

FÍSICA MATEMÁTICA.

Grupos de Lie, rotaciones, unitarios, Poincaré. Monte Carlo.

L. L. Salcedo
Departamento de Física Atómica, Molecular y Nuclear,
Universidad de Granada, E-18071 Granada, Spain
E-mail: salcedo@ugr.es

14 de mayo de 2019

Resumen

Apuntes de la asignatura. Versión v3.11, 2014-2019.
Se ruega comunicar los errores que puedan encontrarse a salcedo@ugr.es
<http://www.ugr.es/local/salcedo/public/fm/curso.pdf>

Índice

1. Grupo de traslaciones en \mathbb{R}^n	7
2. Generalidades sobre grupos de Lie	10
2.1. Grupos de Lie	10
2.2. Ley de composición	12

2.3.	Constantes de estructura	14
2.4.	Elementos infinitesimales	15
2.5.	Representación fiel	15
2.6.	Generadores infinitesimales. Álgebra del grupo.	16
2.7.	Coordenadas canónicas	17
2.8.	Fórmula de Campbell-Hausdorff	19
2.9.	Álgebra de Lie	20
2.10.	Relación entre grupo de Lie y álgebra	23
2.11.	Medida invariante	23
2.12.	Representación adjunta del grupo	25
2.13.	Acción de un grupo de Lie sobre una variedad	25
2.13.1.	Representación escalar	26
2.13.2.	Generadores infinitesimales	26
2.13.3.	Representaciones espinoriales	27
3.	Grupo de rotaciones	28
3.1.	Grupo de rotaciones en \mathbb{R}^n	28
3.2.	Álgebra de Lie de $SO(n)$	30
3.3.	Grupo $SO(2)$	30
3.4.	Grupo $SO(3)$	32
3.4.1.	Ángulos de Euler	34

3.4.2.	Generadores infinitesimales y álgebra de $SO(3)$	35
3.4.3.	Operador momento angular orbital	37
3.4.4.	Operadores escalares y vectoriales	39
3.5.	Grupo $SU(2)$	40
3.5.1.	Grupos $U(n)$ y $SU(n)$	40
3.5.2.	Matrices de $SU(2)$ y matrices de Pauli	42
3.5.3.	Álgebra de $SU(2)$	43
3.5.4.	Relación entre los grupos $SU(2)$ y $SO(3)$	44
3.5.5.	Representaciones enteras y semienteras de $SU(2)$	46
3.6.	Representaciones irreducibles de $SU(2)$ y $SO(3)$	47
3.6.1.	Soluciones del álgebra de $SU(2)$	47
3.6.2.	Matrices de las representaciones irreducibles de $SU(2)$ y $SO(3)$	53
3.7.	Armónicos esféricos	53
3.8.	Serie de Clebsch-Gordan de $SU(2)$	55
3.8.1.	Suma de momentos angulares	55
3.8.2.	Coeficientes de Clebsch-Gordan	57
4.	Grupo de Poincaré	61
4.1.	Transformaciones de Lorentz	61
4.2.	Grupo de Lorentz	63
4.3.	Métricas y aplicaciones lineales	64

4.4.	Estructura del grupo de Lorentz	66
4.5.	Grupo de Poincaré	68
4.5.1.	Tipos de intervalo	68
4.6.	Álgebra de Lie del grupo de Poincaré	69
4.6.1.	Álgebra de Lorentz	69
4.6.2.	Álgebra de Poincaré	71
4.7.	Álgebra de Poincaré en la base espacio-temporal	73
4.8.	Representaciones irreducibles del grupo de Lorentz	75
4.9.	Representaciones irreducibles del grupo de Poincaré	76
5.	Representaciones de $SU(n)$	81
5.1.	Representación tensorial de $GL(n, \mathbb{C})$	81
5.2.	Reducción de V_n^r bajo S_r y $GL(n, \mathbb{C})$	82
5.3.	Serie de Clebsch-Gordan para irreps tensoriales de $GL(n, \mathbb{C})$	87
5.4.	Reducibilidad de representaciones tensoriales bajo $SU(n)$	89
5.5.	Otras representaciones de $GL(n, \mathbb{C})$	90
5.6.	Representaciones de tipo T_s^r de $GL(n, \mathbb{C})$ y $U(n)$	91
5.7.	Representaciones irreducibles de $SU(n)$	94
5.8.	Matrices de Gell-Mann	99
6.	Método Monte Carlo	101
6.1.	Introducción	101

6.1.1.	Ejemplo de cálculo Monte Carlo	101
6.1.2.	Monte Carlo y valores esperados	102
6.1.3.	Fluctuación en estimaciones Monte Carlo	102
6.1.4.	Estimación Monte Carlo de integrales	105
6.2.	Probabilidad. Variables aleatorias.	110
6.2.1.	Probabilidades	110
6.2.2.	Deltas de Dirac	112
6.2.3.	Variables aleatorias	113
6.3.	Método Monte Carlo	119
6.3.1.	Promedios pesados	122
6.4.	Métodos de muestreo	124
6.4.1.	Números pseudo aleatorios	125
6.4.2.	Distribuciones de variables discretas	126
6.4.3.	Distribuciones de variables continuas	130
6.4.4.	Método de inversión	130
6.4.5.	Distribución normal	133
6.4.6.	Puntos uniformemente distribuidos en una región	133
6.4.7.	Método de aceptación-rechazo	136
6.4.8.	Método de reweighting	138
6.5.	Métodos markovianos	141
6.5.1.	Condición de balance detallado	142

6.5.2. Algoritmo de Metropolis	144
6.5.3. Baño térmico	149
6.5.4. Método de búsqueda de mínimos por enfriamiento	150
6.5.5. Termalización y tiempo de autocorrelación	150
7. Problemas resueltos	156
7.1. Relación 1: Grupos de Lie	156
7.2. Relación 2: Grupo de rotaciones	161
7.3. Relación 3: Grupo de rotaciones	166
7.4. Relación 1 de problemas laboriosos.	171
7.5. Relación 4: Monte Carlo	175

1. Grupo de traslaciones en \mathbb{R}^n

El grupo de traslaciones está formado por aplicaciones de \mathbb{R}^n en \mathbb{R}^n , con la composición de aplicaciones, del tipo:

$$G = \{T_a, a \in \mathbb{R}^n\}, \quad T_a: \mathbb{R}^n \longrightarrow \mathbb{R}^n$$

$$x \longmapsto x + a \quad (1.1)$$

Aquí $a = (a^1, a^2, \dots, a^n)$ forman un **sistema de coordenadas** de G . La dimensión del grupo es n , el número de coordenadas. $\dim G = n$.

Ley de composición:

$$x \xrightarrow{a_2} x + a_2 \xrightarrow{a_1} (x + a_2) + a_1 = x + (a_1 + a_2) \quad (1.2)$$

por tanto la ley de composición de este grupo en estas coordenadas es

$$a_{12} = a_1 + a_2. \quad (1.3)$$

Se trata de un grupo abeliano $T_{a_1}T_{a_2} = T_{a_2}T_{a_1}$.

Representación de las traslaciones en $L^2(\mathbb{R}^n)$:

$$\psi(x) \mapsto (U(a)\psi)(x) = \psi(x - a). \quad (1.4)$$

Esto es una representación ya que $U(a_{12}) = U(a_1)U(a_2)$. Además es **unitaria**: es una isometría

$$\langle U(a)\psi_1 | U(a)\psi_2 \rangle = \int d^n x \psi_1^*(x - a)\psi_2(x - a) = \int d^n x \psi_1^*(x)\psi_2(x) = \langle \psi_1 | \psi_2 \rangle, \quad (1.5)$$

y es invertible: $U(a)^{-1} = U(-a)$, como se deduce de la ley de multiplicación. Esto implica $U^{-1}(a) = U^\dagger(a)$.

Esta representación coincide con la representación regular del grupo.¹

Por otro lado, para una **traslación infinitesimal** δa

$$\psi(x - \delta a) = \psi(x) - \delta a^i \partial_i \psi(x). \quad (1.6)$$

¹La representación regular por la izquierda se define como $(T^L(g)\psi)(g') = \psi(g^{-1}g')$ en el espacio de funciones definidas sobre el grupo, $\mathcal{C}(G)$, y se les puede dar estructura de espacio de Hilbert, $L^2(G, d\mu)$.

Aquí $\partial_i = \partial/\partial x^i$ y despreciamos términos de segundo orden $O(\delta a^2)$. La variación puede escribirse

$$\delta\psi(x) = -\delta a \cdot \nabla\psi(x) \equiv \frac{1}{i\hbar}\delta a \cdot P\psi(x) \quad (1.7)$$

donde se ha introducido el **operador momento**

$$P^i = -i\hbar\partial_i, \quad P = -i\hbar\nabla = P^\dagger \quad (1.8)$$

que es el **generador infinitesimal** de las traslaciones. El $1/(i\hbar)$ es convencional, se introduce para que P sea hermítico y con dimensiones de momento lineal. Los operadores $-i\hbar\partial_i$ son los generadores infinitesimales de las traslaciones en esta representación del grupo.

Relaciones de conmutación. Las derivadas conmutan

$$\partial_i\partial_j\psi(x) = \partial_j\partial_i\psi(x) \quad (1.9)$$

y esto implica las relaciones de conmutación entre generadores

$$[P^i, P^j] = 0. \quad (1.10)$$

Que los generadores conmuten es consecuencia directa de que el grupo es abeliano.

Para transformaciones finitas

$$\begin{aligned} \psi(x-a) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} a^{i_1} \dots a^{i_n} \partial_{i_1} \dots \partial_{i_n} \psi(x) \quad (\text{Taylor}) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} (-a \cdot \nabla)^n \psi(x) = e^{-a \cdot \nabla} \psi(x). \end{aligned} \quad (1.11)$$

Por tanto el operador que representa a las traslaciones en $L^2(\mathbb{R}^n)$ es

$$U(a) = e^{-\frac{i}{\hbar}a \cdot P}. \quad (1.12)$$

Esto es general: exponenciando los generadores infinitesimales se obtiene el operador del grupo (para grupos conexos). También se tiene la relación entre generadores hermíticos y representación unitaria

$$P = P^\dagger \iff U(a)^\dagger = U(a)^{-1}. \quad (1.13)$$

Tomando a infinitesimal se recuperan los generadores a partir de los operadores del grupo

$$U(a) = 1 - \frac{i}{\hbar}a \cdot P + O(a^2), \quad P^i = i\hbar \left. \frac{\partial U(a)}{\partial a^i} \right|_0. \quad (1.14)$$

Representaciones irreducibles.² Como el grupo es abeliano sus representaciones irreducibles (o simplemente **irreps**) son unidimensionales, por el lema de Schur.³ Si $D(a)$ (matriz compleja 1×1) es una tal representación

$$D(a_1)D(a_2) = D(a_1 + a_2). \quad (1.15)$$

Las únicas soluciones (con el requerimiento de que sean diferenciables con respecto de a)⁴ son de la forma

$$D(a) = e^{\mu \cdot a}, \quad \mu \in \mathbb{C}^n. \quad (1.16)$$

En efecto, aplicando $\nabla_{a_2}|_0$ a la relación (1.15), se tiene

$$D(a_1)D'(0) = D'(a_1), \quad D'(0) = \nabla \log D(a). \quad (1.17)$$

($D(a) \neq 0$ por ser $D(a)$ invertible.) Llamando μ a $D'(0)$ se obtiene la solución dada. El vector μ determina completamente todas las irreps inequivalentes (en espacios unidimensionales representaciones distintas son inequivalentes) del grupo de traslaciones. En particular

$$e^{-i\frac{a \cdot p}{\hbar}}, \quad p \in \mathbb{R}^n \quad (1.18)$$

son las representaciones unitarias y están contenidas en la representación regular, $L^2(\mathbb{R}^n)$. Si $\psi_p(x)$ y sus múltiplos forman un subespacio unidimensional en la representación p ,

$$\psi_p(x - a) = (U(a)\psi_p)(x) = e^{-i\frac{a \cdot p}{\hbar}} \psi_p(x). \quad (1.19)$$

Tomando $a = x$ se obtiene (salvo normalización)

$$\psi_p(x) = e^{i\frac{p \cdot x}{\hbar}} \quad (\text{onda plana}). \quad (1.20)$$

Es decir, una partícula con momento p cae en la irrep p del grupo de traslaciones y esto nos dice cómo responde su función de onda a una traslación.

En la reducción de $L^2(\mathbb{R}^n)$ bajo traslaciones cada irrep $p \in \mathbb{R}^n$ aparece exactamente una vez

$$\begin{aligned} \mathcal{H} = L^2(\mathbb{R}^n) &= \bigoplus_p V_p, \quad V_p = \{\lambda e^{i\frac{p \cdot x}{\hbar}}, \lambda \in \mathbb{C}\} \\ \psi(x) &= \int \frac{d^n p}{(2\pi\hbar)^n} \tilde{\psi}(p) e^{i\frac{p \cdot x}{\hbar}} \quad (\text{Transformada de Fourier}) \end{aligned} \quad (1.21)$$

Las otras irreps con μ complejo aparecen también en el espacio $\mathcal{C}(\mathbb{R}^n)$ de funciones complejas $f(x)$, $x \in \mathbb{R}^n$ pero no en su restricción a $L^2(\mathbb{R}^n)$.

²Cuando no se diga otra cosa la irreducibilidad se entenderá en el sentido de representaciones complejas. Representaciones irreducibles reales pueden ser reducibles complejas.

³Para un grupo abeliano $[U(g_1), U(g_2)] = 0$ en una irrep implica $U(g) = \lambda(g)1_d$ (identidad en \mathbb{C}^d) y $d = 1$ por ser irreducible.

⁴De hecho hay otras soluciones no continuas de la ecuación pero sólo buscamos representaciones continuas.

2. Generalidades sobre grupos de Lie

2.1. Grupos de Lie

Un **grupo continuo** G es un conjunto que es grupo en sentido algebraico y al mismo tiempo un espacio topológico tal que la aplicación de $G \times G \rightarrow G$ definida por $(g_1, g_2) \mapsto g_1 g_2^{-1}$ es continua.⁵

Se dice que dos espacios topológicos son **homeomorfos** si existe una aplicación biyectiva y bicontinua entre ellos, y en este caso se consideran topológicamente equivalentes. Dos grupos continuos son **isomorfos** si lo son como grupos en sentido algebraico y además son homeomorfos como espacios topológicos.

Un espacio topológico X que sea localmente euclídeo es una **variedad topológica**. Esto quiere decir que admite un atlas de coordenadas locales. Cada carta local del atlas es un abierto de X homeomorfo a un abierto de \mathbb{R}^n , siendo n la dimensión de la variedad topológica.

Un grupo continuo localmente euclídeo es un **grupo de Lie**. La dimensión del grupo de Lie es su dimensión como variedad topológica, es decir, el número de coordenadas necesario para especificar un elemento del grupo.

Propiedades topológicas relevantes de un grupo de Lie son su dimensión, y si se trata o no de un conjunto compacto, o conexo o simplemente conexo.

Ejemplo. El grupo $(\mathbb{R}^2, +)$ tiene dimensión 2, es homeomorfo a (tiene la misma topología que) un plano, por tanto es conexo y simplemente conexo, pero no compacto. Basta una sola carta para cubrir todo el grupo.

Ejemplo. $(U(1), \cdot)$ es el grupo de las fases con el producto

$$U(1) = \{\omega \mid \omega \in \mathbb{C}, |\omega| = 1\}, \quad \omega = e^{i\theta}, \quad \theta \in \mathbb{R}, \quad \omega_{12} = \omega_1 \omega_2. \quad (2.1)$$

Este grupo tiene la topología de una circunferencia, S^1 . Este grupo tiene dimensión 1, es compacto, conexo pero no simplemente conexo. Los elementos se pueden parametrizar con $\theta \in]-\pi, \pi]$ con $\omega = e^{i\theta}$, pero una carta local es un **conjunto abierto**, entonces se puede usar $|\theta| < \pi$ como sistema de coordenadas que incluye al elemento neutro ($\omega = 1$) pero hace falta otra carta que incluya al elemento $\omega = -1$ (por ejemplo $\omega = -e^{i\theta'}$ con $|\theta'| < \pi$).

⁵Ver por ejemplo [4] para definiciones precisas de espacio topológico, continuidad, conjunto compacto, variedad topológica, etc.

Ejemplo. $O(2)$ es el grupo de matrices 2×2 reales y ortogonales

$$O(2) = \left\{ \begin{pmatrix} \cos \theta & -\eta \operatorname{sen} \theta \\ \operatorname{sen} \theta & \eta \cos \theta \end{pmatrix}, 0 \leq \theta < 2\pi, \eta = \pm 1 \right\}. \quad (2.2)$$

Este grupo tiene dimensión 1 y es compacto pero no es conexo: tiene dos componentes conexas ($\eta = \pm 1$) cada una con topología S^1 . Las matrices con $\eta = +1$ forman el grupo $SO(2)$, que es un subgrupo invariante de $O(2)$. Esto es general, en un grupo de Lie G , la componente conexa del neutro, G_c , es un subgrupo invariante y las demás componentes son las clases de equivalencia de G/G_c y son homeomorfas a G_c (ver [4]).

Ejemplo. $GL(1, \mathbb{R})$, el grupo de cambios de base reales en una dimensión, tiene dos componentes conexas.

Teorema. Para grupos **compactos** (lo cual incluye a los grupos finitos) se tiene⁶

- a) Sus representaciones son equivalentes a unitarias y completamente reducibles.
- b) Sus irreps son de dimensión finita.
- c) En la reducción de la representación regular, cada irrep aparece tantas veces como sea su dimensión.

Que el espacio sea **simplemente conexo** significa que, dados dos puntos cualesquiera, todos los caminos que los unen son deformables unos en otros, es decir, esencialmente sólo hay un modo de ir de un punto a otro. Equivalentemente, todo camino cerrado es contráctil a un punto.

Ejemplo. \mathbb{R}^n es simplemente conexo.

Ejemplo. El grupo $GL(1, \mathbb{C})$ de cambios de base complejos en una dimensión. Cada elemento es un punto de $\mathbb{C} \setminus \{0\}$. Este grupo es conexo pero no simplemente conexo: los caminos cerrados (que no pueden pasar por 0) pueden dar $n \in \mathbb{Z}$ vueltas alrededor de 0 y caminos con distinto valor de n no son deformables entre sí. Igualmente las clases de caminos de $U(1)$ (la circunferencia S^1), y de $\mathbb{R} \times U(1)$ (el cilindro $\mathbb{R} \times S^1$) se clasifican por un número entero. Las clases de $U(1) \times U(1)$ (el toro $S^1 \times S^1$) se clasifican por dos enteros, \mathbb{Z}^2 .

⁶Un teorema relacionado: si G es conexo, simple y *no compacto*, su única representación unitaria de dimensión finita es la trivial. Para un grupo de Lie, **simple** quiere decir que es no abeliano y que sus únicos subgrupos invariantes propios son discretos.

