variables sólo por simplicidad)

$$\mathcal{L} = a \frac{\partial^2}{\partial x^2} + 2b \frac{\partial^2}{\partial x \partial y} + c \frac{\partial^2}{\partial y^2} + d \frac{\partial}{\partial x} + e \frac{\partial}{\partial y} + f, \quad (9.2)$$

la cual puede ser reducida a las formas canónicas (i), (ii), (iii) de acuerdo a si el discriminante $D = ac - b^2 > 0$, $= 0$ o < 0 . Si $\xi(x, y)$ es determinada a partir de la ecuación de primer orden, pero no lineal, PDE

$$a \left(\frac{\partial \xi}{\partial x} \right)^2 + 2b \left(\frac{\partial \xi}{\partial x} \right) \left(\frac{\partial \xi}{\partial y} \right) + c \left(\frac{\partial \xi}{\partial y} \right)^2 = 0, \quad (9.3)$$

donde los términos de más bajo orden en \mathcal{L} son ignorados, entonces los coeficientes de $\partial^2/\partial \xi^2$ en \mathcal{L} es cero (*i.e.*, ecuación (9.3)). Si η es una solución independiente de la misma ecuación (9.3), entonces el coeficiente de $\partial^2/\partial \eta^2$ también es cero. El operador remanente $\partial^2/\partial \xi \partial \eta$ en \mathcal{L} es característico del caso hiperbólico (iii) con $D < 0$, donde la forma cuadrática $a\lambda^2 + 2b\lambda + c$ es factorizable y, por lo tanto, la ecuación (9.3) tiene dos soluciones independientes $\xi(x, y)$, $\eta(x, y)$. En el caso elíptico (i) con $D > 0$ las dos soluciones ξ , η son complejos conjugados los cuales, cuando se sustituyeron en la ecuación (9.2), remueven la derivada de segundo orden mezclada en vez de los otros términos de segundo orden produciendo la forma canónica (i). En el caso parabólico (ii) con $D = 0$, solamente $\partial^2/\partial \xi^2$ permanece en \mathcal{L} , mientras que los coeficientes de las otras dos derivadas de segundo orden se anulan.

Si los coeficientes a , b , c en \mathcal{L} son funciones de las coordenadas, entonces esta clasificación es solamente local, *i.e.*, su tipo podría cambiar cuando las coordenadas varían.

Ilustremos la física implícita en el caso hiperbólico mirando la ecuación de onda (en 1 + 1 dimensiones por simplicidad)

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} \right) \psi = 0. \quad (9.4)$$

Ya que la ecuación (9.3) se convierte en

$$\left(\frac{\partial \xi}{\partial t}\right)^2 - c^2 \left(\frac{\partial \xi}{\partial x}\right)^2 = \left(\frac{\partial \xi}{\partial t} - c \frac{\partial \xi}{\partial x}\right) \left(\frac{\partial \xi}{\partial t} + c \frac{\partial \xi}{\partial x}\right) = 0, \quad (9.5)$$

y es factorizable, determinamos la solución de $\partial \xi / \partial t - c \partial \xi / \partial x = 0$. Esta es una función arbitraria $\xi = F(x + ct)$, y $\xi = G(x - ct)$ resuelve $\partial \xi / \partial t + c \partial \xi / \partial x = 0$, la cual se verifica rápidamente. Por superposición lineal una solución general de la ecuación (9.4) es la suma $\psi = F(x + ct) + G(x - ct)$. Para funciones periódicas F, G reconocemos los argumentos $x + ct$ y $x - ct$ como la fase de la onda plana o frente de ondas, donde las soluciones de la ecuación de onda (9.4) cambian abruptamente (de cero a sus valores actuales) y no están únicamente determinadas. Normal al frente de onda están los rayos de la óptica geométrica. De este modo, las soluciones de la ecuación (9.5) o (9.3) más generalmente, son llamadas *características* o algunas veces *bicaracterísticas* (para PDE de segundo orden) en la literatura matemática corresponde a los frente de ondas de la solución de la óptica geométrica de la ecuación de onda completa.

Para el caso elíptico consideremos la ecuación de Laplace

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0, \quad (9.6)$$

para un potencial ψ de dos variables. Aquí la ecuación característica es

$$\left(\frac{\partial \xi}{\partial x}\right)^2 + \left(\frac{\partial \xi}{\partial y}\right)^2 = \left(\frac{\partial \xi}{\partial x} + i \frac{\partial \xi}{\partial y}\right) \left(\frac{\partial \xi}{\partial x} - i \frac{\partial \xi}{\partial y}\right) = 0 \quad (9.7)$$

tiene soluciones complejas conjugadas: $\xi = F(x + iy)$ para $\partial \xi / \partial x + i \partial \xi / \partial y = 0$ y $\xi = G(x - iy)$ para $\partial \xi / \partial x - i \partial \xi / \partial y = 0$. Una solución general de la ecuación de potencial (9.6) es por lo tanto $\psi = F(x + iy) + iG(x - iy)$ Tanto la parte real como la imaginaria de ψ , son llamadas *funciones armónicas*, mientras que las soluciones polinomiales son llamadas *polinomios armónicos*.