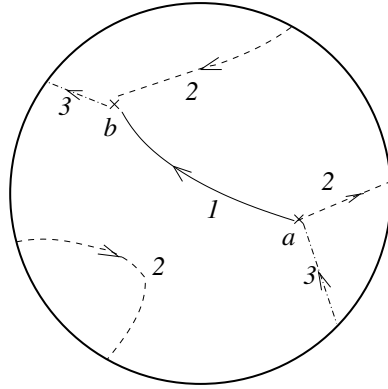


Figura 1: \bar{D}_2/\sim : Disco con los puntos del borde diametralmente opuestos identificados. Los caminos 1 y 2 de a a b son de la misma clase de homotopía (es decir, deformables uno en otro) y 3 de la otra clase.

Ejemplo. El disco abierto $D_2 = \{x \in \mathbb{R}^2, \|x\| < 1\}$ es simplemente conexo. En cambio, si añadimos los puntos del borde e identificamos los que sean diametralmente opuestos, obtenemos

$$\bar{D}_2/\sim = \{x \in \mathbb{R}^2, \|x\| \leq 1, \text{ con } x \sim -x \text{ si } \|x\| = 1\}, \quad (2.3)$$

que es un espacio doblemente conexo: hay dos tipos de caminos cerrados, según que utilicen $x \sim -x$ un número par o impar de veces.

La relevancia es que los espacios simplemente conexos no tienen funciones multivaluadas, pero los múltiplemente conexos sí. Ej: en $U(1)$, $\omega \rightarrow \omega^{1/2}$ (representación unidimensional) es bivaluada.

2.2. Ley de composición

Sea G un grupo de Lie de dimensión n . Se puede elegir un sistema de coordenadas locales que contenga al elemento neutro e . Esto quiere decir que hay un homeomorfismo entre un entorno abierto del grupo, $U_G \ni e$, y un entorno abierto U de \mathbb{R}^n ,

$$g \in U_G \leftrightarrow a = (a^1, a^2, \dots, a^n) \in U. \quad (2.4)$$

Se suele elegir el sistema de coordenadas de modo que el neutro, e , tiene coordenadas $(0, 0, \dots, 0)$.

Si g_1 , g_2 y $g_{12} = g_1 g_2$ están en U_G , con coordenadas a_1 , a_2 y a_{12} , respectivamente, la función

$$a_{12} = f(a_1, a_2) \quad (2.5)$$

se denomina **ley de composición** de G en esas coordenadas. Más explícitamente

$$a_{12}^i = f^i(a_1^1, \dots, a_1^n, a_2^1, \dots, a_2^n), \quad i = 1, \dots, n. \quad (2.6)$$

La ley de composición es una función continua y cumple

$$f(a, f(b, c)) = f(f(a, b), c), \quad a = f(a, 0) = f(0, a), \quad (2.7)$$

correspondientes a propiedad asociativa y existencia de neutro. Existencia de inverso requiere que $0 = f(a, b)$ sea invertible para expresar b en función de a .

Obviamente la ley de composición depende del sistema de coordenadas. Aunque no se ha impuesto, las restricciones sobre f por ser un grupo garantizan que se puede elegir el sistema de coordenadas de modo que f sea **analítica** (real). Elegimos coordenadas analíticas (no únicas) en lo que sigue.

Dos grupos de Lie G y G' son **localmente isomorfos** si admiten la misma ley de composición en sendos entornos del neutro U_G y $U_{G'}$. En general dos grupos localmente isomorfos pueden no ser **globalmente** isomorfos (es decir, no ser realmente isomorfos como grupos de Lie).

Ejemplo. En $U(1)$ $\omega = e^{i\theta}$, $\omega_{12} = \omega_1 \omega_2$, $\theta_{12} = \theta_1 + \theta_2$, y para $(\mathbb{R}, +)$, $x_{12} = x_1 + x_2$. Los dos grupos tienen la misma ley de composición pero no son isomorfos. Por ej. $(e^{i2\pi/3})^3 = 1$ en cambio $3 \times 2\pi/3 \neq 0$ (no son algebraicamente isomorfos). Y tampoco son topológicamente iguales: $U(1)$ es compacto y no simplemente conexo y \mathbb{R} es no compacto y simplemente conexo.

Teorema. Dado un grupo de Lie G , en la clase de equivalencia de los grupos localmente isomorfos a G hay exactamente uno (módulo isomorfismos), \tilde{G} , que es conexo y simplemente conexo, denominado **recubridor universal** de la clase de equivalencia. Además si G es conexo, $G \cong \tilde{G}/N$ donde N es un subgrupo invariante discreto de \tilde{G} .⁷ Si N tiene n elementos, G será un espacio n -conexo.

Ejemplo. \mathbb{R} y $U(1)$ son localmente isomorfos y \mathbb{R} es simplemente conexo, por lo cual es el grupo recubridor y $U(1)$ es un grupo cociente. En efecto, $U(1) \cong \mathbb{R}/(2\pi\mathbb{Z})$ ya que $x \bmod 2\pi$ es una variable angular.

Cualquier representación (univaluada) de \tilde{G} define una representación de G que será también univaluada en un entorno del neutro, sin embargo generalmente será multivaluada para G a nivel

⁷Usamos \cong para indicar isomorfismo.

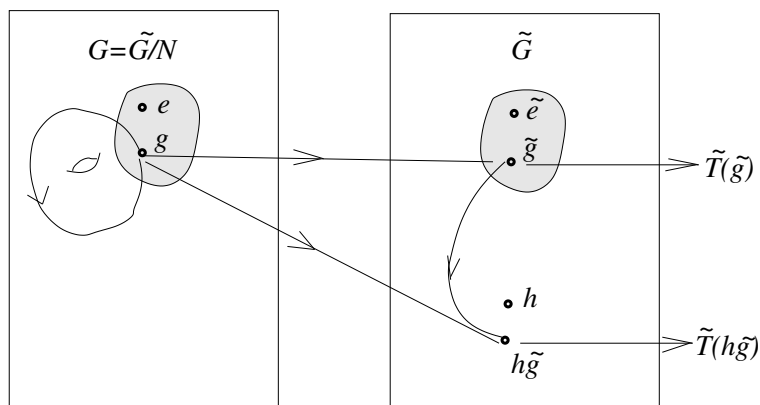


Figura 2: $G = \tilde{G}/N$ donde N es un subgrupo discreto. G y \tilde{G} son localmente isomorfos. Al tomar cociente el camino $\tilde{g} \rightarrow h\tilde{g}$ ($h \in N$) es abierto en \tilde{G} pero cerrado y no contráctil en G que es no simplemente conexo. Una representación \tilde{T} de \tilde{G} en general será multivaluada como representación de G .

global. En efecto, si $D(a)$ son los operadores de la representación del recubridor \tilde{G} , y $h \in N$ ($G \cong \tilde{G}/N$), siempre que $D(h) \neq 1$ se tendrá una multivaluación de D como representación de G ya que $D(e) = 1$ pero $h \equiv e$ en G .

2.3. Constantes de estructura

Dada la ley composición⁸ (usando la segunda ec. (2.7))

$$f^k(x, y) = x^k + y^k + \frac{\partial^2 f^k}{\partial x^i \partial y^j} \Big|_0 x^i y^j + R_3 \quad (2.8)$$

(R_3 representa términos cúbicos o más en el desarrollo en serie.) Los números

$$c_{ij}^k = \frac{\partial^2 f^k}{\partial x^i \partial y^j} \Big|_0 - \frac{\partial^2 f^k}{\partial x^j \partial y^i} \Big|_0 \quad (2.9)$$

se denominan **constantes de estructura** del grupo (en las coordenadas x). Se deduce que se anulan si el grupo es abeliano.

Bajo cambios de coordenadas $x'^i(x)$ las constantes de estructura se transforman como un tensor

⁸**Importante:** Aquí, y a menudo en lo que sigue, usamos el convenio de índices repetidos. Así por ejemplo $B_i = A_{ijk} x^j y^k$ quiere decir $B_i = \sum_{j,k} A_{ijk} x^j y^k$.

(a diferencia de $\frac{\partial^2 f^k}{\partial x^i \partial y^j} \Big|_0$):

$$x^i = A^i_j x'^j + R_2, \quad A^i_j = \frac{\partial x^i}{\partial x'^j} \Big|_0, \quad c_{ij}^k = c'_{lm} A^l_i A^m_j (A^{-1})^k_n. \quad (2.10)$$

Propiedades:

- 1) Las c_{ij}^k son reales.
- 2) $c_{ij}^k = -c_{ji}^k$ (antisimetría). (2.11)
- 3) $c_{ir}^l c_{jk}^r + c_{jr}^l c_{ki}^r + c_{kr}^l c_{ij}^r = 0$ (**identidad de Jacobi**).

La identidad de Jacobi es consecuencia de la propiedad asociativa.

2.4. Elementos infinitesimales

Los elementos infinitesimales son elementos del grupo con coordenadas (analíticas) infinitesimales (de primer orden), δa^i . En particular, usando la ec. (2.8),

$$f(\delta a, -\delta a) = 0, \quad (2.12)$$

ya que despreciamos $O(\delta^2)$. Esto implica que si un elemento infinitesimal tiene coordenadas δa^i , su inverso tiene coordenadas $-\delta a^i$.⁹

2.5. Representación fiel

A menudo es conveniente trabajar con operadores (o matrices) que representen al grupo en lugar de con el grupo directamente. A $g \in G$ la representación le asocia un operador invertible $T(g)$ de $GL(V)$, siendo V un espacio vectorial (o de Hilbert), de modo que $T(g_1)T(g_2) = T(g_1g_2)$, o en coordenadas, $T(a)T(b) = T(f(a,b))$ (usamos $T(a)$ para el operador $T(g)$ siendo a las coordenadas de g .)

La representación $T(g)$ es **fiel** si es inyectiva (es decir, elementos distintos se representan por operadores distintos) y en este caso hay un isomorfismo entre los grupos G y $T(G) = \{T(g), g \in G\}$.

Una representación fiel concreta es la **representación regular por la izquierda** T^L . En este caso $V = \mathcal{C}(G)$ es el espacio de funciones complejas continuas definidas sobre el grupo, $\psi(g) \in \mathbb{C}$, y T^L se define mediante

$$(T^L(g)\psi)(g') = \psi(g^{-1}g'), \quad T^L(g_1)T^L(g_2) = T^L(g_1g_2). \quad (2.13)$$

⁹Para elementos finitos, el inverso coincide con $-a^i$ sólo en ciertos sistemas de coordenadas.

2.6. Generadores infinitesimales. Álgebra del grupo.

$$T(a) = 1 + \left. \frac{\partial T}{\partial a^j} \right|_0 a^j + O(a^2) \quad (2.14)$$

Aquí, 1 representa el operador identidad en el espacio V en el que actúa la representación. El operador

$$X_j := i \left. \frac{\partial T}{\partial a^j} \right|_0 \quad j = 1, \dots, n \quad (2.15)$$

es el **generador infinitesimal** en la dirección j . La unidad imaginaria i se introduce por conveniencia (también se puede introducir un \hbar pero no es usual en este contexto). Por tanto, para un elemento infinitesimal

$$T(\delta a) = 1 - i\delta a^i X_i \equiv 1 - i\delta a \cdot X. \quad (2.16)$$

Los X_i son n operadores que actúan en V , igual que $T(a)$. Si se trata de una **representación unitaria**,

$$T(a)^{-1} = T(a)^\dagger, \quad (2.17)$$

los X_i son hermíticos:

$$\forall \delta a \quad 1 + i\delta a^i X_i^\dagger = T(\delta a)^\dagger = T(\delta a)^{-1} = T(-\delta a) = 1 + i\delta a^i X_i \quad \Rightarrow \quad X_i = X_i^\dagger. \quad (2.18)$$

(Para esto se introduce la i en la definición de generador.)

Los X_i subtienden un espacio vectorial real¹⁰

$$L := \text{lin}_{\mathbb{R}} \{X_i, i = 1, \dots, n\} = \{a^i X_i, a \in \mathbb{R}^n\}. \quad (2.19)$$

Si la representación es fiel, los X_i son linealmente independientes y forman una base. En efecto, si $\delta a^i X_i = 0$, se tendrá $T(\delta a) = 1$ y entonces $\delta a = 0$ por ser a un sistema de coordenadas y T fiel. En este caso, $\dim L = n$.

El espacio L concreto depende de la representación. El espacio asociado a la representación regular, T^L , se denomina **álgebra del grupo** G . Por extensión, el mismo nombre se aplica a los L asociados a representaciones fieles.

¹⁰Esto quiere decir, que tomamos combinaciones lineales reales de los generadores, pero V es un espacio vectorial complejo y los X_i mismos pueden ser matrices complejas, por ejemplo.

La base X_i depende de las coordenadas. Bajo un cambio de coordenadas $a'^i(a)$

$$X'_i = i \frac{\partial T}{\partial a'^i} \Big|_0 = i \frac{\partial a^j}{\partial a'^i} \Big|_0 \frac{\partial T}{\partial a^j} \Big|_0 = A^j{}_i X_j. \quad (2.20)$$

Esto implica que el cambio de coordenadas induce un cambio de base, pero el espacio L mismo no depende de las coordenadas. Por supuesto hay muchos sistemas de coordenadas con los mismos generadores. Los generadores serán los mismos cuando los dos sistemas coincidan hasta primer orden en un entorno del neutro, es decir, $a'^i = a^i + O(a^2)$.

Para estados y operadores cuánticos un grupo de transformaciones actúa según

$$|\psi\rangle \mapsto |\psi^g\rangle = T(g)|\psi\rangle, \quad A \mapsto A^g = T(g)AT(g)^{-1}, \quad (2.21)$$

de modo que $(A|\psi\rangle)^g = A^g|\psi^g\rangle$. (Esta acción define la representación adjunta de T .) Entonces, para la variación infinitesimal se tiene, usando $T(g) = 1 - i\delta X$,

$$\delta|\psi\rangle = -i\delta X|\psi\rangle, \quad \delta A = -i[\delta X, A], \quad \delta X \equiv \delta a^i X_i. \quad (2.22)$$

Se deduce que si un observable A es invariante bajo la acción de un grupo dicho observable conmuta con los generadores infinitesimales.

2.7. Coordenadas canónicas

La exponencial de un operador A se define por su desarrollo en serie

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n. \quad (2.23)$$

Para operadores acotados esta serie es convergente en todo el espacio.

En general si $f(x) = \sum c_n x^n$, se define $f(A) = \sum c_n A^n$, o equivalentemente, si $\{|a\rangle\}$ es una base propia, $A|a\rangle = a|a\rangle$, se define $f(A)|a\rangle = f(a)|a\rangle$. En todo caso se deduce la propiedad $Sf(A)S^{-1} = f(SAS^{-1})$ que se utilizará más adelante.

Sea G un grupo *conexo*. Por conveniencia, en lugar de usar directamente el grupo, podemos usar una representación fiel cualquiera, T , de modo que el grupo $T(G)$ es isomorfo a G .

En general para un sistema de coordenadas a , no se cumple que $T(a)T(a) = T(2a)$ o más generalmente que $T(a)^N = T(Na)$ ($N \in \mathbb{Z}$), sin embargo es posible elegir las coordenadas de modo que esto sea cierto, éstas son las denominadas **coordenadas canónicas**.

Dado un sistema cualquiera de coordenadas $a \in U \subseteq \mathbb{R}^n$, con generadores X_j , veamos primero que

$$e^{-i\hat{a}^j X_j} = e^{-i\hat{a} \cdot X} \in T(G), \quad \hat{a} \in \mathbb{R}^n. \quad (2.24)$$

Partimos de $\hat{a} \in U$,

$$T(G) \ni T(\hat{a}) = 1 - i\hat{a} \cdot X + O(\hat{a}^2) = e^{-i\hat{a} \cdot X + O(\hat{a}^2)}. \quad (2.25)$$

Para $N \in \mathbb{N}$ también se tendrá¹¹

$$T(G) \ni T(\hat{a}/N)^N = \left(e^{-i\frac{1}{N}\hat{a} \cdot X + O(1/N^2)} \right)^N = e^{-i\hat{a} \cdot X + O(1/N)}, \quad (2.26)$$

donde se ha usado la identidad

$$(e^A)^N = e^{NA} \quad N \in \mathbb{Z}. \quad (2.27)$$

Tomando el límite de N grande con \hat{a} fijo

$$e^{-i\hat{a} \cdot X} \in T(G). \quad (2.28)$$

Aquí se suponía $\hat{a} \in U$ pero lo mismo vale para $\hat{a} \in \mathbb{R}^n$: basta escribir $\hat{a} = N(\hat{a}/N)$ para N suficientemente grande de modo que $\hat{a}/N \in U$ y $e^{-i\hat{a} \cdot X} \in T(G)$ por $e^{-i\frac{\hat{a}}{N} \cdot X} \in T(G)$.

Lo que acabamos de ver es que (simbólicamente) $e^{-iL} \subseteq G$. De hecho,

$$G = e^{-iL}, \quad (2.29)$$

es decir, todos los elementos de G se pueden alcanzar variando \hat{a} , y los \hat{a}^j definen un sistema de coordenadas. Un argumento (cualitativo) para ver esto es que, para un grupo de Lie conexo, cualquier entorno U_G del neutro se puede usar para reconstruir todo el grupo, y concretamente

$$\forall g \in G \quad \exists h \in U_G \quad \exists N \in \mathbb{N} \quad g = h^N. \quad (2.30)$$

De modo que basta ver que todos los elementos de U_G se pueden obtener mediante $e^{-i\hat{a} \cdot X}$. Pero esto es evidente cuando se toma U_G formado por elementos infinitesimales¹²

$$h \in U_G \quad T(U_G) \ni T(h) = 1 - i\delta\hat{a} \cdot X = e^{-i\delta\hat{a} \cdot X}. \quad (2.31)$$

Los n números \hat{a}^i forman un sistema de coordenadas locales denominadas **coordenadas canónicas** o **normales**. Todas las $T(g)$ se pueden expresar como $e^{-i\hat{a} \cdot X}$, pero sólo en un entorno

¹¹Sin pérdida de generalidad suponemos U es convexo, es decir, tal que que si $x \in U$ entonces $\lambda x \in U$, para $\lambda \leq 1$.

¹²Para $g \in G$ cualquiera, podemos elegir U_G cada vez más pequeño de modo que $g = h^N$ con un h más próximo al neutro y N mayor.

del neutro está garantizado que es una biyección. Las coordenadas canónicas no dependen de la representación fiel usada en la construcción, son una propiedad del grupo.

Nótese que en general las \hat{a}^i no coinciden con las a^i originales, $T(a) \neq \lim_{N \rightarrow \infty} T(a/N)^N$, pero $\hat{a}^i = a^i + O(a^2)$ y por tanto sí tienen los mismos generadores infinitesimales.

Las coordenadas canónicas son únicas para cada conjunto de generadores (cada base de L) y distintas coordenadas canónicas están relacionadas por transformaciones lineales. La propiedad básica de las coordenadas canónicas es que al hacer dos transformaciones en la *misma dirección* las coordenadas se suman¹³

$$e^{-i\lambda\hat{a}\cdot X} e^{-i\mu\hat{a}\cdot X} = e^{-i(\lambda+\mu)\hat{a}\cdot X}, \quad (2.32)$$

Nótese que las coordenadas *no se suman* para dos transformaciones arbitrarias $e^{-i\hat{a}\cdot X} e^{-i\hat{b}\cdot X}$, a menos que el grupo sea abeliano.

2.8. Fórmula de Campbell-Hausdorff

Una fórmula útil es

$$e^A X e^{-A} = X + [A, X] + \frac{1}{2!} [A, [A, X]] + \frac{1}{3!} [A, [A, [A, X]]] + \dots \equiv e^{[A, \cdot]} X. \quad (2.33)$$

Esta fórmula se demuestra por inducción.¹⁴

Sean A y B dos operadores y C tal que¹⁵

$$e^A e^B = e^C. \quad (2.34)$$

Desarrollando en serie ambos lados, se ve que el operador C puede escribirse como [4]

$$C = A + B + \frac{1}{2} [A, B] + \frac{1}{12} [A, [A, B]] + \frac{1}{12} [B, [B, A]] + \dots \equiv C_1 + C_2 + C_3 + \dots, \quad (2.35)$$

donde $[A, B] \equiv AB - BA$ es el conmutador de A y B . El término C_n tiene exactamente n operadores A ó B y $n - 1$ conmutadores, además A y B sólo aparecen en la forma $[A, \cdot]$, $[B, \cdot]$. Dicho de otra forma

$$C = A + f([A, \cdot], [B, \cdot])B, \quad (2.36)$$

¹³Para que esta condición garantice que \hat{a}^i son canónicas hace falta además que sean coordenadas analíticas.

¹⁴Alternativamente, si ponemos una etiqueta 1 ó 2 según que un operador A esté situado a la izquierda o la derecha de X , se tiene $e^A X e^{-A} = e^{A_1 - A_2} X$ y $A_1 - A_2$ no es más que el conmutador $[A, \cdot]$.

¹⁵Se va aplicar para matrices en un entorno de cero, la multivaluación de tipo $2\pi i n$ en C no es relevante aquí. Se elige la rama tal que $C \rightarrow 0$ cuando $A, B \rightarrow 0$.

donde la operación $[A, \cdot]$ sobre un operador X cualquiera representa $[A, X]$ (idem $[B, \cdot]$) y $f(X, Y)$ es una función conocida de argumentos X e Y para los que se supone válida la propiedad asociativa pero no la conmutativa, y f es desarrollable en serie de potencias de X e Y .

Esta fórmula es consistente con $A \rightarrow A + a1$, $B \rightarrow B + b1$, $C \rightarrow C + (a + b)1$, $a, b \in \mathbb{C}$.

2.9. Álgebra de Lie

Un **álgebra de Lie** es un álgebra con un producto de Lie, esto es, lineal, antisimétrico y que satisfaga la identidad de Jacobi. El espacio formado por todos los operadores definidos sobre un espacio vectorial forma un álgebra de Lie con el conmutador como producto.¹⁶

El espacio subtendido por los generadores infinitesimales, L , forma un álgebra de Lie, es decir, L es cerrado bajo conmutación de operadores

$$X, Y \in L \Rightarrow -i[X, Y] \in L, \quad -i[L, L] \subseteq L. \quad (2.37)$$

Para ver esto, en cualquier representación T , dado un sistema de coordenadas a y sus coordenadas canónicas asociadas \hat{a} ,

$$\begin{aligned} T(a) &= e^{-i\hat{a}^i X_i}, & T(b) &= e^{-i\hat{b}^i X_i} \\ T(a)T(b) &= T(c) = e^{-i\hat{c}^i X_i}, & c^i &= f^i(a, b). \end{aligned} \quad (2.38)$$

Por la fórmula de Campbell-Hausdorff

$$-i\hat{c}^k X_k = -i\hat{a}^k X_k - i\hat{b}^k X_k + \frac{1}{2}[-i\hat{a}^i X_i, -i\hat{b}^j X_j] + R_3 \quad (2.39)$$

y al mismo tiempo

$$-i\hat{c}^k X_k = -i \left(\hat{a}^k + \hat{b}^k + \frac{\partial^2 \hat{f}^k}{\partial \hat{a}^i \partial \hat{b}^j} \Big|_0 \hat{a}^i \hat{b}^j + R'_3 \right) X_k, \quad (2.40)$$

lo cual implica, usando la definición de las constantes de estructura en ec. (2.9),

$$-\frac{1}{2}[X_i, X_j] = -i \frac{\partial^2 \hat{f}^k}{\partial \hat{a}^i \partial \hat{b}^j} \Big|_0 X_k = -\frac{i}{2} \hat{c}_{ij}^k X_k. \quad (2.41)$$

¹⁶La identidad Jacobi en este caso se sigue de

$$0 = [A, [B, C]] + [B, [C, A]] + [C, [A, B]].$$

para tres operadores cualesquiera.

Nótese que $\frac{\partial^2 f^k}{\partial a^i \partial b^j} \Big|_0$ no es antisimétrico en general, pero sí en coordenadas canónicas. Por otro lado $\hat{c}_{ij}^k = c_{ij}^k$ ya que la matriz de cambio de base entre las coordenadas a y \hat{a} es $A^i_j = \delta^i_j$ en ec. (2.10). Finalmente se obtiene, en un sistema de coordenadas arbitrario y para una representación cualquiera

$$[X_i, X_j] = ic_{ij}^k X_k \quad (\text{relaciones de conmutación}), \quad (2.42)$$

que también pueden escribirse como

$$[-iX_i, -iX_j] = c_{ij}^k (-iX_k) \quad \text{o} \quad -i[X_i, X_j] = c_{ij}^k X_k, \quad (2.43)$$

interpretando $-i[,]$ como el producto de Lie.

Si la representación es fiel, las c_{ij}^k (que no dependen de la representación) se pueden obtener calculando $[X_i, X_j]$ por ser $\{X_k\}_{k=1}^n$ una base. Una vez conocidas las relaciones de conmutación se puede obtener el conmutador de dos elementos cualesquiera del álgebra de Lie (sin necesidad de tener los operadores mismos)

$$X, Y \in L \quad X = a^i X_i, \quad Y = b^j X_j, \quad -i[X, Y] = z^k X_k \in L, \quad z^k \equiv c_{ij}^k a^i b^j. \quad (2.44)$$

A veces se usa la notación $z = \mathbf{a} \times \mathbf{b}$. De hecho el producto vectorial usual es un caso particular como se verá.

Por definición una **representación de un álgebra de Lie** L en un espacio L' de operadores quiere decir un homomorfismo de álgebras

$$f : X \in L \mapsto f(X) \in L' \quad f([X, Y]) = [f(X), f(Y)]. \quad (2.45)$$

Puesto que el conmutador depende sólo de las constantes de estructura, se deduce que dos álgebras de Lie son **isomorfas** si y sólo si tienen las mismas constantes de estructura (en sendas bases apropiadas). Cualquier representación $T(a)$ del grupo produce una representación de su álgebra de Lie.

Acabamos de ver que si los operadores de L son los generadores infinitesimales de un grupo de Lie, entonces forman un álgebra de Lie. También se cumple el recíproco: *si un conjunto de operadores L es un álgebra de Lie, entonces su exponencial e^{-iL} es un grupo de Lie.*¹⁷

¹⁷Al menos formalmente. Si los operadores de L son no acotados (lo cual requiere un espacio de dimensión infinita) puede ocurrir que e^{-iL} o la serie de Campbell-Hausdorff no converjan. En este caso se dice que el álgebra L no es integrable a un grupo de Lie.

En efecto, dado L álgebra de Lie, definimos $G = e^{-iL}$. Es evidente que G tiene un neutro y cada elemento e^{-iX} tiene un inverso dado por e^{iX} . Falta ver que G es cerrado bajo multiplicación. Esto se ve usando Campbell-Hausdorff: Sean $X, Y \in L$

$$\begin{aligned} T(x) &\equiv e^{-iX}, \quad T(y) \equiv e^{-iY}, \quad T(x)T(y) = e^{-iX}e^{-iY} \equiv e^{-iZ}, \\ -iZ &= -iX - iY + \frac{1}{2}[-iX, -iY] + \dots, \quad \text{o} \quad Z = X + Y + \frac{1}{2}(-i)[X, Y] + \dots \end{aligned} \quad (2.46)$$

Como toda la serie está formada por conmutadores, $Z \in L$ y $T(x)T(y) \in G$. Las constantes de estructura de G son las del álgebra L .

Al mismo tiempo esta demostración implica que no sólo las constantes de estructura están fijadas por la ley de composición del grupo, sino que a su vez *la ley de composición* $c = f(a, b)$ *está completamente determinada por las constantes de estructura* (módulo cambios de coordenadas). En efecto, las componentes de Z se obtienen calculando conmutadores lo cual sólo requiere conocer las constantes de estructura

$$f^k(x, y) = z^k = x^k + y^k + \frac{1}{2}c_{ij}^k x^i y^j + \dots \quad (2.47)$$

En consecuencia el álgebra de Lie (caracterizada por sus constantes de estructura) determina completamente el grupo a nivel local: dos grupos son localmente isomorfos si y sólo si sus álgebras de Lie son isomorfas. Si se tienen los operadores concretos (no abstractos) del álgebra, su exponencial produce el grupo conexo (la componente conexa del grupo que contiene al neutro) concreto. Así por ejemplo, si $n = 1$ y $X_1 = 1$ (matriz 1×1) su exponencial e^{-iaX_1} produce el grupo $U(1)$, en cambio si $X_1 = i$ su exponencial produce un grupo isomorfo a $(\mathbb{R}, +)$. En ambos casos las relaciones de conmutación son las mismas, $[X_1, X_1] = 0$.

Si L es abeliana ($[L, L] = 0$) el grupo conexo es abeliano. Nótese que $O(2)$ tiene un álgebra de Lie abeliana pero el grupo mismo no es abeliano porque no es conexo. (La componente conexa es el grupo $SO(2)$ de rotaciones en el plano que sí es abeliano.)

Obsérvese que no todos los conjuntos de n^3 números c_{ij}^k definen un álgebra de Lie (es decir, existen n operadores linealmente independientes que cumplan las relaciones de conmutación). La condición necesaria y suficiente es que se satisfaga las condiciones (2.11). Además las constantes de estructura se pueden llevar a una forma canónica mediante cambios de base, por lo que hay menos álgebras de Lie distintas de lo podría parecer. Para cada dimensión n sólo hay un número finito de álgebras de Lie semisimples¹⁸ distintas (no isomorfas) y por tanto de grupos de Lie locales.

¹⁸Un álgebra es **simple** cuando no tiene ideales y **semisimple** cuando es suma directa de simples.

Para $n = 1$ sólo hay un álgebra (álgebra abeliana) y para $n = 2$ sólo hay dos álgebras distintas

$$\begin{aligned} [X_1, X_2] &= 0 && \text{(abeliana)} \\ [X_1, X_2] &= iX_1 && \text{(por ejemplo } X_1 = x, X_2 = -ix\partial_x \text{ en } L^2(\mathbb{R})) \end{aligned} \quad (2.48)$$

Análogamente, para dimensión 3 sólo hay dos álgebras de Lie reales simples (y por tanto semisimples).

2.10. Relación entre grupo de Lie y álgebra

Si H es un subgrupo de G , su álgebra de Lie L_H es una subálgebra de L , es decir, $-i[L_H, L_H] \subseteq L_H$. Y viceversa, la exponencial de una subálgebra genera un subgrupo.

Si H es un subgrupo invariante de G , L_H es un ideal de L , es decir, $-i[L_H, L] \subseteq L_H$. En efecto, si H es invariante (usando ec. (2.33))

$$\begin{aligned} g \in G, h \in H, \quad H \ni ghg^{-1} &= e^{-iX_g} e^{-iX_h} e^{iX_g} = \exp(-ie^{-iX_g} X_h e^{iX_g}) = e^{-iX_h - [X_g, X_h] + \dots} \\ &\Rightarrow -i[X_g, X_h] \in L_H \quad \forall X_g \in L, X_h \in L_H. \end{aligned} \quad (2.49)$$

Y viceversa si L_H es un ideal su exponencial es un subgrupo invariante.

Por otro lado, si $T^{(1)}(g)$ y $T^{(2)}(g)$ son dos representaciones de G en los espacios $V^{(1)}$ y $V^{(2)}$, se puede hacer su producto directo $T(g) = T^{(1)}(g) \otimes T^{(2)}(g)$ que actúa en el espacio producto tensorial $V = V^{(1)} \otimes V^{(2)}$. Los generadores infinitesimales de la representación T son

$$X_i = X_i^{(1)} \otimes 1 + 1 \otimes X_i^{(2)} \equiv X_i^{(1)} + X_i^{(2)}. \quad (2.50)$$

En efecto, para un elemento infinitesimal $g = e^{-i\delta X}$

$$T^{(1)}(g) \otimes T^{(2)}(g) = (1 + \delta X^{(1)}) \otimes (1 + \delta X^{(2)}) = 1 + \delta X^{(1)} \otimes 1 + 1 \otimes \delta X^{(2)}. \quad (2.51)$$

2.11. Medida invariante

En un grupo finito, si $\psi(g)$ es una función definida sobre el grupo, se puede definir su promedio como $\langle \psi \rangle_G = \frac{1}{|G|} \sum_{g \in G} \psi(g)$, y este promedio es invariante por la izquierda y por la derecha, es decir, las funciones $\psi(g'g)$ y $\psi(gg')$ tienen los mismos promedios que la función original. Esto se puede extender a grupos continuos.

Todo grupo de Lie tiene una **medida invariante**, también denominada medida de Haar, que es invariante por la izquierda y única salvo normalización. Y lo mismo por la derecha. Si $\psi(g)$ es una función definida sobre el grupo

$$\begin{aligned} \psi : G \rightarrow \mathbb{C}, \quad \int_G d\mu_L(g) \psi(g) &= \int_G d\mu_L(g) \psi(g'g) \\ \int_G d\mu_R(g) \psi(g) &= \int_G d\mu_R(g) \psi(gg') \end{aligned} \quad (2.52)$$

En un sistema de coordenadas a las medidas toman la forma

$$d\mu_{L,R}(g) = \rho_{L,R}(g) d^n a, \quad (2.53)$$

para ciertas densidades no negativas $\rho_{L,R}(g)$.

Por ejemplo, para el grupo de traslaciones en \mathbb{R}^n $\int d^n a \psi(a)$ es la medida invariante.

Con la medida invariante por la izquierda se puede definir el espacio de Hilbert $L^2(G, d\mu_L)$ de funciones complejas de cuadrado integrable definidas sobre el grupo y la representación regular por la izquierda es una representación unitaria con ese producto escalar. (Ídem por la derecha.)

Si el grupo es compacto o abeliano (y más casos) las medidas por la derecha e izquierda coinciden (se dice que el grupo es **unimodular**). Además si el grupo es compacto, se puede (y se suele) elegir

$$\int_G d\mu(g) = 1. \quad (2.54)$$

Por ejemplo para $U(1) \ni \omega = e^{-i\phi}$, la medida invariante normalizada es $\langle \psi \rangle_{U(1)} = \int_0^{2\pi} \frac{d\phi}{2\pi} \psi(\omega)$.

De acuerdo con el **teorema de Peter-Weyl**, para un grupo compacto G con irreps $D^\mu(g)$ de dimensión n_μ , las funciones $\sqrt{n_\mu} D^\mu(g^{-1})^i_j$ forman una **base ortonormal** de $L^2(G, d\mu)$. En consecuencia si $\psi(g)$ es de cuadrado integrable

$$\psi(g) = \sum_{\mu} \sum_{i,j=1}^{n_\mu} \psi_{\mu i}^j \sqrt{n_\mu} D^\mu(g^{-1})^i_j, \quad \psi_{\mu i}^j = \int_G d\mu(g) (\sqrt{n_\mu} D^\mu(g^{-1})^i_j)^* \psi(g). \quad (2.55)$$

Bajo la representación regular por la izquierda, cada subespacio $V_{\mu i} = \text{lin} \{D^\mu(g^{-1})^i_j\}_{j=1}^{n_\mu}$ lleva una irrep μ . En efecto, si $|j\rangle \equiv D^\mu(g^{-1})^i_j$,

$$T^L(g')|j\rangle = D^\mu((g'^{-1}g)^{-1})^i_j = D^\mu(g^{-1}g')^i_j = D^\mu(g^{-1})^i_k D^\mu(g')^k_j = D^\mu(g')^k_j |k\rangle. \quad (2.56)$$

2.12. Representación adjunta del grupo

Dada una representación fiel cualquiera T , se puede construir la **representación adjunta del grupo**, usando como espacio de representación

$$L^c = \{a \cdot X, a \in \mathbb{C}^n\}. \quad (2.57)$$

La representación $g \mapsto \text{Ad}g$ está definida por la acción

$$X \in L \xrightarrow{g} \text{Ad}g(X) \equiv T(g)XT(g)^{-1}. \quad (2.58)$$

Cuando $X \in L$ también $T(g)XT(g)^{-1} \in L$ por la relación

$$G \ni T(g)e^{-iX}T(g)^{-1} = \exp(-iT(g)XT(g)^{-1}) \quad (2.59)$$

(todo está definido en el grupo de modo que el resultado no depende de la representación usada). Por tanto $\text{Ad}L^c \subset L^c$. Además es una representación del grupo ya que aplicar g_2 y luego g_1 produce el mismo efecto que g_1g_2 . Esta representación es real: Para $X = a^i X_i$, $a^i \mapsto D(g)^i_j a^j$ y $D(g)$ una matriz real $n \times n$ ya que $\text{Ad}L \subset L$.

Para grupos abelianos la representación adjunta es equivalente a la representación trivial.

La representación adjunta del grupo induce la **representación adjunta del álgebra**, formada por operadores que actúan sobre L^c . Usando (2.22), es inmediato que X_i está representado por $\text{Ad}X_i$ definido por $[X_i, \cdot]$,

$$X \in L \mapsto \text{Ad}X \quad \text{Ad}X(Y) \equiv [X, Y] \quad \forall Y \in L^c. \quad (2.60)$$

Por la identidad de Jacobi, es inmediato comprobar que es una representación del álgebra, es decir $\text{Ad}(-i)[X, Y] = (-i)[\text{Ad}X, \text{Ad}Y]$. Esta representación es puramente imaginaria, ya que si $X, Y \in L$ $[X, Y] \in iL$. Las relaciones de conmutación ec. (2.42) implican que la matriz $n \times n$ correspondiente al generador X_i del álgebra en la representación adjunta es $(\text{Ad}X_i)^k_j = ic_{ij}^k$.¹⁹

2.13. Acción de un grupo de Lie sobre una variedad

Sea G un grupo de Lie de dimensión n , y M una variedad diferenciable de dimensión m . Una **acción** A de G sobre M es una aplicación

$$A : G \times M \ni (g, x) \mapsto A(g, x) \in M. \quad (2.61)$$

¹⁹En efecto, si $[X_i, X] = X'$ con $X = a^j X_j$ y $X' = a'^k X_k$, buscamos la matriz $\text{Ad}X_i$ tal que $(\text{Ad}X_i)^k_j a^j = a'^k$. Entonces, $[X_i, a^j X_j] = a^j ic_{ij}^k X_k = a'^k X_k$ implica $(\text{Ad}X_i)^k_j = ic_{ij}^k$.

tal que

$$A(e, x) = x, \quad A(g_1, A(g_2, x)) = A(g_1 g_2, x). \quad (2.62)$$

Por ejemplo, el propio grupo como espacio, $M = G$, y gx como producto en G . Por tanto todo grupo se puede ver como un grupo de transformaciones.