En mecánica cuántica la forma de Wentzel-Kramers-Brillouin (WKB) de $\psi = \exp(-iS/\hbar)$ para la solución de la ecuación de Schrödinger

$$\left(-\frac{\hbar^2}{2m} \nabla^2 + V\right) \psi = i\hbar \frac{\partial \psi}{\partial t}, \quad (9.8)$$

conduce a la ecuación Hamilton-Jacobi de la mecánica clásica,

$$\frac{1}{2m} (\vec{\nabla} S)^2 + V = \frac{\partial S}{\partial t}, \quad (9.9)$$

en el límite $\hbar \rightarrow 0$. La acción clásica de S entonces llega a ser la característica de la ecuación de Schrödinger. Sustituyendo $\vec{\nabla} \psi = -i\psi \vec{\nabla} S / \hbar$, $\partial \psi / \partial t = -i\psi \partial S / \partial t / \hbar$ en la ecuación (9.8), dejando la totalidad de los factores de ψ no nulos, y aproximando el Laplaciano $\nabla^2 \psi = -i\psi \nabla^2 S / \hbar - \psi (\nabla S)^2 / \hbar^2 \simeq -\psi (\nabla S)^2$, *i.e.*, despreciando $-i\nabla^2 \psi / \hbar$, realmente obtenemos la ecuación (9.9).

Resolver las características es una de las técnicas generales de encontrar las soluciones de PDE. Para más ejemplos y tratamientos detallados de las características, las cuales no perseguimos aquí, nos referimos a H. Bateman, *Partial Differential Equations of Mathematical Physics*. New York: Dover (1994); K.E. Gustafson, *Partial Differential Equations and Hilbert Space Methods*, 2nd ed. New York: Wiley (1987).

9.1.3. Las PDE no lineales.

Las ODEs y PDEs no lineales son un campo importante y de rápido crecimiento. Encontramos más arriba la ecuación de onda lineal más simple

$$\frac{\partial \psi}{\partial t} + c \frac{\partial \psi}{\partial x} = 0 ,$$

como la PDE de primer orden a partir de la característica de la ecuación de onda. La ecuación de onda no lineal más simple

$$\frac{\partial \psi}{\partial t} + c(\psi) \frac{\partial \psi}{\partial x} = 0 , \quad (9.10)$$

resulta si la velocidad local de propagación, c , no es constante sino que depende de la onda ψ . Cuando una ecuación no lineal tiene una solución de la forma $\psi(x, t) = A \cos(kx - \omega t)$, donde $\omega(k)$ varía con k tal que $\omega''(k) \neq 0$, entonces ella es llamada *dispersiva*. Quizás la ecuación dispersiva no lineal más conocida de segundo orden es la ecuación de Korteweg-de Vries

$$\frac{\partial \psi}{\partial t} + \psi \frac{\partial \psi}{\partial x} + \frac{\partial^3 \psi}{\partial x^3} = 0 , \quad (9.11)$$

la cual modela la propagación sin pérdidas de las ondas de agua superficiales y otros fenómenos. Esta es ampliamente conocida por sus soluciones *solitón*. Un solitón es una onda viajera con la propiedad de persistir a través de una interacción con otro solitón: después de que ellos pasan uno a través del otro, ellos emergen en la misma forma y con la misma velocidad y no adquieren más que un cambio de fase. Sea $\psi(\xi = x - ct)$ tal onda viajera. Cuando es sustituida en la ecuación (9.11) esta produce la ODE no lineal

$$(\psi - c) \frac{d\psi}{d\xi} + \frac{d^3 \psi}{d\xi^3} = 0 , \quad (9.12)$$

la cual puede ser integrada dando

$$\frac{d^2 \psi}{d\xi^2} = c\psi - \frac{\psi^2}{2} . \quad (9.13)$$

No hay constantes de integración aditivas en la ecuación (9.13) para asegurar que se satisfaga la condición $d^2 \psi / d\xi^2 \rightarrow 0$ con $\psi \rightarrow 0$ para ξ grande, tal que ψ está localizado en la característica $\xi = 0$, o $x = ct$. Multiplicando la ecuación (9.13) por $d\psi/d\xi$ e integrando nuevamente tenemos

$$\left(\frac{d\psi}{d\xi} \right)^2 = c\psi^2 - \frac{\psi^3}{3} , \quad (9.14)$$

donde $d\psi/d\xi \rightarrow 0$ para ξ grande. Tomando la raíz de la ecuación (9.14) e integrando una vez más encontramos la solución solitónica

$$\psi(x - ct) = \frac{3c}{\cosh^2 \left(\sqrt{c} \frac{x - ct}{2} \right)} . \quad (9.15)$$

9.1.4. Condiciones de borde.

Usualmente, cuando conocemos un sistema físico en algún momento y la ley que rige ese proceso físico, entonces somos capaces de predecir el desarrollo subsecuente. Tales valores iniciales son las más comunes condiciones de borde asociadas con ODEs y PDEs. Encontrando soluciones que calcen con los puntos, curvas o superficies dados correspondientes al problema de valores de contorno. Las autofunciones usualmente requieren que satisfagan ciertas condiciones de borde impuestas (*e.g.*, asintóticas). Estas condiciones pueden ser tomadas de tres formas:

1. Condiciones de borde de Cauchy. El valor de una función y su derivada normal es especificada en el borde. En electrostática estas significarían φ , el potencial, y E_n la componente normal del campo eléctrico.
2. Condiciones de borde de Dirichlet. El valor específico en el borde.
3. Condiciones de borde de Neumann. La derivada normal (gradiente normal) de una función específica en el borde. En el caso electrostático este sería E_n y por lo tanto σ , la densidad de carga superficial.

Un resumen de las relaciones de estos tres tipos de condiciones de borde con los tres tipos de ecuaciones diferenciales parciales bidimensionales están dadas en la tabla 9.1. Para discusiones más extensas de estas ecuaciones diferenciales parciales puede consultar Sommerfeld, capítulo 2, o Morse y Feshbach, capítulo 6.

Partes de la tabla 9.1 son simplemente un asunto de mantener la consistencia interna, o sentido común. Por ejemplo, para la ecuación de Poisson con una superficie cerrada, las condiciones de Dirichlet conducen a una solución única y estable. Las condiciones de Neumann, independiente de las condiciones de Dirichlet, del mismo modo conducen a una solución única y estable independiente de la solución de Dirichlet. Por lo tanto las condiciones de borde de Cauchy (lo que significa la de Dirichlet más la de Neumann) conducen a una inconsistencia.

El término de condiciones de borde incluye como un caso especial el concepto de condiciones iniciales. Por ejemplo, especificando la posición inicial x_0 y la velocidad inicial v_0 en algunos problemas de dinámica correspondería a condiciones de borde de Cauchy. La única diferencia en el presente uso de las condiciones de borde en estos problemas unidimensionales es que estamos aplicando las condiciones en *ambos* extremos del intervalo permitido de la variable.

9.2. Ecuaciones diferenciales de primer orden.

La física involucra algunas ecuaciones diferenciales de primer orden, ellas fueron estudiadas en el curso de ecuaciones diferenciales. Por completitud parece ser deseable revisarlas brevemente.

Consideremos aquí ecuaciones diferenciales de la forma general

$$\frac{dy}{dx} = f(x, y) = -\frac{P(x, y)}{Q(x, y)}. \quad (9.16)$$

Condiciones de borde	Tipo de ecuación diferencial parcial		
	Elípticas	Hiperbólicas	Parabólicas
	Laplace, Poisson en (x, y)	Ecuación de Ondas en (x, t)	Ecuación de difusión en (x, t)
Cauchy			
Superficie Abierta	Resultados no físicos (inestabilidades)	<i>Solución única y estable</i>	Demasiado restrictivo
Superficie Cerrada	Demasiado restrictivo	Demasiado restrictivo	Demasiado restrictivo
Dirichlet			
Superficie Abierta	Insuficiente	Insuficiente	<i>Solución única y estable en 1 dim</i>
Superficie Cerrada	<i>Solución única y estable</i>	Solución no única	Demasiado restrictivo
Neumann			
Superficie Abierta	Insuficiente	Insuficiente	<i>Solución única y estable en 1 dim</i>
Superficie Cerrada	<i>Solución única y estable</i>	Solución no única	Demasiado restrictivo

Cuadro 9.1:

La ecuación (9.16) es claramente una ecuación de primer orden ordinaria. Es de *primer orden* ya que contiene la primera derivada y no mayores. Es *Ordinaria* ya que la derivada dy/dx es una derivada ordinaria o total. La ecuación (9.16) puede o no puede ser *lineal*, aunque trataremos el caso lineal explícitamente más adelante.

9.2.1. Variables separables.

Frecuentemente la ecuación (9.16) tendrá la forma especial

$$\frac{dy}{dx} = f(x, y) = -\frac{P(x)}{Q(y)}. \quad (9.17)$$

Entonces la podemos reescribir como

$$P(x)dx + Q(y)dy = 0.$$

Integrando de (x_0, y_0) a (x, y) tiende a

$$\int_{x_0}^x P(x')dx' + \int_{y_0}^y Q(y')dy' = 0. \quad (9.18)$$

Ya que los límites inferiores x_0 e y_0 contribuyen en unas constantes, podríamos ignorar los límites inferiores de integración y simplemente añadir una constante de integración al final.

Note que esta técnica de separación de variables *no* requiere que la ecuación diferencial sea lineal.

Ejemplo Ley de Boyle.

Una forma diferencial de la ley de los gases de Boyle es

$$\frac{dV}{dP} = -\frac{V}{P},$$

para el volumen V de una cantidad fija de gas a presión P (y temperatura constante). Separando variables, tenemos

$$\frac{dV}{V} = -\frac{dP}{P}$$

o

$$\ln V = -\ln P + C.$$

Con dos logaritmos presentes, es más conveniente reescribir la constante de integración C como $\ln k$. Entonces

$$\ln V + \ln P = \ln PV = \ln k$$

y

$$PV = k.$$

9.2.2. Ecuaciones diferenciales exactas.

Reescribimos la ecuación (9.16) como

$$P(x, y)dx + Q(x, y)dy = 0. \quad (9.19)$$

Esta ecuación se dice que es *exacta* si podemos calzar el lado izquierdo de ella a un diferencial $d\varphi$,

$$d\varphi = \frac{\partial\varphi}{\partial x}dx + \frac{\partial\varphi}{\partial y}dy. \quad (9.20)$$

Ya que la ecuación (9.19) tiene un cero a la derecha, buscamos una función desconocida $\varphi(x, y) = \text{constante}$, tal que $d\varphi = 0$. Tenemos (si tal función $\varphi(x, y)$ existe)

$$P(x, y)dx + Q(x, y)dy = \frac{\partial\varphi}{\partial x}dx + \frac{\partial\varphi}{\partial y}dy \quad (9.21)$$

y

$$\frac{\partial\varphi}{\partial x} = P(x, y), \quad \frac{\partial\varphi}{\partial y} = Q(x, y). \quad (9.22)$$

La condición necesaria y suficiente para que la ecuación sea exacta es que la segunda derivada parcial mezclada de $\varphi(x, y)$ (supuesta continua) es independiente del orden de diferenciación:

$$\frac{\partial^2\varphi}{\partial y\partial x} = \frac{\partial P(x, y)}{\partial y} = \frac{\partial Q(x, y)}{\partial x} = \frac{\partial^2\varphi}{\partial x\partial y}. \quad (9.23)$$

Si la ecuación (9.19) corresponde a un rotor (igual cero), entonces un potencial, $\varphi(x, y)$, debiera existir.