Cuando la acción a que nos referimos se sobreentiende, se suelen usar las notaciones

$$A(g, x) = x^g = gx. \quad (2.63)$$

2.13.1. Representación escalar

Sea $\mathcal{C}(M)$ el conjunto de funciones complejas definidas sobre la variedad M , en la que actúa el grupo G . Se puede definir una representación de G en $\mathcal{C}(M)$ mediante

$$g \mapsto U(g) \quad \psi \mapsto \psi^g = U(g)\psi, \quad \psi^g(x) \equiv \psi(g^{-1}x) \quad (2.64)$$

o equivalentemente

$$\psi^g(gx) = \psi(x) \quad (\text{transformación escalar}). \quad (2.65)$$

Podemos comprobar que $U(g)$ así definido es una representación del grupo:

$$\psi(x) \xrightarrow{g_2} \psi^{g_2}(x) = \psi(g_2^{-1}x) \xrightarrow{g_1} \psi^{g_2}(g_1^{-1}x) = \psi(g_2^{-1}(g_1^{-1}x)) = \psi((g_1 g_2)^{-1}x) = \psi^{g_1 g_2}(x) \quad (2.66)$$

que implica $U(g_1)U(g_2) = U(g_1 g_2)$.

2.13.2. Generadores infinitesimales

Para g infinitesimal con coordenadas δa^i , $i = 1, \dots, n$ y x^μ , $\mu = 1, \dots, m$,

$$(gx)^\mu = x^\mu + \delta x^\mu, \quad \delta x^\mu = \delta a^i f_i^\mu(x). \quad (2.67)$$

Igualmente,

$$g^{-1}x = x - \delta x \quad (2.68)$$

ya que $-\delta a^i$ son las coordenadas de g^{-1} . Las f_i^μ se pueden obtener mediante

$$(gx)^\mu = F^\mu(x, a), \quad f_i^\mu(x) = \left. \frac{\partial F^\mu}{\partial a_i} \right|_{a=0}. \quad (2.69)$$

Queremos obtener los generadores infinitesimales correspondientes a la representación escalar. Para ello consideramos al acción de g infinitesimal sobre una función ψ

$$\psi^g(x) = \psi(x - \delta x) \equiv \psi(x) + \delta\psi(x), \quad \delta\psi(x) = -\delta x \cdot \nabla\psi(x) = -\delta x^\mu \partial_\mu \psi(x). \quad (2.70)$$

Por otro lado

$$\psi^g = U(g)\psi = (1 - i\delta X)\psi, \quad \delta\psi = -i\delta X\psi = -i\delta a^i X_i \psi. \quad (2.71)$$

Se deduce

$$\delta X = -i\delta x^\mu \partial_\mu, \quad X_i = -if_i^\mu(x) \partial_\mu. \quad (2.72)$$

Por ejemplo para traslaciones $\delta x = \delta a$, $f_i^\mu = \delta_i^\mu$, $X_i = -i\partial_i$, $X = -i\nabla$.

2.13.3. Representaciones espinoriales

Si $D(g)$ es una representación de G en un espacio vectorial V de dimensión d (por ejemplo si $V = \mathbb{C}^d$, los D son matrices $d \times d$) se puede construir una representación en el espacio $\mathcal{H} = \mathcal{C}(M) \otimes V$. Los vectores de \mathcal{H} están descritos por funciones $\psi_\alpha(x)$, donde α etiqueta una base de V . Considerado como un vector de V para cada x , $\Psi(x) \in V$ (un vector columna de longitud d).

La representación $U(g)$ en \mathcal{H} se puede definir mediante

$$\Psi^g(x) = D(g)\Psi(g^{-1}x), \quad \psi_\alpha^g(x) = (D(g))_{\alpha\beta} \psi_\beta(g^{-1}x). \quad (2.73)$$

En efecto,

$$\begin{aligned} \Psi(x) &\xrightarrow{g_2} D(g_2)\Psi(g_2^{-1}x) \xrightarrow{g_1} D(g_1)(D(g_2)\Psi(g_2^{-1}(g_1^{-1}x))) = D(g_1)D(g_2)\Psi((g_1g_2)^{-1}x) \\ &= D(g_1g_2)\Psi((g_1g_2)^{-1}x) = \Psi^{g_1g_2}(x). \end{aligned} \quad (2.74)$$

A nivel infinitesimal, usando

$$D(g) = 1 - i\delta a^i S_i, \quad (2.75)$$

se encuentra

$$\begin{aligned} \Psi^g(x) &= (1 - i\delta a^i S_i)(1 - \delta x^\mu \partial_\mu)\Psi(x) = (1 - i\delta a^i S_i - \delta x^\mu \partial_\mu)\Psi(x) \\ &\equiv (1 - i\delta a^i X_i)\Psi(x) \end{aligned} \quad (2.76)$$

que implica

$$X_i = -if_i^\mu(x) \partial_\mu + S_i \equiv L_i + S_i. \quad (2.77)$$

El operador L_i sólo actúa en $\mathcal{C}(M)$ y S_i sólo actúa en V y ambos operadores conmutan. Este resultado es consistente con ec. (2.50).

3. Grupo de rotaciones

3.1. Grupo de rotaciones en \mathbb{R}^n

$O(n, \mathbb{R})$ u $O(n)$ es el grupo de las matrices ortogonales reales $n \times n$

$$O(n) = \{A \text{ matriz } n \times n, A^* = A, A^T A = 1\} \quad (3.1)$$

(Compruébese que forman grupo.) Equivalentemente, $A^{-1} = A^T$ ó $A^{-1T} = A$.

La ortogonalidad de la matriz implica que como aplicación lineal conserva la norma euclídea en \mathbb{R}^n (o equivalentemente el producto escalar)

$$\mathbf{x} \in \mathbb{R}^n \quad \|\mathbf{Ax}\|^2 = \mathbf{Ax} \cdot \mathbf{Ax} = A^i_j x^j A^i_k x^k = x^j (A^T A)_{jk} x^k = \|\mathbf{x}\|^2 \quad (3.2)$$

o también $\|\mathbf{Ax}\|^2 = (\mathbf{Ax})^T (\mathbf{Ax}) = x^T A^T A x = x^T x = \|x\|^2$, donde x es un vector columna, es decir, una matriz $n \times 1$.

Que A sea ortogonal equivale a decir que sus n columnas son las componentes de n vectores que forman una base ortonormal, y lo mismo las n filas.

La propiedad de conservar la norma se puede tomar como definición de $O(n)$. En general, *un conjunto de transformaciones invertibles que dejan una propiedad invariante siempre define un grupo.*

El grupo $O(n)$ no es conexo:

$$1 = \det(A^T A) = \det(A)^2 \Rightarrow \det(A) = \pm 1. \quad (3.3)$$

Como siempre, la componente conexa del neutro forma un subgrupo invariante, el grupo de las **rotaciones** en n dimensiones:

$$SO(n) = \{R \text{ matriz } n \times n, R^* = R, R^T R = 1, \det(R) = 1\}. \quad (3.4)$$

Los grupos $SO(n)$ son simples (para $n > 2$), compactos y conexos pero no simplemente conexos. $SO(2)$ es abeliano.

$$O(n)/SO(n) \cong \mathbb{Z}_2, \quad O(n) = SO(n) \cup O^-(n). \quad (3.5)$$

Por ejemplo $A = \text{diag}(-1, 1, 1, \dots, 1) \in O^-(n)$ representa una **reflexión** respecto del primer eje. $O^-(n)$ está formado por transformaciones del tipo rotaciones seguidas de una reflexión.

Para *dimensiones impares* se puede definir inversión espacial o **paridad** $P \equiv \text{diag}(-1, -1, \dots, -1) = -1 \in O^-(n)$ de modo que $\{1, P\} \cong \mathbb{Z}_2$ es un subgrupo invariante de $O(n)$ y de hecho

$$O(n) \cong SO(n) \otimes \mathbb{Z}_2 \quad (n \text{ impar}). \quad (3.6)$$

En el **punto de vista pasivo** (cambian las componentes porque cambia la base)

$$x'^i = A^i_j x^j \quad (3.7)$$

A es la matriz del cambio de base,

$$A \in GL(n, \mathbb{R}) = \{A, \text{ matriz } n \times n, A^* = A, \det(A) \neq 0\} \quad (3.8)$$

Para los vectores de la base

$$\mathbf{x} = x^j \mathbf{e}_j = x'^i \mathbf{e}'_i = A^i_j x^j \mathbf{e}'_i \Rightarrow \mathbf{e}_i = A^i_j \mathbf{e}'_i, \quad \mathbf{e}'_i = (A^{-1})^i_j \mathbf{e}_j. \quad (3.9)$$

Los objetos/índices que se transforman como la base se dice que son **covariantes** [índice abajo, ec. (3.9)], y los que se transforman como las componentes se dice que son **contravariantes** [índice arriba, ec. (3.7)].

Por definición dos bases tienen la misma **orientación** si $\det(A) > 0$. Por tanto, las rotaciones se pueden definir como las transformaciones lineales que dejan invariante la norma y la orientación,²⁰ es decir, las rotaciones transforman entre sí bases ortonormales con orientación positiva. Hay una biyección entre rotaciones y bases ortonormales positivamente orientadas: dada una tal base fija B_0 , cualquier otra base B está fijada por la rotación que lleva B_0 a B .

Las transformaciones en $O^-(n)$ se denominan **rotaciones impropias** y cambian la orientación.

Como la matriz $R^T R$ es automáticamente simétrica, la ecuación $R^T R = 1$ sobre $R \in GL(n, \mathbb{R})$ impone $n(n+1)/2$ ecuaciones para n^2 incógnitas, en consecuencia

$$\dim SO(n) = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}. \quad (3.10)$$

Otra forma de verlo es contar bases ortonormales positivas. Para especificar el primer vector de la base se necesitan $n-1$ parámetros (n componentes pero la norma está fijada a 1). Para el vector k -ésimo se requerirán $n-k$ parámetros: estar normalizado quita un parámetro y ser ortogonal a cada uno de los $k-1$ vectores anteriores quitan otros $k-1$ parámetros. En total $\sum_{k=1}^n (n-k) = \sum_{j=1}^{n-1} j = n(n-1)/2$.

²⁰En realidad la condición de que sean lineales se puede deducir de conservación de la norma.

3.2. Álgebra de Lie de $SO(n)$

Considerando transformaciones infinitesimales, $R = 1 - i\delta X$

$$1 = R^T R = (1 - i\delta X^T)(1 - i\delta X) = 1 - i(\delta X^T + \delta X) \Rightarrow \delta X^T + \delta X = 0. \quad (3.11)$$

Se deduce que el álgebra de Lie $so(n)$ está formada por matrices imaginarias puras antisimétricas

$$so(n) = \{X, \text{ matriz } n \times n, X^* = -X, X^T = -X\}. \quad (3.12)$$

Se comprueba que forman álgebra:

$$X, Y \in so(n), \quad (-i[X, Y])^T = -i[Y^T, X^T] = -i[Y, X] = i[X, Y] \Rightarrow -i[X, Y] \in so(n). \quad (3.13)$$

Estas matrices son hermíticas, $X^\dagger = X$ y las R son unitarias (las matrices ortogonales reales son automáticamente unitarias). Exponenciando el álgebra se obtiene el grupo (ya que es conexo)

$$SO(n) = \{R = e^{-iX}, \quad X \in so(n)\}, \quad R^T = e^{-iX^T} = e^{iX} = R^{-1}. \quad (3.14)$$

La condición de que la matriz X sea antisimétrica automáticamente implica que el número de parámetros (reales) libres es $n(n-1)/2$ (este es el número de elementos de matriz por encima de la diagonal).

3.3. Grupo $SO(2)$

$SO(2)$ son las rotaciones en el plano,

$$SO(2) = \left\{ R = \begin{pmatrix} \cos \phi & -\text{sen } \phi \\ \text{sen } \phi & \cos \phi \end{pmatrix}, \quad -\pi < \phi \leq \pi \right\}. \quad (3.15)$$

(Compruébese usando $R^T R = 1$, $\det(R) = 1$.) Este grupo es isomorfo al grupo de las fases

$$SO(2) \cong U(1) = \{e^{i\phi}, \quad -\pi < \phi \leq \pi\}. \quad (3.16)$$

La ley de composición es simplemente $\phi_{12} = \phi_1 + \phi_2$.

Como el grupo es abeliano sus irreps $D^\mu(\phi)$ son de dimensión 1. La condición $D(R_1)D(R_2) = D(R_1 R_2)$ implica

$$D(R) = e^{-i\alpha\phi}, \quad \alpha \in \mathbb{C} \quad (3.17)$$

α es una constante característica de la representación. Para que la representación sea univaluada, $\alpha = m \in \mathbb{Z}$, por ejemplo $1 = D(\pi/2)^4 = (e^{-i\alpha\pi/2})^4 = e^{-i2\pi\alpha}$.

$$D^m(R) = e^{-im\phi}, \quad m \in \mathbb{Z} \quad (\text{irreps de SO}(2)). \quad (3.18)$$

Alternativamente, por la teoría general, y teniendo en cuenta que ϕ es directamente una coordenada canónica, $D(R) = e^{-i\phi J}$ donde J es el generador infinitesimal ($J \in \mathbb{C}$, es una matriz 1×1 compleja) que automáticamente cumple el álgebra de Lie, $[J, J] = 0$. Para que sea una representación de $U(1)$ y no sólo de su recubridor $(\mathbb{R}, +)$, $J = m \in \mathbb{Z}$.

La propia representación R que define al grupo $SO(2)$ es bidimensional y por tanto es reducible (como representación compleja). Para ver cómo se reduce es más cómodo trabajar con el álgebra. Para una rotación infinitesimal (J_z denota el generador infinitesimal correspondiente)

$$R = \begin{pmatrix} 1 & -\delta\phi \\ \delta\phi & 1 \end{pmatrix} = 1 - iJ_z\delta\phi, \quad J_z = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \equiv \sigma_2 \quad (3.19)$$

(σ_2 es una de las matrices de Pauli). $J_z = J_z^\dagger$, $J_z^2 = 1$, de aquí

$$e^{-i\phi J_z} = \cos(\phi J_z) - i \operatorname{sen}(\phi J_z) = \cos(\phi) - i \operatorname{sen}(\phi) \sigma_2 = \begin{pmatrix} \cos \phi & -\operatorname{sen} \phi \\ \operatorname{sen} \phi & \cos \phi \end{pmatrix}. \quad (3.20)$$

La representación $J_z = \sigma_2$ se puede reducir (se puede diagonalizar)

$$e_{\pm 1} \equiv \mp \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \end{pmatrix}, \quad J_z e_\lambda = \lambda e_\lambda, \quad \langle e_\lambda | e_{\lambda'} \rangle = e_\lambda^* \cdot e_{\lambda'} = \delta_{\lambda\lambda'}, \quad \lambda, \lambda' = \pm 1. \quad (3.21)$$

Esto implica

$$R(\phi) e_\lambda = e^{-i\phi J_z} e_\lambda = e^{-i\lambda\phi} e_\lambda, \quad \lambda = \pm 1 \quad (3.22)$$

y los dos vectores $e_{\pm 1}$ generan sendos subespacios irreducibles unidimensionales. Por tanto la reducción es

$$R = D^{m=1} \oplus D^{m=-1}. \quad (3.23)$$

En la base $e_{\pm 1}$ la rotación toma la forma $\begin{pmatrix} e^{-i\phi} & 0 \\ 0 & e^{i\phi} \end{pmatrix}$.

$SO(2)$ es compacto, tiene una medida biinvariante normalizable

$$\langle f \rangle_{SO(2)} \equiv \int_0^{2\pi} \frac{d\phi}{2\pi} f(\phi) = \oint \frac{d\omega}{2\pi i} \frac{1}{\omega} f(\omega) \quad (\omega = e^{i\phi}). \quad (3.24)$$

(La integral de contorno es sobre $|\omega| = 1$ con orientación positiva.) Esta medida es invariante

$$\langle f(\omega) \rangle_{\text{SO}(2)} = \langle f(\omega' \omega) \rangle_{\text{SO}(2)}. \quad (3.25)$$

También se aplica el teorema de Peter-Weyl, las irreps $e^{-im\phi}$, $m \in \mathbb{Z}$ forman una base ortonormal de las funciones periódicas $f(\phi)$

$$\begin{aligned} f(\phi) &= \sum_{m \in \mathbb{Z}} f_m e^{im\phi}, & f_m &= \int_0^{2\pi} \frac{d\phi}{2\pi} e^{-im\phi} f(\phi) \\ \int_0^{2\pi} \frac{d\phi}{2\pi} e^{-im\phi} e^{im'\phi} &= \delta_{mm'}, \\ \sum_{m \in \mathbb{Z}} e^{im\phi} e^{-im'\phi} &= \sum_{n \in \mathbb{Z}} 2\pi \delta(\phi - \phi' - 2\pi n) = 2\pi \delta(\phi - \phi') \quad (\phi, \phi' \text{ mód } 2\pi). \end{aligned} \quad (3.26)$$

Esta última relación es la **identidad de Poisson**.

Con otra notación

$$\begin{aligned} \langle m|m' \rangle &= \delta_{mm'}, & \langle \phi|\phi' \rangle &= 2\pi \delta(\phi - \phi') \quad (\phi, \phi' \text{ mód } 2\pi), \\ |f\rangle &= \sum_{m \in \mathbb{Z}} f_m |m\rangle = \int_0^{2\pi} \frac{d\phi}{2\pi} f(\phi) |\phi\rangle, & f_m &= \langle m|f\rangle, & f(\phi) &= \langle \phi|f\rangle, & \langle \phi|m\rangle &= e^{im\phi}. \end{aligned} \quad (3.27)$$

Sobre funciones (escalares) definidas sobre el plano \mathbb{R}^2 , las rotaciones actúan según $\psi(\mathbf{x}) \mapsto \psi(R^{-1}\mathbf{x})$. Podemos calcular el generador infinitesimal en esta representación (que denotamos L_z):

$$\begin{aligned} \delta \mathbf{x} &= \begin{pmatrix} \delta x^1 \\ \delta x^2 \end{pmatrix} = \begin{pmatrix} 0 & -\delta\phi \\ \delta\phi & 0 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} -\delta\phi x^2 \\ \delta\phi x^1 \end{pmatrix}, \\ -i\delta\phi L_z &= -\delta \mathbf{x} \cdot \nabla = -\delta\phi (-x^2 \partial_1 + x^1 \partial_2) \Rightarrow L_z = -i(x^1 \partial_2 - x^2 \partial_1), \\ e^{-i\phi L_z} \psi(\mathbf{x}) &= \psi(R^{-1}\mathbf{x}). \end{aligned} \quad (3.28)$$

3.4. Grupo $\text{SO}(3)$

Una rotación en \mathbb{R}^3 se puede parametrizar mediante un eje $\hat{\mathbf{n}}$, $\hat{\mathbf{n}}^2 = 1$, $\hat{\mathbf{n}} \in S^2$, y un ángulo θ , $R(\hat{\mathbf{n}}, \theta)$. En total tres parámetros, de acuerdo con la fórmula general de $\text{SO}(n)$.