Si $\varphi(x, y)$ existe entonces a partir de las ecuaciones (9.19) y (9.21) nuestra solución es

$$\varphi(x, y) = C . \quad (9.24)$$

Podemos construir $\varphi(x, y)$ a partir de sus derivadas parciales de la misma manera que construimos un potencial magnético vectorial en el capítulo de vectores a partir de su rotor.

Podemos volver a la ecuación (9.19) y ver qué pasa si no es exacta: la ecuación (9.23) no es satisfecha. Sin embargo, siempre existe al menos una o quizás una infinidad de factores de integración, $\alpha(x, y)$, tales que

$$\alpha(x, y)P(x, y)dx + \alpha(x, y)Q(x, y)dy = 0$$

es exacta. Desafortunadamente, un factor de integración no siempre es obvio o fácil de encontrar. Diferente es el caso de la ecuación diferencial de primer orden *lineal* considerada a continuación, no hay una manera sistemática de desarrollar un factor de integración para la ecuación (9.19).

Una ecuación diferencial en la cual las variables han sido separadas es automáticamente exacta. Una ecuación diferencial exacta *no* es necesariamente separable.

9.2.3. Ecuaciones diferenciales ordinarias de primer orden lineales.

Si $f(x, y)$ en la ecuación (9.16) tiene la forma $-p(x)y + q(x)$, entonces la ecuación (9.16) se convierte en

$$\frac{dy}{dx} + p(x)y = q(x) . \quad (9.25)$$

La ecuación (9.25) es la ODE de primer orden *lineal* más general. Si $q(x) = 0$, la ecuación (9.25) es homogénea (en y). Un $q(x)$ distinto de cero puede representar una fuente o un término de forzamiento. La ecuación (9.25) es *lineal*; cada término es lineal en y o dy/dx . No hay potencias mayores; esto es, no hay y^2 , ni productos, $y(dy/dx)$. Note que la linealidad se refiere a y y a la dy/dx ; $p(x)$ y $q(x)$ no es necesario que sean lineales en x . La ecuación (9.25), es la más importante de estas ecuaciones diferenciales de primer orden para los físicos y puede ser resuelta exactamente.

Busquemos un *factor de integración* $\alpha(x)$ tal que

$$\alpha(x)\frac{dy}{dx} + \alpha(x)p(x)y = \alpha(x)q(x) , \quad (9.26)$$

puede ser reescrito como

$$\frac{d}{dx}[\alpha(x)y] = \alpha(x)q(x) . \quad (9.27)$$

El propósito de esto es hacer el lado izquierdo de la ecuación (9.25) una derivada total que pueda ser integrada por inspección. Esto también, incidentalmente, hace la ecuación (9.25) exacta. Expandiendo la ecuación (9.27), obtenemos

$$\alpha(x)\frac{dy}{dx} + \frac{d\alpha}{dx}y = \alpha(x)q(x) .$$

La comparación con la ecuación (9.26) muestra que debemos requerir que

$$\frac{d\alpha(x)}{dx} = \alpha(x)p(x) . \quad (9.28)$$

Aquí hay una ecuación diferencial para $\alpha(x)$, con las variables α y x separables. Separamos variables, integramos, y obtenemos

$$\alpha(x) = \exp \left[\int^x p(x') dx' \right] \quad (9.29)$$

como nuestro factor de integración.

Con $\alpha(x)$ conocida procedemos a integrar la ecuación (9.27). Esto, por supuesto, fue el objetivo de introducir α en primer lugar. Tenemos

$$\int^x \frac{d}{dx'} [\alpha(x')y] dx' = \int^x \alpha(x')q(x') dx' .$$

Ahora integrando por inspección, tenemos

$$\alpha(x)y = \int^x \alpha(x')q(x') dx' + C .$$

Las constantes a partir del límite inferior de integración constante son reunidas en la constante C . Dividiendo por $\alpha(x)$, obtenemos

$$y(x) = \frac{1}{\alpha(x)} \left\{ \int^x \alpha(x')q(x') dx' + C \right\} .$$

Finalmente, sustituyendo en la ecuación (9.29) por α conduce

$$y(x) = \exp \left[- \int^x p(t) dt \right] \left\{ \int^x \exp \left[\int^s p(t) dt \right] q(s) ds + C \right\} . \quad (9.30)$$

Aquí las variables mudas de integración han sido reescritas para hacerlas inambiguas. La ecuación (9.30) es la solución general completa de la ecuación diferencial lineal, de primer orden, la ecuación (9.25). La porción

$$y_1(x) = C \exp \left[- \int^x p(t) dt \right] \quad (9.31)$$

corresponde al caso $q(x) = 0$ y es solución general de la ecuación diferencial homogénea. El otro término en la ecuación (9.30),

$$y(x) = \exp \left[- \int^x p(t) dt \right] \int^x \exp \left[\int^s p(t) dt \right] q(s) ds , \quad (9.32)$$

es una solución particular que corresponde al término específico de fuente $q(x)$.

Podemos notar que si nuestra ecuación diferencial de primer orden es homogénea ($q = 0$), entonces ella es separable. De lo contrario, salvo casos especiales tal como $p = \text{constante}$, $q = \text{constante}$, o $q(x) = ap(x)$, la ecuación (9.25) no es separable.

Ejemplo Circuito RL.

Para un circuito resistencia-inductancia las leyes de Kirchhoff producen

$$L \frac{dI(t)}{dt} + RI(t) = V(t) ,$$

para la corriente $I(t)$, donde L es la inductancia y R es la resistencia, ambas constantes. $V(t)$ es el voltaje aplicado tiempo dependiente.

De la ecuación (9.29) nuestro factor de integración $\alpha(t)$ es

$$\begin{aligned} \alpha(t) &= \exp \int^t \frac{R}{L} dt \\ &= e^{Rt/L} . \end{aligned}$$

Entonces por la ecuación (9.30)

$$I(t) = e^{-Rt/L} \left[\int^t e^{Rt/L} \frac{V(t)}{L} dt + C \right] ,$$

con la constante C es determinada por una condición inicial (una condición de borde).

Para el caso especial $V(t) = V_0$, una constante,

$$\begin{aligned} I(t) &= e^{-Rt/L} \left[\frac{V_0}{L} \frac{L}{R} e^{Rt/L} + C \right] \\ &= \frac{V_0}{R} + C e^{-Rt/L} . \end{aligned}$$

Si la condición inicial es $I(0) = 0$, entonces $C = -V_0/R$ y

$$I(t) = \frac{V_0}{R} [1 - e^{-Rt/L}] .$$

9.2.4. Conversión a una ecuación integral.

Nuestra ecuación diferencial de primer orden, ecuación (9.16), puede ser convertida a una ecuación integral por integración directa:

$$y(x) - y(x_0) = \int_{x_0}^x f[x, y(x)] dx . \quad (9.33)$$

Como una ecuación integral hay una posibilidad de una solución en serie de Neumann (se verá en el próximo curso) con la aproximación inicial $y(x) \approx y(x_0)$. En la literatura de ecuaciones diferenciales esto es llamado el “método de Picard de aproximaciones sucesivas”.

Ecuaciones diferenciales de primer orden las encontraremos de nuevo en conexión con las transformadas de Laplace y de Fourier.

9.3. Separación de variables.

Las ecuaciones de la física matemática listada en la sección 9.1 son todas ecuaciones diferenciales parciales. Nuestra primera técnica para su solución es dividir la ecuación diferencial parcial en n ecuaciones diferenciales ordinarias de n variables. Cada separación introduce una constante de separación arbitraria. Si tenemos n variables, tenemos que introducir $n - 1$ constantes, determinadas por las condiciones impuestas al resolver el problema.

9.3.1. Coordenadas cartesianas.

En coordenadas cartesianas las ecuaciones de Helmholtz llegan a ser

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + k^2 \psi = 0 , \quad (9.34)$$

usando la forma cartesiana para el Laplaciano. Por el momento, k^2 será una constante. Quizás la manera más simple de tratar una ecuación diferencial parcial tal como la ecuación (9.34) es dividirla en un conjunto de ecuaciones diferenciales ordinarias. Esto puede ser hecho como sigue. Sea

$$\psi(x, y, z) = X(x)Y(y)Z(z) , \quad (9.35)$$

y sustituir de vuelta en la ecuación (9.34). ¿Cómo sabemos que la ecuación (9.35) es válida?. La respuesta es muy simple: ¡No sabemos si es válida!. Mejor dicho, estamos procediendo en este espíritu y tratando de ver si trabaja. Si nuestro intento es exitoso, entonces la ecuación (9.35) será justificada. Si no es exitoso, lo descubriremos pronto y luego trataremos otro ataque tal como las funciones de Green, transformadas integral, o análisis numérico a la fuerza bruta. Con ψ supuestamente dada por la ecuación (9.35), la ecuación (9.34) llega a ser

$$YZ \frac{d^2 X}{dx^2} + XZ \frac{d^2 Y}{dy^2} + XY \frac{d^2 Z}{dz^2} + k^2 XYZ = 0 . \quad (9.36)$$

Dividiendo por $\psi = XYZ$ y reorganizando los términos, obtenemos

$$\frac{1}{X} \frac{d^2 X}{dx^2} = -k^2 - \frac{1}{Y} \frac{d^2 Y}{dy^2} - \frac{1}{Z} \frac{d^2 Z}{dz^2} . \quad (9.37)$$

La ecuación (9.37) exhibe una separación de variables. El lado izquierdo es sólo función de x , mientras que el lado derecho depende solamente de y y z . Así la ecuación (9.37) es una clase de paradoja. Una función de x es igualada a una función de y y z , pero x , y y z son todas coordenadas independientes. Esta independencia significa que el comportamiento de x como una variable independiente no está determinada ni por y ni por z . La paradoja está resuelta fijando cada lado igual a una constante, una constante de separación. Escogemos³

$$\frac{1}{X} \frac{d^2 X}{dx^2} = -l^2 , \quad (9.38)$$

³La elección de signo es completamente arbitraria, será fijada en un problema específico por la necesidad de satisfacer las condiciones de borde.

$$-k^2 - \frac{1}{Y} \frac{d^2 Y}{dy^2} - \frac{1}{Z} \frac{d^2 Z}{dz^2} = -l^2 . \quad (9.39)$$

Ahora, volviendo nuestra atención a la ecuación (9.39), obtenemos

$$\frac{1}{Y} \frac{d^2 Y}{dy^2} = -k^2 + l^2 - \frac{1}{Z} \frac{d^2 Z}{dz^2} , \quad (9.40)$$

y una segunda separación ha sido realizada. Aquí tenemos una función de y igualada a una función de z y aparece la misma paradoja. La resolvemos como antes igualando cada lado a otra constante de separación, $-m^2$,

$$\frac{1}{Y} \frac{d^2 Y}{dy^2} = -m^2 , \quad (9.41)$$

$$\frac{1}{Z} \frac{d^2 Z}{dz^2} = -k^2 + l^2 + m^2 = -n^2 , \quad (9.42)$$

introduciendo una constante n^2 por $k^2 = l^2 + m^2 + n^2$ para producir un conjunto simétrico de ecuaciones. Ahora tenemos tres ecuaciones diferenciales ordinarias ((9.38), (9.41), y (9.42)) para reemplazar en la ecuación (9.34). Nuestra suposición (ecuación (9.35)) ha sido exitosa y es por lo tanto justificada.