Prescindiendo de conocimientos previos sobre rotaciones, usando sólo $R \in \text{SO}(3)$, se ve que el polinomio característico $\det(R - \lambda) = 0$ es cúbico y debe tener una solución real, es decir, $R\mathbf{x} = \lambda\mathbf{x}$

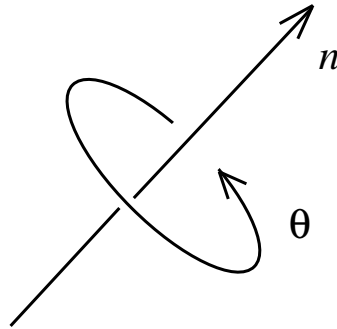


Figura 3: Convenio eje-ángulo: El ángulo se toma en dirección positiva relativa al eje (regla del tornillo).

($x \neq 0$). Por la conservación de norma $\lambda = \pm 1$ y por continuidad con $R = 1$ (el grupo es conexo) $\lambda = 1$. Por tanto hay un $\hat{n} \in S^2$ tal que $R\hat{n} = \hat{n}$. Por otro lado, un $x \in \mathbb{R}^3$ cualquiera se puede escribir $x = x_{\parallel} + x_{\perp}$ (según \hat{n}) y $Rx = x_{\parallel} + Rx_{\perp}$, por conservación del producto escalar $Rx_{\perp} \perp \hat{n}$ y además $\|Rx_{\perp}\| = \|x_{\perp}\|$ (y conserva la orientación por continuidad). En consecuencia x_{\perp} y Rx_{\perp} están relacionados por una rotación del plano perpendicular a \hat{n} , es una rotación de $SO(2)$ y tiene asociado un ángulo θ :

$$R(\hat{n}, \theta)x = x_{\parallel} + \cos(\theta)x_{\perp} + \sin(\theta)\hat{n} \times x_{\perp}. \quad (3.29)$$

De esta expresión se deduce que

$$R(\hat{n}, \theta) = R(\hat{n}, \theta + 2\pi) = R(-\hat{n}, 2\pi - \theta) \quad (3.30)$$

por lo cual el ángulo se puede restringir a $0 \leq \theta \leq \pi$. Incluso así hay casos repetidos por $R(\hat{n}, \pi) = R(-\hat{n}, \pi)$ pero forman un conjunto de medida nula.

De la ec. (3.29) se puede deducir que, para una rotación cualquiera R' ,

$$R'R(\hat{n}, \theta)R'^{-1} = R(\hat{n}', \theta), \quad \hat{n}' = R'\hat{n}. \quad (3.31)$$

Esto permite llevar cualquier eje a cualquier otro y demuestra que las clases de conjugación de $SO(3)$ está unívocamente determinadas por el ángulo de rotación, $0 \leq \theta \leq \pi$.

Las rotaciones sobre un eje fijo, digamos el eje z , $\hat{n} = e_3$, forman un subgrupo isomorfo a $SO(2)$. Por tanto, $SO(3) \supset SO(2)$ con $R_3 = \begin{pmatrix} R_2 & 0 \\ 0 & 1 \end{pmatrix}$.

Dada una rotación $R \neq 1$ con eje \hat{n} , se tiene $R\hat{n} = \hat{n}$ y de hecho éste es el único vector propio real.²¹ Por ello, si $Rx = x$ entonces x es paralelo a \hat{n} . Una consecuencia es que la representación de $SO(3)$ definida por las matrices R es irreducible (como representación compleja). En efecto, si fuera reducible habría un vector propio (complejo) común a todas las rotaciones, pero los únicos vectores propios de una rotación (que no sea la identidad) son el eje \hat{n} y los $\hat{e}_{\pm 1}$ asociados al plano perpendicular al eje, y estos vectores dependen de la rotación.

Los tres números

$$\theta = \theta \hat{n} \quad (3.32)$$

definen un sistema de coordenadas (que son analíticas, como se sigue de ec. (3.29)). El grupo no es abeliano pero si se hacen dos rotaciones sucesivas sobre el mismo eje los ángulos se suman

$$R(\hat{n}, \theta_1)R(\hat{n}, \theta_2) = R(\hat{n}, \theta_1 + \theta_2). \quad (3.33)$$

Esto implica que θ define un sistema de **coordenadas canónicas** para $SO(3)$. El sistema de coordenadas vale en el abierto $|\theta| < \pi/2$, con $\hat{n} \in S^2$. Es importante notar que

$$R(\theta_1)R(\theta_2) \neq R(\theta_1 + \theta_2) \quad \text{en general.} \quad (3.34)$$

La ley de composición $\theta_{12} = f(\theta_1, \theta_2)$ es complicada y no se necesita en forma explícita.

3.4.1. Ángulos de Euler

Sea e_i , $i = 1, 2, 3$, una base ortonormal positiva, y $R_i(\theta) \equiv R(e_i, \theta)$. La rotación más general se puede escribir como

$$R = R_3(\alpha)R_2(\beta)R_3(\gamma), \quad 0 \leq \alpha, \gamma < 2\pi, \quad 0 \leq \beta \leq \pi. \quad (3.35)$$

En efecto, dada una rotación R , si el nuevo eje z es $e'_3 = (\beta, \alpha)$ (ángulos polar y acimutal, respectivamente)

$$Re_3 = e'_3 = R_3(\alpha)R_2(\beta)e_3 \Rightarrow (R_3(\alpha)R_2(\beta))^{-1}Re_3 = e_3 \Rightarrow (R_3(\alpha)R_2(\beta))^{-1}R = R_3(\gamma). \quad (3.36)$$

Hay que señalar que los parámetros (α, β, γ) , ángulos de Euler, no forman un auténtico sistema de coordenadas ya que $e = (\alpha, 0, -\alpha)$ (no es una biyección local entre elementos del grupo y parámetros).

²¹Los $e_{\pm 1}$ de ec. (3.21) sí serían vectores propios pero complejos.

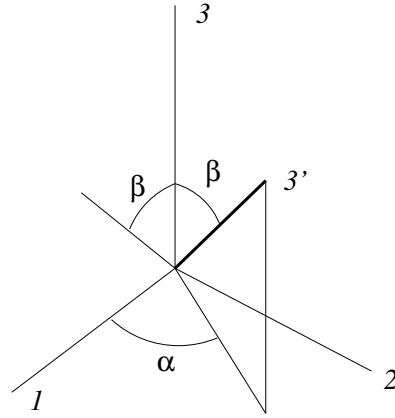


Figura 4: La rotación $R_2(\beta)$ seguida de $R_3(\alpha)$, $R_3(\alpha)R_2(\beta)$, lleva el eje 3 a $3'$, con ángulos polar β y acimutal α .

3.4.2. Generadores infinitesimales y álgebra de $SO(3)$

Dado que θ son coordenadas canónicas podemos escribir, en cualquier representación

$$U(R) = e^{-i\theta \cdot J}. \quad (3.37)$$

En la propia representación R (que coincide con la representación adjunta del grupo $SO(3)$)

$$R(e_3, \theta) = \begin{pmatrix} \cos \theta & -\text{sen } \theta & 0 \\ \text{sen } \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.38)$$

que implica

$$J_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.39)$$

análogamente, considerando rotaciones según los otros dos ejes

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}. \quad (3.40)$$

Estas matrices son hermíticas y sin traza y satisfacen las relaciones de conmutación

$$[J_1, J_2] = iJ_3, \quad [J_2, J_3] = iJ_1, \quad [J_3, J_1] = iJ_2, \quad (3.41)$$

es decir,

$$[J_i, J_j] = iJ_k, \quad ijk \text{ permutación cíclica de } 123. \quad (3.42)$$

Como todos los generadores se mezclan no hay ningún ideal y el álgebra y el grupo son simples. Las relaciones pueden también escribirse

$$[J_i, J_j] = i\varepsilon_{ijk}J_k \quad (\text{Álgebra del momento angular}). \quad (3.43)$$

Aquí ε_{ijk} es el **tensor de Levi-Civita**, que es completamente antisimétrico

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{si } ijk \text{ es una permutación par de } 123 \\ -1 & \text{si } ijk \text{ es una permutación impar de } 123 \\ 0 & \text{si se repite algún índice} \end{cases} \quad (3.44)$$

En particular

$$\varepsilon_{ijk} = -\varepsilon_{jik}, \quad \varepsilon_{ijk} = \varepsilon_{jki}, \quad (\mathbf{A} \times \mathbf{B})_i = \varepsilon_{ijk}A_jB_k. \quad (3.45)$$

Otra forma de obtener los generadores J_i es considerar una rotación infinitesimal sobre $\mathbf{x} \in \mathbb{R}^3$. Usando ec. (3.29) se deduce

$$\delta \mathbf{x} = \delta \boldsymbol{\theta} \times \mathbf{x}, \quad (3.46)$$

y por definición de generador infinitesimal [ec. (2.22)] se tendrá

$$\delta \mathbf{x} = -i(\delta \boldsymbol{\theta} \cdot \mathbf{J})\mathbf{x}. \quad (3.47)$$

Comparando ambas expresiones (en componentes):

$$\begin{aligned} \delta x^i &= (\delta \boldsymbol{\theta} \times \mathbf{x})^i = \varepsilon_{ilj} \delta \theta^l x^j, \\ \delta x^i &= (-i \delta \boldsymbol{\theta} \cdot \mathbf{J})^i_j x^j = -i \delta \theta^l (J_l)^i_j x^j, \end{aligned} \quad (3.48)$$

se deduce

$$(J_i)_{jk} = -i\varepsilon_{ijk}, \quad (3.49)$$

que coincide con el resultado en (3.39) y (3.40).²²

La ec. (3.43) dice que las constantes de estructura de $\text{SO}(3)$ son $c_{ij}^k = \varepsilon_{ijk}$ en esta base. En cualquier representación de $\text{SO}(3)$ es convencional elegir la base del álgebra de modo que las

²²Para reconstruir una matriz A_{ij} a partir de sus elementos de matriz el convenio es que el primer índice es la fila y el segundo la columna. Así $(J_3)_{12} = -i\varepsilon_{312} = -i$ coincide con el elemento de matriz (12) en (3.39).

constantes de estructura sean ε_{ijk} (esto es automático si la base son los generadores asociados a las coordenadas θ).

Para todo grupo de Lie las constantes de estructura proporcionan las matrices de la representación adjunta mediante $(\text{Ad} X_i)^k_j = ic_{ij}^k$, y la ec. (3.49) verifica esta propiedad para $\text{SO}(3)$.

El tensor de Levi-Civita satisface las relaciones

$$\begin{aligned}\varepsilon_{ijk}\varepsilon_{abc} &= \delta_{ia}\delta_{jb}\delta_{kc} \pm \text{permutaciones de } abc \\ \varepsilon_{ijk}\varepsilon_{abk} &= \delta_{ia}\delta_{jb} - \delta_{ib}\delta_{ja} \\ \varepsilon_{ijk}\varepsilon_{ajk} &= 2\delta_{ia}\end{aligned}\tag{3.50}$$

Entonces

$$\varepsilon_{ijk}[J_i, J_j] = i\varepsilon_{ijk}\varepsilon_{ijc}J_c = 2iJ_k \Rightarrow \mathbf{J} \times \mathbf{J} = i\mathbf{J}.\tag{3.51}$$

(Para c-números $\mathbf{a} \times \mathbf{a} = 0$ pero \mathbf{J} es un operador y las componentes J_i y J_j no conmutan si $i \neq j$.)

La representación definida por las matrices R es unitaria y correspondientemente $\mathbf{J}^\dagger = \mathbf{J}$. Como ya se dijo la representación es irreducible y ello se comprueba igualmente en el álgebra (por ejemplo, $[J_i, X] = 0 \Rightarrow X = \lambda 1$, X matriz 3×3).

3.4.3. Operador momento angular orbital

En $L^2(\mathbb{R}^3)$ tenemos la representación escalar

$$\psi(\mathbf{x}) \xrightarrow{R} (U(R)\psi)(\mathbf{x}) = \psi(R^{-1}\mathbf{x}).\tag{3.52}$$

Esta representación es unitaria:

$$\|U(R)\psi\|^2 = \int d^3x |\psi(R^{-1}\mathbf{x})|^2 = \int d^3x |\psi(\mathbf{x})|^2 = \|\psi\|^2.\tag{3.53}$$

Se ha usado que la medida (el elemento de volumen) es invariante bajo rotaciones. En efecto, si $\mathbf{y} \equiv R^{-1}\mathbf{x}$, y cambiamos de variable $\mathbf{x} = R\mathbf{y}$, $d^3x = \det(R)d^3y = d^3y$.²³

El operador $U(R)$ que representa a la rotación R se puede escribir

$$U(R) = e^{-i\theta \cdot \mathbf{L}}, \quad \mathbf{L} = \mathbf{L}^\dagger.\tag{3.54}$$

²³El grupo de transformaciones lineales en \mathbb{R}^n que dejan el volumen invariante, es decir, $\det A = 1$, forman el grupo $\text{SL}(n, \mathbb{R})$.

L , momento angular orbital, es el nombre usual de J en esta representación. Para determinar este operador necesitamos $\delta\mathbf{x}$, que ya hemos obtenido en ec. (3.46). Usando entonces la relación (2.72)

$$\delta\boldsymbol{\theta} \cdot \mathbf{L} = -i\delta\mathbf{x} \cdot \nabla = -i(\delta\boldsymbol{\theta} \times \mathbf{x}) \cdot \nabla = -i\delta\boldsymbol{\theta} \cdot (\mathbf{x} \times \nabla). \quad (3.55)$$

Finalmente

$$\mathbf{L} = -i\mathbf{x} \times \nabla = \mathbf{x} \times \mathbf{p} = -\mathbf{p} \times \mathbf{x} \quad \mathbf{p} \equiv -i\nabla. \quad (3.56)$$

Aquí \mathbf{x} es un operador multiplicativo, el operador posición. (Nótese que en general para operadores $\mathbf{A} \times \mathbf{B}$ no coincide con $-\mathbf{B} \times \mathbf{A}$, pero x^i conmuta con p_j si $i \neq j$.) En componentes

$$L_i = -i\epsilon_{ijk}x^j\partial_k = \epsilon_{ijk}x^j p_k \quad (3.57)$$

o también

$$L_i = x^j p_k - x^k p_j, \quad (ijk) \text{ permutación cíclica de } (123), \quad (3.58)$$

explícitamente

$$L_x = yp_z - zp_y, \quad L_y = zp_x - xp_z, \quad L_z = xp_y - yp_x. \quad (3.59)$$

Por supuesto L_z es el mismo operador ya obtenido en ec. (3.28) para $SO(2)$.

Para derivar las relaciones de conmutación en esta representación, se pueden usar las relaciones

$$[\partial_i, f(\mathbf{x})] = (\partial_i f)(\mathbf{x}), \quad [\partial_i, x^j] = \delta_i^j, \quad [x^i, p_j] = i\delta_j^i, \quad [x^i, x^j] = [p_i, p_j] = 0. \quad (3.60)$$

Y también

$$[A, BC] = [A, B]C + B[A, C], \quad (3.61)$$

que expresa que $[A,]$ es una derivación (satisface la regla de Leibniz), e igualmente $[, A] = -[A,]$. Así, si ijk es una permutación cíclica de 123

$$\begin{aligned} [L_i, L_j] &= -[x^j\partial_k - x^k\partial_j, x^k\partial_i - x^i\partial_k] = -[x^j\partial_k, x^k\partial_i] - [x^k\partial_j, x^i\partial_k] \\ &= -x^j\partial_i + x^i\partial_j = iL_k. \end{aligned} \quad (3.62)$$

Alternativamente, para índices i, j, \dots arbitrarios

$$\begin{aligned} [L_i, L_j] &= -\epsilon_{iab}\epsilon_{jcd}[x^a\partial_b, x^c\partial_d] = -\epsilon_{iab}\epsilon_{jcd}([x^a\partial_b, x^c]\partial_d + x^c[x^a\partial_b, \partial_d]) \\ &= -\epsilon_{iab}\epsilon_{jcd}(x^a\delta_b^c\partial_d - x^c\delta_d^a\partial_b) = -\epsilon_{iab}\epsilon_{jbd}x^a\partial_d + \epsilon_{iab}\epsilon_{jca}x^c\partial_b \\ &= (\delta_i^d\delta_a^j - \delta_a^d\delta_i^j)x^a\partial_d - (\delta_j^b\delta_c^i - \delta_c^b\delta_j^i)x^c\partial_b = x^j\partial_i - x^i\partial_j \\ &= -\epsilon_{ijk}\epsilon_{abk}x^a\partial_b = i\epsilon_{ijk}L_k \end{aligned} \quad (3.63)$$

De hecho, dada cualquier álgebra de Lie n -dimensional, el conjunto de operadores $X_i = c_{ij}^k x^j p_k$ en $L^2(\mathbb{R}^n)$ satisface la relaciones de conmutación $[X_i, X_j] = ic_{ij}^k X_k$, como consecuencia de la identidad de Jacobi para las constantes de estructura. Esta realización del álgebra se denomina **álgebra de Poisson-Lie**.

3.4.4. Operadores escalares y vectoriales

En una representación cualquiera de $SO(3)$, se dice que S es un **operador escalar** bajo rotaciones si

$$S \mapsto U(R)SU(R)^{-1} = S \quad (3.64)$$

igualmente, A es un **operador vectorial** si

$$A \mapsto U(R)AU(R)^{-1} = R^{-1}A, \quad U(R)A^iU(R)^{-1} = (R^{-1})^i_j A^j = R^j_i A^j. \quad (3.65)$$

Equivalentemente, $R^i_j U(R)A^j U(R)^{-1} = A^i$. Obsérvese que $U(R)AU(R)^{-1} = RA$ no sería consistente con $U(R_1 R_2) = U(R_1)U(R_2)$.

Por ejemplo el operador ∇ en $L^2(\mathbb{R}^3)$ es un vector:

$$\begin{aligned} U(R)\partial_i U(R)^{-1}\psi(\mathbf{r}) &= U(R)\partial_i(\psi(R\mathbf{r})) = U(R)\frac{\partial(R\mathbf{r})^j}{\partial r^i}(\partial_j\psi)(R\mathbf{r}) = U(R)R^j_i(\partial_j\psi)(R\mathbf{r}) \\ &= R^j_i(\partial_j\psi)(\mathbf{r}) = (R^{-1})^i_j \partial_j\psi(\mathbf{r}). \end{aligned} \quad (3.66)$$

Igualmente el operador posición \mathbf{x} ²⁴

$$\begin{aligned} U(R)|\mathbf{r}\rangle &= |R\mathbf{r}\rangle, \quad \mathbf{x}|\mathbf{r}\rangle = \mathbf{r}|\mathbf{r}\rangle \\ U(R)\mathbf{x}U(R)^{-1}|\mathbf{r}\rangle &= U(R)\mathbf{x}|R^{-1}\mathbf{r}\rangle = U(R)R^{-1}\mathbf{r}|R^{-1}\mathbf{r}\rangle = R^{-1}\mathbf{r}|\mathbf{r}\rangle = R^{-1}\mathbf{x}|\mathbf{r}\rangle. \end{aligned} \quad (3.67)$$

Se comprueba que si A y B son operadores vectoriales, su producto escalar es un escalar y su producto vectorial un vector:

$$\begin{aligned} UA \cdot BU^{-1} &= UA^i U^{-1} U B^j U^{-1} = (R^{-1})^i_j A^j (R^{-1})^k_l B^l = (R^T R)_{jk} A^j B^k \\ &= \delta_{jk} A^j B^k = \mathbf{A} \cdot \mathbf{B}, \\ R^i_j U(\mathbf{A} \times \mathbf{B})^j U^{-1} &= (R^{-1})^j_i \varepsilon_{jkl} (R^{-1})^k_a A^a (R^{-1})^l_b B^b = \det(R^{-1}) \varepsilon_{iab} A^a B^b = (\mathbf{A} \times \mathbf{B})^i. \end{aligned} \quad (3.68)$$

En la última igualdad se ha utilizado la identidad

$$\varepsilon_{i_1, \dots, i_n} A^{i_1}_{j_1} \cdots A^{i_n}_{j_n} = \det(A) \varepsilon_{j_1, \dots, j_n}, \quad (3.69)$$

²⁴ $|\psi\rangle = \int d^3x \psi(\mathbf{x})|\mathbf{x}\rangle$ implica $U(R)|\psi\rangle = \int d^3x \psi(R^{-1}\mathbf{x})|\mathbf{x}\rangle = \int d^3x \psi(\mathbf{x})|R\mathbf{x}\rangle$.

válida para una matriz cuadrada A arbitraria.

Nótese que no hace falta que los operadores involucrados conmuten entre sí.

Si A y B se transforman según ec. (3.65) incluso cuando R es una rotación *impropia* se dice que son **vectores polares** o auténticos vectores. El operador producto $C = A \times B$ en cambio será un **pseudo-vector** o **vector axial**, es decir, bajo rotaciones impropias $C \mapsto -R^{-1}C$, $R \in O^-(3)$. En efecto, en la demostración anterior salía un factor $\det(R)$ que antes era 1 pero para rotaciones impropias es -1 .

A nivel infinitesimal, $U = 1 - i\delta\theta \cdot \mathbf{J}$, la transformación de un operador bajo rotaciones es

$$A \mapsto UAU^{-1} = (1 - i\delta\theta \cdot \mathbf{J})A(1 + i\delta\theta \cdot \mathbf{J}) \Rightarrow \delta A = -i\delta\theta \cdot [\mathbf{J}, A] \quad (3.70)$$

Para un escalar

$$\delta S = 0 \Leftrightarrow [\mathbf{J}, S] = 0. \quad (3.71)$$

Para un operador vectorial

$$R^{-1}\mathbf{A} = \mathbf{A} - \delta\theta \times \mathbf{A}, \quad \delta\mathbf{A} = -\delta\theta \times \mathbf{A} = -i[\delta\theta \cdot \mathbf{J}, \mathbf{A}]. \quad (3.72)$$

En componentes

$$-\varepsilon_{ijk}\delta\theta^j A^k = -i\delta\theta^j [J_j, A^i] \Rightarrow [J_i, A^j] = i\varepsilon_{ijk} A^k. \quad (3.73)$$

La última relación también se puede escribir $[A^i, J_j] = i\varepsilon_{ijk} A^k$. Así en particular \mathbf{J} es un vector. En $L^2(\mathbb{R}^3)$, \mathbf{x} , \mathbf{p} y \mathbf{L} son vectores.

3.5. Grupo SU(2)

3.5.1. Grupos U(n) y SU(n)

Las matrices unitarias forman el grupo U(n) ($n = 1, 2, \dots$)

$$U(n) = \{U, \text{matriz compleja } n \times n, U^\dagger U = 1\}. \quad (3.74)$$

Una matriz U es unitaria sii sus columnas forman una base ortonormal de \mathbb{C}^n . En efecto,

$$(\hat{e}_j)_i \equiv U_{ij}, \quad \delta_{jk} = (U^\dagger U)_{jk} = (U^\dagger)_{ji} U_{ik} = U_{ij}^* U_{ik} = (\hat{e}_j)_i^* (\hat{e}_k)_i = \langle \hat{e}_j | \hat{e}_k \rangle. \quad (3.75)$$

Análogamente la condición equivalente $UU^\dagger = 1$ implica la ortonormalidad por filas.

Los elementos de $U(n)$ son las matrices de cambio de base entre dos bases ortonormales complejas de \mathbb{C}^n .

El subgrupo de matrices unitarias con determinante unidad forman el grupo unitario especial $SU(n)$

$$SU(n) = \{U, \text{ matriz compleja } n \times n, U^\dagger U = 1, \det(U) = 1\}. \quad (3.76)$$

La aplicación $U \mapsto \det(U)$ define un homomorfismo de $U(n)$ en $U(1)$ (por ser U unitaria su determinante es una fase) con núcleo $SU(n)$, de modo que $U(n)/SU(n) \cong U(1)$.²⁵

$U(n)$ es un grupo compacto y conexo pero no simplemente conexo ni simple (contiene un factor $U(1)$). El grupo $SU(n)$ es compacto, conexo y simplemente conexo, y simple.

Los elementos de $U(n)$ se pueden escribir en la forma

$$U = e^{-iX}, \quad U^\dagger = e^{iX^\dagger} = U^{-1} = e^{+iX}, \quad (3.77)$$

y se concluye que su álgebra es²⁶

$$\mathfrak{u}(n) = \{X, \text{ matrices complejas } n \times n, X^\dagger = X\}. \quad (3.78)$$

La dimensión de $U(n)$ o $\mathfrak{u}(n)$ es n^2 , ya que cada matriz hermítica queda especificada por n elementos de matriz reales en la diagonal y $n(n-1)/2$ elementos de matriz complejos por encima de la diagonal, en total $n + 2 \times n(n-1)/2 = n^2$ parámetros reales.

De la **igualdad de Jacobi**, válida para cualquier matriz cuadrada compleja A

$$\det(e^A) = e^{\text{tr}(A)} \quad (3.79)$$

se deduce (de nuevo la multivaluación se puede obviar) que

$$1 = \det(e^{-iX}) = e^{-i\text{tr}(X)} \Leftrightarrow \text{tr}(X) = 0. \quad (3.80)$$

En consecuencia, el álgebra de $SU(n)$ está formada por matrices hermíticas sin traza,

$$\mathfrak{su}(n) = \{X, \text{ matriz compleja } n \times n, X^\dagger = X, \text{tr}(X) = 0\}. \quad (3.81)$$

²⁵Por otro lado, el centro de $U(n)$ es el subgrupo isomorfo a $U(1)$ formado por las matrices $e^{-i\theta} 1$. El centro de $SU(n)$ es \mathbb{Z}_n (formado por las raíces n -ésimas de la unidad) de modo que $U(n)/U(1) \cong SU(n)/\mathbb{Z}_n$.

²⁶Elegimos X por continuidad desde 0 por lo que no afecta la multivaluación $2\pi n$.

En efecto, $\mathfrak{su}(n)$ forma un álgebra. Si $X, Y \in \mathfrak{su}(n)$, $-i[X, Y]$ es otra vez hermítico. Por otro lado para dos matrices $n \times n$ cualesquiera $\text{tr}[A, B] = 0$, por la **propiedad cíclica de la traza**²⁷

$$\text{tr}(AB) = A^i_j B^j_i = \text{tr}(BA). \quad (3.82)$$

Si X es hermítica, $\text{tr}(X)$ es automáticamente real e imponer $\text{tr}(X) = 0$ sólo elimina un parámetro real, en consecuencia, la dimensión de $SU(n)$ o $\mathfrak{su}(n)$ es $n^2 - 1$.

3.5.2. Matrices de $SU(2)$ y matrices de Pauli

No es difícil ver que la matriz más general de $SU(2)$ se puede escribir como²⁸

$$U = \begin{pmatrix} a_0 - ia_3 & -ia_1 - a_2 \\ -ia_1 + a_2 & a_0 + ia_3 \end{pmatrix}, \quad a \in \mathbb{R}^4, \quad \sum_{\mu=0}^3 a_\mu^2 = 1. \quad (3.83)$$

Hay una biyección entre los elementos de $SU(2)$ y los puntos a de la esfera $S^3 = \{a \in \mathbb{R}^4, \|a\| = 1\}$, y ambos espacios son homeomorfos. Esto implica que $SU(2)$ es compacto, conexo y simplemente conexo. De hecho la medida invariante de $SU(2)$ no es más que la medida uniforme sobre S^3 (inducida por la medida de Lebesgue en \mathbb{R}^4) [4].

Las matrices U de (3.83) se pueden escribir en la forma

$$U = a_0 1 - ia_1 \sigma_1 - ia_2 \sigma_2 - ia_3 \sigma_3 = a_0 - ia \cdot \sigma, \quad (3.84)$$

donde σ_i , $i = 1, 2, 3$ son las **matrices de Pauli**,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.85)$$

Estas matrices tienen las siguientes propiedades

$$\sigma_i^\dagger = \sigma_i, \quad \sigma_i \sigma_j = \delta_{ij} + i \varepsilon_{ijk} \sigma_k. \quad (3.86)$$

²⁷Puesto que la traza del conmutador siempre se anula, las matrices sin traza forman un ideal del álgebra de Lie, y las matrices con determinante unidad forman un subgrupo invariante.

²⁸En efecto, si $U = \begin{pmatrix} z_1 & z_3 \\ z_2 & z_4 \end{pmatrix}$ es unitaria, los dos vectores $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ y $\begin{pmatrix} z_3 \\ z_4 \end{pmatrix}$ forman una base ortonormal de \mathbb{C}^2 . Entonces, $|z_1|^2 + |z_2|^2 = |z_3|^2 + |z_4|^2 = 1$ y $z_1^* z_3 + z_2^* z_4 = 0$, ó $z_3 / (-z_2^*) = z_4 / z_1^* \equiv \lambda$. Entonces $U = \begin{pmatrix} z_1 & -\lambda z_2^* \\ z_2 & \lambda z_1^* \end{pmatrix}$ con $|\lambda| = 1$, para que $U \in U(2)$. Si además $1 = \det(U) = \lambda$, se obtiene la forma en ec. (3.83).

La segunda relación equivale a

$$\begin{aligned}\sigma_i^2 &= 1 && \text{(no hay suma en } i), \\ \sigma_i\sigma_j &= -\sigma_j\sigma_i = i\sigma_k && \text{(}ijk\text{ permutación cíclica de }123\text{)}.\end{aligned}\tag{3.87}$$

Estas propiedades implican que $\{\pm 1, \pm i\sigma\}$ es el grupo de cuaterniones y el conjunto

$$Q = \{a_0 - i\mathbf{a} \cdot \boldsymbol{\sigma}, a \in \mathbb{R}^3\}\tag{3.88}$$

es el álgebra de los cuaterniones, introducida por Hamilton.

Para el grupo se obtiene

$$SU(2) = \{U = a_0 - i\mathbf{a} \cdot \boldsymbol{\sigma}, a_0^2 + \mathbf{a}^2 = 1\}.\tag{3.89}$$

El vector \mathbf{a} define un sistema de coordenadas locales en $\|\mathbf{a}\| < 1$ con $a_0 = +\sqrt{1 - \mathbf{a}^2}$ (hemisferio norte –abierto– de S^3). Estas coordenadas no son canónicas.

Es interesante notar que las matrices de Pauli forman un conjunto irreducible ya que no admiten un vector propio común a las tres. En consecuencia las matrices U forman una irrep de $SU(2)$.

3.5.3. Álgebra de $SU(2)$

Tomando un elemento infinitesimal, $U = 1 - i\delta\mathbf{a} \cdot \boldsymbol{\sigma}$, se ve que $\boldsymbol{\sigma}$ forma una base del álgebra de Lie de $SU(2)$ (matrices 2×2 hermíticas sin traza). Es convencional tomar la base reescalada

$$J_i = \frac{1}{2}\sigma_i, \quad \mathbf{J} = \frac{1}{2}\boldsymbol{\sigma}.\tag{3.90}$$

De este modo, usando la identidad $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k$,

$$[J_i, J_j] = i\epsilon_{ijk}J_k,\tag{3.91}$$

que coincide con el álgebra de $SO(3)$.

Exponenciando las matrices del álgebra se reobtiene el grupo $SU(2)$

$$U = e^{-i\boldsymbol{\psi} \cdot \mathbf{J}} = e^{-i\boldsymbol{\psi} \cdot \boldsymbol{\sigma}/2}.\tag{3.92}$$

El vector $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3)$ define las tres coordenadas canónicas para este grupo (asociadas a los generadores \mathbf{J}).

Para calcular explícitamente la exponencial conviene expresar ψ en la forma

$$\psi = \psi \hat{n}, \quad \psi = \|\psi\|, \quad \hat{n} = \frac{\psi}{\psi}. \quad (3.93)$$

Las identidades

$$(\psi \cdot \sigma)^2 = \psi^2, \quad (\psi \cdot \sigma)^n = \begin{cases} \psi^n & n \text{ par} \\ \psi^n \hat{n} \cdot \sigma & n \text{ impar} \end{cases} \quad (3.94)$$

permiten expresar U en función de ψ y \hat{n} :

$$U(\psi) = U(\hat{n}, \psi) = e^{-i\psi \hat{n} \cdot \sigma / 2} = \cos\left(\frac{\psi}{2}\right) - i \operatorname{sen}\left(\frac{\psi}{2}\right) \hat{n} \cdot \sigma \quad (3.95)$$

En consecuencia la relación con las coordenadas a es

$$a = \operatorname{sen}\left(\frac{\psi}{2}\right) \hat{n}, \quad a_0 = \cos\left(\frac{\psi}{2}\right). \quad (3.96)$$

El grupo queda cubierto tomando $\hat{n} \in S^2$ cualquiera y $0 \leq \psi \leq 2\pi$.²⁹ Esto es el doble que para $SO(3)$, que sólo requería $0 \leq \psi \leq \pi$.

Más en detalle, expresando \hat{n} en polares, $\hat{n} = (\operatorname{sen} \theta \cos \varphi, \operatorname{sen} \theta \operatorname{sen} \varphi, \cos \theta)$, se tiene

$$U(\psi) = \begin{pmatrix} \cos\left(\frac{\psi}{2}\right) - i \operatorname{sen}\left(\frac{\psi}{2}\right) \cos(\theta) & -ie^{-i\varphi} \operatorname{sen}\left(\frac{\psi}{2}\right) \operatorname{sen}(\theta) \\ -ie^{i\varphi} \operatorname{sen}\left(\frac{\psi}{2}\right) \operatorname{sen}(\theta) & \cos\left(\frac{\psi}{2}\right) + i \operatorname{sen}\left(\frac{\psi}{2}\right) \cos(\theta) \end{pmatrix}. \quad (3.97)$$

Como caso particular

$$R(\hat{e}_3, \psi) = \begin{pmatrix} \cos \psi & -\operatorname{sen} \psi & 0 \\ \operatorname{sen} \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad U(\hat{e}_3, \psi) = \begin{pmatrix} e^{-i\psi/2} & 0 \\ 0 & e^{i\psi/2} \end{pmatrix}. \quad (3.98)$$

3.5.4. Relación entre los grupos $SU(2)$ y $SO(3)$

Como $SU(2)$ y $SO(3)$ comparten álgebra de Lie abstracta (tienen las mismas constantes de estructura) estos grupos son localmente isomorfos en un entorno del neutro

$$SO(3) \cong_{\text{loc}} SU(2), \quad (3.99)$$

²⁹Y cada elemento del grupo aparece sólo una vez, excepto un conjunto de medida nula.

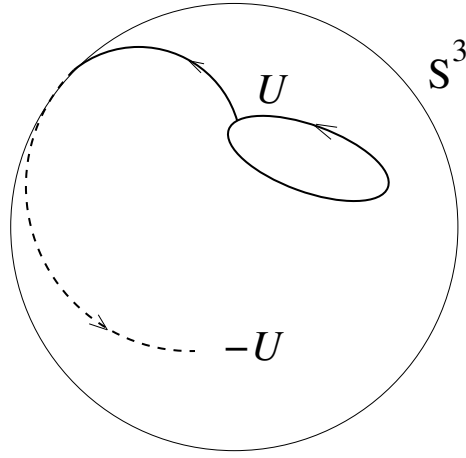


Figura 5: Dos caminos cerrados en $SO(3)$.

además $SU(2)$ es simplemente conexo, lo cual implica que es el grupo recubridor universal de su clase, y $SO(3)$ un grupo cociente. De hecho,

$$SU(2)/\mathbb{Z}_2 \cong SO(3), \quad \mathbb{Z}_2 = \{1, -1\}. \quad (3.100)$$

\mathbb{Z}_2 es el **centro** de $SU(2)$ (más generalmente el centro de $SU(n)$ está formado por las raíces n -ésimas de la unidad y es isomorfo a \mathbb{Z}_n).

Para ver esto basta mostrar que existe un homomorfismo de grupos de $SU(2)$ en $SO(3)$ con núcleo \mathbb{Z}_2 . En efecto, las relaciones de conmutación (3.91) indican que σ es un operador vectorial, por tanto

$$U(\hat{n}, \psi)\sigma U(\hat{n}, \psi)^{-1} = R(\hat{n}, \psi)^{-1}\sigma, \quad (3.101)$$

y el homomorfismo indicado es simplemente

$$\pi : U(\hat{n}, \psi) \mapsto R(\hat{n}, \psi). \quad (3.102)$$

En efecto es un homomorfismo ya que

$$(U_1 U_2)\sigma(U_1 U_2)^{-1} = U_1(R_2^{-1}\sigma)U_1^{-1} = R_2^{-1}R_1^{-1}\sigma = (R_1 R_2)^{-1}\sigma, \quad (3.103)$$

es decir, $\pi(U_1 U_2) = \pi(U_1)\pi(U_2)$. Por otro lado, si U es del núcleo $U\sigma U^{-1} = \sigma$, que implica que U es un múltiplo de la identidad por ser σ un conjunto irreducible, y la condición $\det(U) = 1$ implica $U = \pm 1$.