Nuestra solución sería etiquetada de acuerdo a la elección de nuestras constantes l , m , n , esto es,

$$\psi_{lmn}(x, y, z) = X_l(x)Y_m(y)Z_n(z) . \quad (9.43)$$

Sujeto a las condiciones del problema que se resuelve y a la condición $k^2 = l^2 + m^2 + n^2$, podemos escoger l , m , n como queramos, y la ecuación (9.43) será todavía una solución de la ecuación (9.34), dado que $X_l(x)$ es una solución de la ecuación (9.38) y así seguimos. Podemos desarrollar la solución más general de la ecuación (9.34) tomando una combinación lineal de soluciones ψ_{lmn} ,

$$\Psi = \sum_{l,m,n} a_{lmn} \psi_{lmn} . \quad (9.44)$$

Los coeficientes constantes a_{lmn} finalmente son escogidos para permitir que Ψ satisfaga las condiciones de borde del problema.

9.3.2. Coordenadas cilíndricas circulares.

Si consideramos que nuestra función desconocida ψ depende de ρ , φ , z la ecuación de Helmholtz se convierte en

$$\nabla^2 \psi(\rho, \varphi, z) + k^2 \psi(\rho, \varphi, z) = 0 , \quad (9.45)$$

o

$$\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \psi}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2 \psi}{\partial \varphi^2} + \frac{\partial^2 \psi}{\partial z^2} + k^2 \psi = 0 . \quad (9.46)$$

Como antes, suponemos una forma factorizada para ψ ,

$$\psi(\rho, \varphi, z) = P(\rho)\Phi(\varphi)Z(z) . \quad (9.47)$$

Sustituyendo en la ecuación (9.46), tenemos

$$\frac{\Phi Z}{\rho} \frac{d}{d\rho} \left(\rho \frac{dP}{d\rho} \right) + \frac{PZ}{\rho^2} \frac{d^2\Phi}{d\varphi^2} + P\Phi \frac{d^2Z}{dz^2} + k^2 P\Phi Z = 0 . \quad (9.48)$$

Todas las derivadas parciales han llegado a ser derivadas ordinarias. Dividiendo por $P\Phi Z$ y moviendo la derivada z al lado derecho conduce a

$$\frac{1}{P\rho} \frac{d}{d\rho} \left(\rho \frac{dP}{d\rho} \right) + \frac{1}{\rho^2\Phi} \frac{d^2\Phi}{d\varphi^2} + k^2 = -\frac{1}{Z} \frac{d^2Z}{dz^2} . \quad (9.49)$$

De nuevo, tenemos la paradoja. Una función de z en la derecha aparece dependiendo de una función de ρ y φ en el lado izquierdo. Resolvemos la paradoja haciendo cada lado de la ecuación (9.49) igual a una constante, la misma constante. Escojamos⁴ $-l^2$. Entonces

$$\frac{d^2Z}{dz^2} = l^2 Z , \quad (9.50)$$

y

$$\frac{1}{P\rho} \frac{d}{d\rho} \left(\rho \frac{dP}{d\rho} \right) + \frac{1}{\rho^2\Phi} \frac{d^2\Phi}{d\varphi^2} + k^2 = -l^2 . \quad (9.51)$$

Ajustando $k^2 + l^2 = n^2$, multiplicando por ρ^2 , y reordenando términos, obtenemos

$$\frac{\rho}{P} \frac{d}{d\rho} \left(\rho \frac{dP}{d\rho} \right) + n^2 \rho^2 = -\frac{1}{\Phi} \frac{d^2\Phi}{d\varphi^2} . \quad (9.52)$$

Podemos ajustar el lado derecho a m^2 y

$$\frac{d^2\Phi}{d\varphi^2} = -m^2\Phi \quad (9.53)$$

Finalmente, para la dependencia en ρ tenemos

$$\rho \frac{d}{d\rho} \left(\rho \frac{dP}{d\rho} \right) + (n^2 \rho^2 - m^2)P = 0 . \quad (9.54)$$

Esta es la ecuación diferencial de Bessel. La solución y sus propiedades serán presentadas en el próximo curso. La separación de variables de la ecuación de Laplace en coordenadas parabólicas también conduce a ecuaciones de Bessel. Puede notarse que la ecuación de Bessel es notable por la variedad de formas que puede asumir.

La ecuación original de Helmholtz, una ecuación diferencial parcial tridimensional, ha sido reemplazada por tres ecuaciones diferenciales ordinarias, las ecuaciones (9.50), (9.53) y (9.54). Una solución de la ecuación de Helmholtz es

$$\psi(\rho, \varphi, z) = P(\rho)\Phi(\varphi)Z(z) . \quad (9.55)$$

Identificando las soluciones específicas P , Φ , Z por subíndices, vemos que la solución más general de la ecuación de Helmholtz es una combinación lineal del producto de soluciones:

$$\psi(\rho, \varphi, z) = \sum_{m,n} a_{mn} P_{mn}(\rho)\Phi_m(\varphi)Z_n(z) . \quad (9.56)$$

⁴La elección del signo de la constante de separación es arbitraria. Sin embargo, elegimos un signo menos para la coordenada axial z en espera de una posible dependencia exponencial en z . Un signo positivo es elegido para la coordenada azimutal φ en espera de una dependencia periódica en φ .