El grupo $SU(2)$ recubre dos veces a $SO(3)$ ya que (usando las ecs. (3.30) y (3.95))

$$U(\hat{n}, \psi + 2\pi) = -U(\hat{n}, \psi), \quad R(\hat{n}, \psi + 2\pi) = R(\hat{n}, \psi). \quad (3.104)$$

La relación $SO(3) \cong SU(2)/\mathbb{Z}_2$ indica que $SO(3)$ es doblemente conexo. Su espacio topológico es la esfera S^3 con puntos diametralmente opuestos identificados, $U = -U \pmod{\mathbb{Z}_2}$, es decir, $U \cong -U$ para $SO(3)$. Hay una biyección entre rotaciones y diámetros de S^3 (rectas que pasan por el origen en \mathbb{R}^4).

3.5.5. Representaciones enteras y semienteras de $SU(2)$

El centro de $SU(2)$ está formado por $\{1, -1\}$, de modo que si $U \in SU(2)$, también $-U$ está en $SU(2)$. Se denominan representaciones **enteras** de $SU(2)$ aquellas que son funciones pares de U y **semienteras** las que son funciones impares:³⁰

$$\begin{aligned} T(-1) = +1 & \quad (\text{representación entera}) & \quad T(-U) = T(U), \\ T(-1) = -1 & \quad (\text{representación semientera}) & \quad T(-U) = -T(U). \end{aligned} \quad (3.105)$$

Una representación genérica T de $SU(2)$ tendrá una componente entera y otra semientera. Estas componentes se obtienen simplemente proyectando en la componentes par e impar respecto de $U \rightarrow -U$:

$$T = T_+ \oplus T_-, \quad T_+(U) \equiv \frac{1}{2}(T(U) + T(-U)), \quad T_-(U) \equiv \frac{1}{2}(T(U) - T(-U)). \quad (3.106)$$

Es inmediato comprobar que T_{\pm} son también representaciones de $SU(2)$

Se deduce que las representaciones *irreducibles* de $SU(2)$ son necesariamente enteras o semienteras, ya que si tuvieran ambas componentes a la vez se podrían reducir.

Las representaciones **enteras** de $SU(2)$ son a su vez representaciones (**univaluadas**) de $SO(3)$ ($R \mapsto \pm U \mapsto T(U)$), y las **semienteras** son **bivaluadas** para $SO(3)$ ($R \mapsto \pm U \mapsto \pm T(U)$).

Así la representación bidimensional de $SU(2)$, $U = e^{-i\psi\hat{n}\cdot\sigma/2}$ es irreducible y semientera. Considerada como representación de $SO(3)$, $R(\hat{n}, \psi) \mapsto U(\hat{n}, \psi)$, es una representación bivaluada, ya que según se elija ψ o $\psi + 2\pi$ sale un signo de diferencia, aunque la rotación es la misma.

³⁰Para $SU(n)$ el centro está formado por las n n -ésimas raíces complejas de 1, en consecuencia hay n tipos de representaciones, según sea el valor de r en $T(e^{2\pi i/N}) = e^{2\pi i r/N}$, con $r = 0, \dots, n-1$. Todas las demás representaciones son suma directa de éstas.

Por otro lado la representación tridimensional $U \rightarrow R$ es una representación entera de $SU(2)$ e irreducible. Como representación de $SO(3)$, $R \mapsto R$ es univaluada.

En principio $SO(3)$ es el grupo físico y sólo las representaciones univaluadas deberían aparecer, ya que queremos que una rotación de 2π no tenga ningún efecto físico (el sistema se queda como estaba). Sin embargo, en mecánica cuántica el estado está representado por vectores del espacio de Hilbert normalizados *módulo una fase*: $|\psi\rangle$ y $e^{i\varphi}|\psi\rangle$ representan el mismo estado físico. Esto hace que sistemas que caigan en representaciones bivaluadas de $SO(3)$ también son admisibles. Los estados en representaciones enteras son bosónicos y aquellos en representaciones semienteras son fermiónicos. Éstos responden con una fase -1 a una rotación de 2π . Por consistencia, no hay estados físicos que sean suma coherente de estados bosónicos y fermiónicos ya que para estas sumas una rotación de 2π no produciría sólo una fase (**regla de superselección**).³¹

3.6. Representaciones irreducibles de $SU(2)$ y $SO(3)$

3.6.1. Soluciones del álgebra de $SU(2)$

Buscar irreps de $SU(2)$, siendo conexo y simplemente conexo, equivale a buscar irreps de su álgebra de Lie

$$[J_i, J_j] = i\varepsilon_{ijk}J_k. \quad (3.107)$$

Como el grupo es compacto, sus irreps son unitarias y de dimensión finita. Por tanto buscamos las tres matrices \mathbf{J} más generales que sean *matrices hermiticas irreducibles* y que satisfagan las *relaciones de conmutación*, y sólo nos interesan soluciones inequivalentes (es decir, que no difieran sólo por un cambio de base).

Primero definimos el **operador \mathbf{J}^2**

$$\mathbf{J}^2 := \mathbf{J} \cdot \mathbf{J} = \sum_{i=1}^3 J_i^2. \quad (3.108)$$

Nótese que \mathbf{J}^2 no pertenece al álgebra de Lie. Puesto que \mathbf{J} es un vector, \mathbf{J}^2 es un escalar, es decir, satisface

$$[J_i, \mathbf{J}^2] = 0, \quad (3.109)$$

³¹Se deduce que la *paridad* del número de fermiones, que determina si un estado es bosónico o fermiónico, se conserva absolutamente: los fermiones sólo pueden crearse o aniquilarse por pares.

que también se sigue directamente de las relaciones de conmutación:

$$[J_i, J_j J_j] = \{J_j, [J_i, J_j]\} = i\epsilon_{ijk}\{J_j, J_k\} = 0. \quad (3.110)$$

Por ser \mathbf{J} irreducible (por hipótesis) el lema de Schur implica que \mathbf{J}^2 toma un valor constante K en toda la representación y además positivo ya que $\mathbf{J}^2 \geq 0$ por $\mathbf{J} = \mathbf{J}^\dagger$:

$$\mathbf{J}^2 = K, \quad K \geq 0. \quad (3.111)$$

También definimos los **operadores escalera**

$$J_\pm := J_1 \pm iJ_2, \quad J_\pm^\dagger = J_\mp. \quad (3.112)$$

Nótese que estrictamente hablando tampoco pertenecen al álgebra ya que no son hermíticos. Las relaciones de conmutación se puede reexpresar equivalentemente usando J_3 y los operadores escalera

$$[J_3, J_\pm] = \pm J_\pm, \quad [J_+, J_-] = 2J_3. \quad (3.113)$$

Otras relaciones útiles son

$$\mathbf{J}^2 = \frac{1}{2}(J_+ J_- + J_- J_+) + J_3^2 = J_\mp J_\pm + J_3(J_3 \pm 1). \quad (3.114)$$

Lema. Si el vector $|\psi\rangle$ es propio de J_3 con valor propio m , los vectores $J_\pm |\psi\rangle$ son propios de J_3 con valores propios $m \pm 1$ (de ahí el nombre operadores escalera). En efecto,

$$\begin{aligned} J_3 |\psi\rangle &= m |\psi\rangle, \\ J_3 (J_\pm |\psi\rangle) &= (J_\pm J_3 + [J_3, J_\pm]) |\psi\rangle = (J_\pm m \pm J_\pm) |\psi\rangle = (m \pm 1) (J_\pm |\psi\rangle). \end{aligned} \quad (3.115)$$

J_3 es una matriz hermítica, entonces admite una base ortonormal de vectores propios con valores propios reales. Sea j el valor propio *máximo* de J_3 y $|j\rangle$ un vector propio normalizado asociado al autovalor j

$$J_3 |j\rangle = j |j\rangle \quad \langle j | j \rangle = 1 \quad J_3 \leq j. \quad (3.116)$$

Por el lema,

$$J_+ |j\rangle = 0, \quad (3.117)$$

(ya que no hay vectores propios con valor propio $j + 1$). Además

$$\mathbf{J}^2 |j\rangle = j(j+1) |j\rangle, \quad K = j(j+1). \quad (3.118)$$

En efecto, ya que

$$\mathbf{J}^2|j\rangle = (J_-J_+ + J_3(J_3 + 1))|j\rangle = j(j+1)|j\rangle. \quad (3.119)$$

Usamos el siguiente resultado: Dada una representación T de un grupo G en un espacio V , es inmediato que el espacio generado al aplicar el grupo sobre un vector cualquiera $|\psi\rangle$ de V es un subespacio invariante:

$$V_\psi \equiv \text{lin}\{T(g)|\psi\rangle, g \in G\}, \quad T(g)V_\psi = V_\psi. \quad (3.120)$$

Además T irreducible (V es mínimo) sii $V_\psi = V$ para cualquier $|\psi\rangle$ no nulo.³²

Volviendo al problema de las irreps de $SU(2)$, como la representación es irreducible por hipótesis, usando $|j\rangle$ como pivote se genera todo el espacio al aplicar repetidamente \mathbf{J} . Si se aplica J_- repetidamente sobre $|j\rangle$ se obtendrán vectores propios con valores propios $m = j - n$, $n = 0, 1, 2, \dots, n_{\text{máx}}$. Si denotamos por $|m\rangle$ a los vectores normalizados así obtenidos,

$$J_3|m\rangle = m|m\rangle, \quad m = j, j-1, j-2, \dots, j', \quad (3.121)$$

y el último autovalor debe cumplir

$$J_-|j'\rangle = 0, \quad (3.122)$$

dado que el espacio es de dimensión finita.

Hay que comprobar que si aplicamos J_+ sobre estos vectores no se generan otros vectores nuevos. En efecto, usando ec. (3.114),

$$J_+(J_-|m\rangle) = (\mathbf{J}^2 - J_3(J_3 - 1))|m\rangle = (j(j+1) - m(m-1))|m\rangle \propto |m\rangle. \quad (3.123)$$

En definitiva, $\{|m\rangle, m = j, j-1, j-2, \dots, j'\}$ es una base ortonormal del espacio (no hay otros valores propios de J_3 ni éstos están degenerados, y son linealmente independientes por corresponder a autovalores distintos de J_3).

Por otro lado

$$j(j+1)|j'\rangle = \mathbf{J}^2|j'\rangle = (J_+J_- + J_3(J_3 - 1))|j'\rangle = j'(j'-1)|j'\rangle \quad (3.124)$$

³² Nótese que $V_\psi = V$ para algún $|\psi\rangle$ no garantiza que T sea irreducible. Por ejemplo, $U = \begin{pmatrix} e^{i\varphi} & 0 \\ 0 & e^{-i\varphi} \end{pmatrix}$ define una representación de $U(1)$ en $V = \mathbb{C}^2$. Para $|\psi\rangle = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, es inmediato que $V_\psi = \mathbb{C}^2$, aunque la representación es reducible. Sin embargo, si T es completamente descomponible (por ejemplo unitaria) y $|\psi\rangle$ es un vector propio de algún $T(g)$ con autovalor λ no degenerado, sí se deduce que V_ψ es mínimo. En efecto, V se reduce como suma directa de espacios mínimos y $|\psi\rangle$ estará en el subespacio que lleva el autovalor λ .

implica $j'(j' - 1) = j(j + 1)$, con solución $j' = -j$ (la otra solución $j' = j + 1$ está excluida por $j' \leq j$). Además $j \geq j' = -j$ implica $j \geq 0$. Finalmente, $m = j, j - 1, \dots, -j$ indica que la dimensión del espacio es $2j + 1 = 1, 2, 3, \dots$, en consecuencia $j = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots$. Éstos son los únicos valores permitidos para representaciones unitarias irreducibles de $SU(2)$.

Denotamos los estados de la base de la irrep j por $|jm\rangle$. Tenemos

$$\begin{aligned} J_3|jm\rangle &= m|jm\rangle, & \mathbf{J}^2|jm\rangle &= j(j+1)|jm\rangle, & m &= -j, -j+1, \dots, j, \\ \langle jm|jm'\rangle &= \delta_{mm'}, & j &= 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots, & \dim &= 2j+1. \end{aligned} \quad (3.125)$$

Veamos que cada j determina exactamente una irrep inequivalente de $SU(2)$. Al aplicar el operador J_- bajando desde $|jj\rangle$ y normalizando, se tiene

$$J_-|jm\rangle = N_-(j, m)|j, m-1\rangle, \quad N_-(j, m) \geq 0. \quad (3.126)$$

Que la fase sea positiva es una elección (de la fase de $|j, m-1\rangle$), es el **convenio de Condon-Shortley**. Después de fijar las fases relativas la única ambigüedad es una fase global para toda la irrep. La base así elegida es la **base estándar**.

Los elementos de matriz $N_-(j, m)$ están completamente determinados por las relaciones de conmutación:

$$\begin{aligned} \|J_-|jm\rangle\|^2 &= \langle jm|J_+J_-|jm\rangle = N_-(j, m)^2 \\ &= \langle jm|(\mathbf{J}^2 - J_3(J_3 - 1))|jm\rangle = j(j+1) - m(m-1). \end{aligned} \quad (3.127)$$

Es decir,

$$J_-|jm\rangle = +\sqrt{j(j+1) - m(m-1)}|j, m-1\rangle \quad (3.128)$$

Por otro lado

$$\langle j, m+1|J_+|j, m\rangle = \langle j, m|J_-|j, m+1\rangle^* = \sqrt{j(j+1) - (m+1)m} \quad (3.129)$$

Finalmente

$$\begin{aligned} J_\pm|jm\rangle &= N_\pm(j, m)|j, m\pm 1\rangle, \\ N_\pm(j, m) &= \sqrt{j(j+1) - m(m\pm 1)} = \sqrt{(j\mp m)(j\pm m+1)}. \end{aligned} \quad (3.130)$$

En conclusión, la representación está unívocamente determinada por j .³³

³³Es interesante notar que suponiendo sólo que la representación es unitaria y que J_3 admite un vector propio ya se deduce que debe ser de dimensión finita: la relación ec. (3.114) implica que m está acotado ya que K es fijo y $J_\pm J_\mp = J_\mp^\dagger J_\pm$ es definido positivo.

Una vez construidas explícitamente las matrices se puede comprobar que se cumplen las relaciones de conmutación, que la representación es unitaria y que es irreducible. Esto último se puede comprobar usando el lema de Schur, o bien por la observación en la nota a pie de página en 49, dado que el espectro de J_3 es no degenerado y entonces el espacio generado por $|j, j\rangle$ debe ser mínimo.

Aplicando el lema de Schur, supongamos que la matriz A conmuta con \mathbf{J} :

$$\begin{aligned} 0 &= \langle jm' | [J_3, A] | jm \rangle = (m' - m) \langle jm' | A | jm \rangle \Rightarrow \langle jm' | A | jm \rangle = a_m \delta_{m'm}, \\ A | jm \rangle &= a_m | jm \rangle \quad (A \text{ es diagonal}) \\ 0 &= [J_-, A] | jm \rangle = (a_m - a_{m-1}) N_-(j, m) | j, m-1 \rangle \\ &\Rightarrow a_m = a_{m-1} = \dots \equiv a, \quad A = a. \end{aligned} \tag{3.131}$$

A es múltiplo de la identidad y la representación es irreducible.

En $SU(2)$ hay exactamente una irrep por cada dimensión.

$j = 0$ (dim. 1) es la **representación trivial**, $\mathbf{J} = 0$, $\mathbf{J}|\psi\rangle = 0$, corresponde a estados invariantes bajo rotaciones.

$j = 1/2$ (dim. 2) es la irrep formada por el propio $SU(2)$ (**representación fundamental**). Si se calculan los elementos de matriz con las fórmulas previas se obtiene

$$\begin{aligned} J_3 &= \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}, \quad J_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad J_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \\ J_1 &= \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & -\frac{i}{2} \\ \frac{i}{2} & 0 \end{pmatrix}. \end{aligned} \tag{3.132}$$

Es decir, $\mathbf{J} = \frac{1}{2}\boldsymbol{\sigma}$. Los dos estados $|j, m\rangle = |\frac{1}{2}, \pm\frac{1}{2}\rangle$ se representan también

$$|\frac{1}{2}, +\frac{1}{2}\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |\uparrow\rangle, \quad |\frac{1}{2}, -\frac{1}{2}\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |\downarrow\rangle, \tag{3.133}$$

(espín hacia arriba y espín hacia abajo, respectivamente).

$j = 1$ (dim. 3) es la representación formada por las propias matrices de $SO(3)$,

$$J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad J_+ = J_-^\dagger = \begin{pmatrix} 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.134}$$

Los objetos que caen en esta representación $j = 1$ son **vectores** bajo rotaciones.

En las ecs. (3.39),(3.40) se usaba la base cartesiana \hat{e}_i , $i = 1, 2, 3$. Esta es la misma representación pero expresada en la denominada **base esférica**, $\hat{e}_m = |j = 1, m\rangle$, $m = 0, \pm 1$. La base esférica (que es compleja) es la formada por los estados estándar (propios de J_3 y con fases relativas fijadas por el convenio de Condon-Shortley). La relación entre ambas bases es

$$\hat{e}_{\pm 1} = \mp \frac{1}{\sqrt{2}}(\hat{e}_1 \pm i\hat{e}_2) = \mp \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \pm i \\ 0 \end{pmatrix}, \quad \hat{e}_0 = \hat{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (3.135)$$

Las representaciones con $2j$ par/impar son enteras/semienteras, respectivamente. En efecto, tomando por simplicidad $\hat{n} = e_3$ (eje z) se tiene

$$\begin{aligned} U(e_3, \phi)|jm\rangle &= e^{-i\phi J_3}|jm\rangle = e^{-im\phi}|jm\rangle, \\ U(e_3, 2\pi)|jm\rangle &= e^{-2\pi im}|jm\rangle = (-1)^{2m}|jm\rangle = (-1)^{2j}|jm\rangle. \end{aligned} \quad (3.136)$$

Es decir,

$$U(\hat{n}, 2\pi) = (-1)^{2j} = \begin{cases} +1 & j \text{ entero} \\ -1 & j \text{ semientero} \end{cases} \quad (3.137)$$

Puesto que la representación escalar $\psi(\mathbf{x}) \mapsto \psi(R^{-1}\mathbf{x})$ de $\text{SO}(3)$ en $L^2(\mathbb{R}^3)$ es univaluada, se deduce que el operador momento angular orbital $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ sólo tiene valores $j = l$ enteros, $l = 0, 1, 2, \dots$. Así, si tenemos una distribución de carga eléctrica, $\rho(\mathbf{x})$ se podrá descomponer (por ejemplo aplicando operadores de proyección) en suma de funciones con l definido, $\rho(\mathbf{x}) = \sum_{l=0}^{\infty} \rho_l(\mathbf{x})$. Éste es el desarrollo multipolar. La componente $l = 0$ es la contribución monopolar, tiene simetría esférica y lleva toda la carga. La componente $l = 1$ es la contribución dipolar, el momento dipolar eléctrico de la distribución es un vector (es decir, $j = 1$). La componente $l = 2$ lleva el momento cuadrupolar, etc.