9.3.3. Coordenadas polares esféricas.

Tratemos de separar la ecuación de Helmholtz, de nuevo con k^2 constante, en coordenadas polares esféricas. Usando la expresión del Laplaciano en estas coordenadas obtenemos

$$\frac{1}{r^2 \operatorname{sen} \theta} \left[\operatorname{sen} \theta \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{\partial}{\partial \theta} \left(\operatorname{sen} \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{\operatorname{sen} \theta} \frac{\partial^2 \psi}{\partial \varphi^2} \right] = -k^2 \psi . \quad (9.57)$$

Ahora, en analogía con la ecuación (9.35) tratamos

$$\psi(r, \theta, \varphi) = R(r)\Theta(\theta)\Phi(\varphi) . \quad (9.58)$$

Sustituyendo de vuelta en la ecuación (9.57) y dividiendo por $R\Theta\Phi$, tenemos

$$\frac{1}{Rr^2} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{1}{\Theta r^2 \operatorname{sen} \theta} \frac{d}{d\theta} \left(\operatorname{sen} \theta \frac{d\Theta}{d\theta} \right) + \frac{1}{\Phi r^2 \operatorname{sen}^2 \theta} \frac{d^2 \Phi}{d\varphi^2} = k^2 . \quad (9.59)$$

Note que todas las derivadas son ahora derivadas ordinarias más que parciales. Multiplicando por $r^2 \operatorname{sen}^2 \theta$, podemos aislar $(1/\Phi)(d^2\Phi/d\varphi^2)$ para obtener⁵

$$\frac{1}{\Phi} \frac{d^2 \Phi}{d\varphi^2} = r^2 \operatorname{sen}^2 \theta \left[-k^2 - \frac{1}{r^2 R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) - \frac{1}{r^2 \operatorname{sen} \theta \Theta} \frac{d}{d\theta} \left(\operatorname{sen} \theta \frac{d\Theta}{d\theta} \right) \right] . \quad (9.60)$$

La ecuación (9.60) relaciona una función únicamente de φ con una función de r y θ . Ya que r , θ , y φ son variables independientes, igualamos cada lado de la ecuación (9.60) a una constante. Aquí una pequeña consideración puede simplificar el análisis posterior. En casi todos los problemas físicos φ aparecerá como un ángulo azimutal. Esto sugiere una solución periódica más que una exponencial. Con esto en mente, usemos $-m^2$ como la constante de separación. Cualquier constante lo hará, pero ésta hará la vida un poquito más fácil. Entonces

$$\frac{1}{\Phi} \frac{d^2 \Phi}{d\varphi^2} = -m^2 \quad (9.61)$$

y

$$\frac{1}{r^2 R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{1}{r^2 \operatorname{sen} \theta \Theta} \frac{d}{d\theta} \left(\operatorname{sen} \theta \frac{d\Theta}{d\theta} \right) - \frac{m^2}{r^2 \operatorname{sen}^2 \theta} = -k^2 . \quad (9.62)$$

Multiplicando la ecuación (9.62) por r^2 y reordenando términos, tenemos

$$\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + r^2 k^2 = -\frac{1}{\operatorname{sen} \theta \Theta} \frac{d}{d\theta} \left(\operatorname{sen} \theta \frac{d\Theta}{d\theta} \right) + \frac{m^2}{\operatorname{sen}^2 \theta} . \quad (9.63)$$

Nuevamente, las variables son separadas. Igualamos cada lado a una constante Q y finalmente obtenemos

$$\frac{1}{\operatorname{sen} \theta} \frac{d}{d\theta} \left(\operatorname{sen} \theta \frac{d\Theta}{d\theta} \right) - \frac{m^2}{\operatorname{sen}^2 \theta} \Theta + Q\Theta = 0 , \quad (9.64)$$

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + k^2 R - \frac{QR}{r^2} = 0 . \quad (9.65)$$

⁵El orden en el cual las variables son separadas aquí no es único. Muchos textos de mecánica cuántica separan la dependencia en r primero.

Una vez más hemos reemplazado una ecuación diferencial parcial de tres variables por tres ecuaciones diferenciales ordinarias. Las soluciones de estas tres ecuaciones diferenciales ordinarias son discutidas en el próximo curso. Por ejemplo, la ecuación (9.64) es identificada como la ecuación de asociada de Legendre en la cual la constante Q llega a ser $l(l+1)$; con l entero. Si k^2 es una constante (positiva), la ecuación (9.65) llega a ser la ecuación de Bessel esférica.

Nuevamente, nuestra solución más general puede ser escrita

$$\psi_{Qm}(r, \theta, \varphi) = \sum_{q,m} R_Q(r) \Theta_{Qm}(\theta) \Phi_m(\varphi) . \quad (9.66)$$

La restricción que k^2 sea una constante es innecesariamente severa. El proceso de separación será todavía posible para k^2 tan general como

$$k^2 = f(r) + \frac{1}{r^2}g(\theta) + \frac{1}{r^2 \sin^2 \theta}h(\varphi) + k'^2 . \quad (9.67)$$

En el problema del átomo de hidrógeno, uno de los ejemplos más importantes de la ecuación de onda de Schrödinger con una forma cerrada de solución es $k^2 = f(r)$. La ecuación (9.65) para el átomo de hidrógeno llega a ser la ecuación asociada de Laguerre.

La gran importancia de esta separación de variables en coordenadas polares esféricas deriva del hecho que el caso $k^2 = k^2(r)$ cubre una tremenda cantidad de física: las teorías de gravitación, electrostática, física atómica y física nuclear. Y, con $k^2 = k^2(r)$, la dependencia angular es aislada en las ecuaciones (9.61) y (9.64), *la cual puede ser resuelta exactamente*.

Finalmente, una ilustración de cómo la constante m en la ecuación (9.61) es restringida, notamos que φ en coordenadas polares esféricas y cilíndricas es un ángulo azimutal. Si esto es un problema clásico, ciertamente requeriremos que la solución azimutal $\Phi(\varphi)$ sea univaluada, esto es,

$$\Phi(\varphi + 2\pi) = \Phi(\varphi) . \quad (9.68)$$

Esto es equivalente a requerir que la solución azimutal tenga un período de 2π o algún múltiplo entero de él. Por lo tanto m debe ser un entero. Cuál entero, depende de los detalles del problema. Cada vez que una coordenada corresponda a un eje de translación o a un ángulo azimutal la ecuación separada siempre tendrá la forma

$$\frac{d^2\Phi(\varphi)}{d\varphi^2} = -m^2\Phi(\varphi)$$

para φ , el ángulo azimutal, y

$$\frac{d^2Z}{dz^2} = \pm a^2 Z(z) \quad (9.69)$$

para z , un eje de translación en un sistema de coordenadas cilíndrico. Las soluciones, por supuesto, son $\sin az$ y $\cos az$ para $-a^2$ y la correspondiente función hiperbólica (o exponencial) $\sinh az$ y $\cosh az$ para $+a^2$.

Otras ecuaciones diferenciales ordinarias encontradas ocasionalmente incluyen las ecuaciones de Laguerre y la asociada de Laguerre del importante problema del átomo de hidrógeno en mecánica cuántica:

$$x \frac{d^2y}{dx^2} + (1-x) \frac{dy}{dx} + \alpha y = 0 , \quad (9.70)$$

$$x \frac{d^2 y}{dx^2} + (1 + k - x) \frac{dy}{dx} + \alpha y = 0 . \tag{9.71}$$

De la teoría de la mecánica cuántica del oscilador armónico lineal tenemos la ecuación de Hermite,

$$\frac{d^2 y}{dx^2} - 2x \frac{dy}{dx} + 2\alpha y = 0 . \tag{9.72}$$

Finalmente, de vez en vez encontramos la ecuación diferencial de Chebyshev

$$(1 - x^2) \frac{d^2 y}{dx^2} - x \frac{dy}{dx} + n^2 y = 0 . \tag{9.73}$$

Para una referencia conveniente, las formas de la solución de la ecuación de Laplace, la ecuación de Helmholtz y la ecuación de difusión en coordenadas polares esféricas son resumidas en la tabla 9.2. Las soluciones de la ecuación de Laplace en coordenadas circulares cilíndricas son representadas en la tabla 9.3.

$\psi = \sum_{l,m} a_{lm} \psi_{lm}$		
1.	$\nabla^2 \psi = 0$	$\psi_{lm} = \left\{ \begin{array}{l} r^l \\ r^{-l-1} \end{array} \right\} \left\{ \begin{array}{l} P_l^m(\cos \theta) \\ Q_l^m(\cos \theta) \end{array} \right\} \left\{ \begin{array}{l} \cos m\varphi \\ \text{sen } m\varphi \end{array} \right\}$
2.	$\nabla^2 \psi + k^2 \psi = 0$	$\psi_{lm} = \left\{ \begin{array}{l} j_l(kr) \\ n_l(kr) \end{array} \right\} \left\{ \begin{array}{l} P_l^m(\cos \theta) \\ Q_l^m(\cos \theta) \end{array} \right\} \left\{ \begin{array}{l} \cos m\varphi \\ \text{sen } m\varphi \end{array} \right\}$
3.	$\nabla^2 \psi - k^2 \psi = 0$	$\psi_{lm} = \left\{ \begin{array}{l} i_l(kr) \\ k_l(kr) \end{array} \right\} \left\{ \begin{array}{l} P_l^m(\cos \theta) \\ Q_l^m(\cos \theta) \end{array} \right\} \left\{ \begin{array}{l} \cos m\varphi \\ \text{sen } m\varphi \end{array} \right\}$

Cuadro 9.2: Soluciones en coordenadas polares esféricas

$\psi = \sum_{m,\alpha} a_{m\alpha} \psi_{m\alpha} , \quad \nabla^2 \psi = 0$		
a.		$\psi_{m\alpha} = \left\{ \begin{array}{l} J_m(\alpha\rho) \\ N_m(\alpha\rho) \end{array} \right\} \left\{ \begin{array}{l} \cos m\varphi \\ \text{sen } m\varphi \end{array} \right\} \left\{ \begin{array}{l} e^{-\alpha z} \\ e^{\alpha z} \end{array} \right\}$
b.		$\psi_{m\alpha} = \left\{ \begin{array}{l} I_m(\alpha\rho) \\ K_m(\alpha\rho) \end{array} \right\} \left\{ \begin{array}{l} \cos m\varphi \\ \text{sen } m\varphi \end{array} \right\} \left\{ \begin{array}{l} \cos \alpha z \\ \text{sen } \alpha z \end{array} \right\}$
c.	$\alpha = 0$ (no hay dependencia en z)	$\psi_m = \left\{ \begin{array}{l} \rho^m \\ \rho^{-m} \end{array} \right\} \left\{ \begin{array}{l} \cos m\varphi \\ \text{sen } m\varphi \end{array} \right\}$

Cuadro 9.3: Soluciones en coordenadas cilíndricas circulares

Para las ecuaciones de Helmholtz y de difusión la constante $\pm k^2$ se agrega a la constante de separación $\pm\alpha^2$ para definir un nuevo parámetro γ^2 o $-\gamma^2$. Para la elección de $+\gamma^2$ (con $\gamma^2 > 0$) obtenemos $J_m(\gamma\rho)$ y $N_m(\gamma\rho)$. Para la elección $-\gamma^2$ (con $\gamma^2 > 0$) obtenemos $I_m(\gamma\rho)$ y $K_m(\gamma\rho)$ como previamente.

Estas ecuaciones diferenciales ordinarias y sus generalizaciones serán examinadas y sistematizadas en el próximo curso.