De cara a estudiar otros grupos es interesante notar que lo que se ha hecho es usar \mathbf{J}^2 , J_3 como **conjunto completo de operadores compatibles** para clasificar los estados. Esto se basa en la **cadena canónica**

$$\text{SU}(2) \supset \text{U}(1), \quad \text{ó} \quad \text{SO}(3) \supset \text{SO}(2), \quad (3.138)$$

donde $\text{SO}(2)$ está generado por J_3 . Esta cadena es canónica porque $J_3 = \text{diag}(j, j-1, \dots, -j)$ y todos los autovalores (cada valor m es una irrep de $\text{SO}(2)$) son distintos.

\mathbf{J}^2 es un ejemplo de **operador de Casimir**. Para un álgebra de Lie cualquiera, los operadores de Casimir son operadores de su álgebra envolvente universal³⁴ que sean invariantes (es decir, que

³⁴El **álgebra envolvente universal** es el conjunto de polinomios construidos con los generadores.

conmuten con todos los generadores) y algebraicamente independientes.³⁵ Para álgebras semisimples el número de operadores de Casimir coincide con su **rango**: el rango es el número de generadores que conmutan entre sí. SU(2) es un grupo de rango 1 ya que sólo un generador, por ejemplo J_3 , se puede incluir en el conjunto completo de operadores compatibles.

3.6.2. Matrices de las representaciones irreducibles de SU(2) y SO(3)

Las matrices de representación del grupo se pueden obtener mediante

$$U(\hat{\mathbf{n}}, \psi) = e^{-i\psi \cdot \mathbf{J}}, \quad U(\psi)|jm\rangle = D^j(\psi)^m_m |jm'\rangle, \quad (3.139)$$

calculando la exponencial de la matriz $\hat{\mathbf{n}} \cdot \mathbf{J}$, pero es más práctico usar los ángulos de Euler

$$U(\alpha, \beta, \gamma) = e^{-i\alpha J_3} e^{-i\beta J_2} e^{-i\gamma J_3}. \quad (3.140)$$

En la base estándar

$$U(\alpha, \beta, \gamma)|jm\rangle = e^{-i\alpha m'} d^j(\beta)^m_m e^{-i\gamma m} |jm'\rangle, \quad d^j(\beta)^m_m = \langle jm'|e^{-i\beta J_2}|jm\rangle. \quad (3.141)$$

Por ejemplo, para $j = \frac{1}{2}$,

$$d^{\frac{1}{2}}(\beta) = e^{-i\beta \sigma_2/2} = \cos\left(\frac{\beta}{2}\right) - i\sigma_2 \sin\left(\frac{\beta}{2}\right) = \begin{pmatrix} \cos\left(\frac{\beta}{2}\right) & -\sin\left(\frac{\beta}{2}\right) \\ \sin\left(\frac{\beta}{2}\right) & \cos\left(\frac{\beta}{2}\right) \end{pmatrix}, \quad (3.142)$$

$$D^{\frac{1}{2}}(\alpha, \beta, \gamma) = \begin{pmatrix} e^{-i(\alpha+\gamma)/2} \cos\left(\frac{\beta}{2}\right) & -e^{-i(\alpha-\gamma)/2} \sin\left(\frac{\beta}{2}\right) \\ e^{i(\alpha-\gamma)/2} \sin\left(\frac{\beta}{2}\right) & e^{i(\alpha+\gamma)/2} \cos\left(\frac{\beta}{2}\right) \end{pmatrix}.$$

3.7. Armónicos esféricos

Como ya se ha dicho, el grupo SO(3) actúa en $\mathcal{H} = L^2(\mathbb{R}^3)$ mediante $\psi(\mathbf{x}) \mapsto \psi(R^{-1}\mathbf{x})$ y las rotaciones están generadas por el operador momento angular orbital $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ que sólo toma valores enteros $l = 0, 1, 2, \dots$, es decir,

$$\mathcal{H} = \bigoplus_{l=0}^{\infty} \mathcal{H}_l, \quad \mathcal{H}_l = \{ \psi(\mathbf{x}), \mathbf{L}^2|\psi\rangle = l(l+1)|\psi\rangle \}. \quad (3.143)$$

³⁵Independientes quiere decir sólo se consideran de Casimir los bloques básicos a partir de los cuales se forman todos los demás operadores invariantes.

Cada uno de estos espacios \mathcal{H}_l es reducible, es la suma directa de infinitos espacios mínimos de tipo l , por ejemplo, $\mathcal{H}_{l=0}$ es el espacio de todas las funciones radiales $\psi(\mathbf{x}) = f(r)$. Una forma práctica de proceder es usar coordenadas esféricas, (r, θ, ϕ) ó equivalentemente $(r, \hat{\mathbf{x}})$, siendo

$$r = \|\mathbf{x}\|, \quad \hat{\mathbf{x}} = \frac{\mathbf{x}}{r}. \quad (3.144)$$

Por separación de variables las funciones se pueden escribir en la forma

$$\psi(\mathbf{x}) = \sum_{\nu} R_{\nu}(r) A_{\nu}(\hat{\mathbf{x}}). \quad (3.145)$$

Las rotaciones sólo actúan sobre la parte angular. De hecho los operadores L_z y L^2 en coordenadas esféricas son de la forma

$$L_z = -i \frac{\partial}{\partial \phi}, \quad L^2 = -\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \sin \theta \frac{\partial}{\partial \theta} - \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2}, \quad (3.146)$$

y la coordenada radial no interviene (igualmente para L_x y L_y).

El espacio de funciones tridimensionales se puede expresar como un producto tensorial de funciones radiales, \mathcal{H}_R , y funciones angulares, \mathcal{H}_A . El producto escalar de $L^2(\mathbb{R}^3)$ tiene la forma

$$\|\psi\|^2 = \int d^3x |\psi(\mathbf{x})|^2 = \int_0^{\infty} dr r^2 \int_{S^2} d\Omega |\psi(r, \hat{\mathbf{x}})|^2, \quad (3.147)$$

por tanto los productos escalares en \mathcal{H}_R y \mathcal{H}_A se pueden elegir como

$$\|R\|^2 = \int_0^{\infty} dr r^2 |R(r)|^2, \quad \|A\|^2 = \int_{S^2} d\Omega |A(\hat{\mathbf{x}})|^2 = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin \theta |A(\hat{\mathbf{x}})|^2. \quad (3.148)$$

Se tiene entonces que

$$\mathcal{H} = L^2(\mathbb{R}^3, d^3x) = \mathcal{H}_R \otimes \mathcal{H}_A, \quad \mathcal{H}_R = L^2(\mathbb{R}^+, dr r^2), \quad \mathcal{H}_A = L^2(S^2, d\Omega). \quad (3.149)$$

Las rotaciones son la identidad en \mathcal{H}_R .

\mathcal{H}_A es el espacio de funciones definidas sobre la esfera unitaria S^2 . Este espacio se reduce bajo rotaciones $\mathcal{H}_A = \bigoplus_{l=0}^{\infty} \mathcal{H}_{A,l}$, y cada $\mathcal{H}_{A,l}$ ya es mínimo. Equivalentemente, L^2 y L_z forman un conjunto completo para \mathcal{H}_A . En el caso de \mathcal{H} hay que añadir otro operador escalar bajo rotaciones, tal como \mathbf{P}^2 , para tener un conjunto completo de operadores compatibles.

Las funciones propias normalizadas de L^2 y L_z en \mathcal{H}_A se denominan **armónicos esféricos**, $Y_{lm}(\hat{\mathbf{x}})$,

$$\begin{aligned} L^2 Y_{lm}(\hat{\mathbf{x}}) &= l(l+1)Y_{lm}(\hat{\mathbf{x}}), \\ L_z Y_{lm}(\hat{\mathbf{x}}) &= mY_{lm}(\hat{\mathbf{x}}) \end{aligned}, \quad m = l, l-1, \dots, -l, \quad l = 0, 1, 2, \dots \quad (3.150)$$

Los armónicos esféricos de orden más bajo son

$$Y_{0,0} = \sqrt{\frac{1}{4\pi}}, \quad Y_{1,0} = \sqrt{\frac{3}{4\pi}} \cos \theta, \quad Y_{1,\pm 1} = \mp \sqrt{\frac{3}{8\pi}} \sin \theta e^{\pm i\phi}. \quad (3.151)$$

Los armónicos esféricos forman una base ortonormal de las funciones angulares. Además, para un l dado, los Y_{lm} forman una base estándar de $\mathcal{H}_{A,l}$. Esta condición determina los Y_{lm} de forma única salvo por una fase relativa entre distintos valores de l . El convenio que se usa es $Y_{l,0}(0,0) > 0$.

Una función cualquiera de $L^2(\mathbb{R}^3)$ se puede escribir en la forma

$$\psi(\mathbf{x}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l R_{lm}(r) Y_{lm}(\hat{\mathbf{x}}), \quad R_{lm}(r) = \int_{S^2} d\Omega Y_{lm}^*(\hat{\mathbf{x}}) \psi(\mathbf{x}). \quad (3.152)$$

Si se elige una base $R_n(r)$ en el espacio de funciones radiales, \mathcal{H}_R ,

$$\psi(\mathbf{x}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_n \psi_{nlm} R_n(r) Y_{lm}(\hat{\mathbf{x}}). \quad (3.153)$$

En la práctica es frecuente tomar una base radial *distinta* para cada valor de l , $R_{nl}(r)$.

3.8. Serie de Clebsch-Gordan de SU(2)

3.8.1. Suma de momentos angulares

Tenemos el siguiente resultado general: Si V_μ y V_ν son dos espacios con irreps μ y ν de un grupo

$$V_\mu \otimes V_\nu = \bigoplus_{\lambda} \bigoplus_{\alpha=1}^{a_\lambda} V_{\alpha\lambda} \equiv \bigoplus_{\lambda} a_\lambda V_\lambda \quad (\text{Serie de Clebsch-Gordan}), \quad (3.154)$$

donde $a_\lambda = \langle \lambda | \mu, \nu \rangle = 0, 1, 2, \dots$ es la multiplicidad de la irrep λ al reducir $\mu \otimes \nu$. Correspondientemente, para los caracteres

$$\chi^\mu(g) \chi^\nu(g) = \sum_{\lambda} \langle \lambda | \mu, \nu \rangle \chi^\lambda(g). \quad (3.155)$$

La descomposición (es decir el valor de las multiplicidades $\langle \lambda | \mu, \nu \rangle$) queda unívocamente determinada ya que los caracteres forman una base ortogonal en el espacio de clases de conjugación.

En el caso de $SU(2)$ el ángulo ψ determina la clase de conjugación. En efecto, para $U_1 \in SU(2)$ cualquiera y R_1 su representación en $SO(3)$,

$$U_1 e^{-i\psi \cdot J} U_1^{-1} = e^{-iU_1 \psi \cdot J} U_1^{-1} = e^{-i\psi \cdot (R_1^{-1} J)} = e^{-i(R_1 \psi) \cdot J}. \quad (3.156)$$

Por tanto, los elementos con coordenadas ψ y $R_1 \psi$ son conjugados. Aplicando un R_1 adecuado cualquier rotación de ángulo ψ y eje \hat{n} se puede llevar a otra de igual ángulo y eje z . Se deduce que ψ caracteriza la clase de conjugación.

Esta observación nos permite calcular los caracteres eligiendo rotaciones alrededor del eje z ,

$$\chi^j(\psi) = \text{tr} D^j(e_3, \psi) = \sum_{m=-j}^j e^{-i\psi m} = \frac{\text{sen}((j + \frac{1}{2})\psi)}{\text{sen}(\frac{1}{2}\psi)}, \quad (3.157)$$

donde se ha usado la identidad $\sum_{n=a}^b x^n = \frac{x^{b+1} - x^a}{x - 1} = \frac{x^{b+1/2} - x^{a-1/2}}{x^{1/2} - x^{-1/2}}$.

A partir de este resultado, y haciendo uso de ec. (3.155), veamos que

$$V^{j_1} \otimes V^{j_2} = \bigoplus_{j=|j_1-j_2|}^{j_1+j_2} V^j \quad (\text{Serie de Clebsch-Gordan de } SU(2)), \quad (3.158)$$

es decir, en la reducción de $j_1 \otimes j_2$

$$j = j_{\text{mín}}, j_{\text{mín}} + 1, \dots, j_{\text{máx}}, \quad j_{\text{mín}} = |j_1 - j_2|, \quad j_{\text{máx}} = j_1 + j_2, \quad (3.159)$$

y cada irrep j de la serie aparece exactamente una vez. En total $2 \text{mín}(j_1, j_2) + 1$ irreps. Como puede comprobarse esto es consistente con el cálculo de dimensiones:

$$(2j_1 + 1)(2j_2 + 1) = \sum_{j=|j_1-j_2|}^{j_1+j_2} (2j + 1). \quad (3.160)$$

Para comprobar las multiplicidades basta verlo a nivel de caracteres. Como la serie de C-G es

simétrica ($\langle \lambda | \mu, \nu \rangle = \langle \lambda | \nu, \mu \rangle$) podemos suponer $j_1 \geq j_2$

$$\begin{aligned}
\sum_{j=j_{\min}}^{j_{\max}} \chi^j(\psi) &= \text{Im} \frac{\sum_j e^{i(j+\frac{1}{2})\psi}}{\text{sen}(\frac{\psi}{2})} = \text{Im} \frac{e^{i\frac{\psi}{2}}(e^{i(j_{\max}+\frac{1}{2})\psi} - e^{i(j_{\min}-\frac{1}{2})\psi})}{\text{sen}(\frac{\psi}{2})(e^{i\frac{\psi}{2}} - e^{-i\frac{\psi}{2}})} \\
&= \text{Im} \frac{e^{i(j_1+j_2+1)\psi} - e^{i(j_1-j_2)\psi}}{\text{sen}(\frac{\psi}{2})2i \text{sen}(\frac{\psi}{2})} = -\frac{\cos((j_1+j_2+1)\psi) - \cos((j_1-j_2)\psi)}{2 \text{sen}^2(\frac{\psi}{2})} \quad (3.161) \\
&= \frac{\text{sen}((j_1+\frac{1}{2})\psi) \text{sen}((j_2+\frac{1}{2})\psi)}{\text{sen}(\frac{\psi}{2}) \text{sen}(\frac{\psi}{2})} = \chi^{j_1}(\psi) \chi^{j_2}(\psi).
\end{aligned}$$

En el último paso se ha usado la identidad $\cos(\alpha + \beta) - \cos(\alpha - \beta) = -2 \text{sen}(\alpha) \text{sen}(\beta)$. La relación en ec. (3.160) corresponde al caso particular $\psi = 0$.

Un caso especial interesante es

$$(j_1 = 1) \otimes (j_2 = 1) = (j = 0) \oplus (j = 1) \oplus (j = 2), \quad 3 \times 3 = 1 + 3 + 5. \quad (3.162)$$

Indica que multiplicando dos vectores se puede construir un escalar y un vector (el producto escalar y el producto vectorial) y otro objeto con $j = 2$, que es un tensor de rango 2 sin traza (5 grados de libertad)

$$A^i B^j = \frac{1}{3} \delta_{ij} \mathbf{A} \cdot \mathbf{B} + \frac{1}{2} \varepsilon_{ijk} (\mathbf{A} \times \mathbf{B})^k + \frac{1}{2} (A^i B^j + A^j B^i - \frac{2}{3} \delta_{ij} \mathbf{A} \cdot \mathbf{B}). \quad (3.163)$$

3.8.2. Coeficientes de Clebsch-Gordan

Notemos que el generador de $SU(2)$ en $V^{j_1} \otimes V^{j_2}$ es

$$\mathbf{J}_{\text{tot}} = \mathbf{J}_1 \otimes \mathbf{1} + \mathbf{1} \otimes \mathbf{J}_2 \equiv \mathbf{J}_1 + \mathbf{J}_2 \quad (\text{Momento angular total}). \quad (3.164)$$

La serie de C-G también se puede obtener contando estados. Si $|j_1, m_1\rangle$ y $|j_2, m_2\rangle$ son bases estándar de V^{j_1} y V^{j_2} (únicas salvo fase global), su producto directo

$$|j_1, m_1\rangle \otimes |j_2, m_2\rangle \quad (\text{base desacoplada}), \quad (3.165)$$

es una base de $V^{j_1} \otimes V^{j_2}$ denominada **base desacoplada**. La dimensión de $V^{j_1} \otimes V^{j_2}$ es $(2j_1 + 1)(2j_2 + 1)$.

Para estudiar la reducción notemos que los estados desacoplados son propios de J_3^{tot} con valor propio $m = m_1 + m_2$,

$$J_3^{\text{tot}} |j_1, m_1\rangle \otimes |j_2, m_2\rangle = ((J_1)_3 + (J_2)_3) |j_1, m_1\rangle \otimes |j_2, m_2\rangle = (m_1 + m_2) |j_1, m_1\rangle \otimes |j_2, m_2\rangle. \quad (3.166)$$

Veamos un ejemplo con $j_1 = 1$, $j_2 = \frac{3}{2}$. En la tabla se recogen los valores de $m = m_1 + m_2$ por cada estado de la base desacoplada

$m = m_1 + m_2$		m_2			
		$-3/2$	$-1/2$	$1/2$	$3/2$
m_1	1	$-1/2$	$1/2$	$3/2$	$5/2$
	0	$-3/2$	$-1/2$	$1/2$	$3/2$
	-1	$-5/2$	$-3/2$	$-1/2$	$1/2$

El estado con m máximo tiene $m = 5/2$ y este valor es único. Se deduce que en el espacio producto no hay valores de $j > 5/2$ y además el valor $j = 5/2$ aparece exactamente una vez. Esa entrada en la tabla corresponde a un estado $|j = 5/2, m = 5/2\rangle$. Por tanto hay un estado $|j = 5/2, m = 3/2\rangle$ que corresponde a cierta combinación lineal de las dos entradas $3/2$. Este estado se obtienen aplicando J_- al $m = 5/2$. La combinación ortogonal a esta debe corresponder a $|j = 3/2, m = 3/2\rangle$ ($m = 3/2$ implica $j \geq 3/2$ pero ya no hay más estados con $j = 5/2$). Igualmente hay tres estados con $m = 1/2$, una combinación de ellos es $|j = 5/2, m = 1/2\rangle$ y otra (ortogonal) será $|j = 3/2, m = 1/2\rangle$. Entonces la tercera combinación ortogonal a las otras dos será $|j = 1/2, m = 1/2\rangle$. Y esto ya da cuenta de los valores con m negativo:

$$\begin{aligned}
 m = \frac{5}{2} & \quad |j = \frac{5}{2}, m = \frac{5}{2}\rangle \\
 m = \frac{3}{2} & \quad |j = \frac{5}{2}, m = \frac{3}{2}\rangle, \quad |j = \frac{3}{2}, m = \frac{3}{2}\rangle \\
 m = \frac{1}{2} & \quad |j = \frac{5}{2}, m = \frac{1}{2}\rangle, \quad |j = \frac{3}{2}, m = \frac{1}{2}\rangle, \quad |j = \frac{1}{2}, m = \frac{1}{2}\rangle \\
 m = -\frac{1}{2} & \quad |j = \frac{5}{2}, m = -\frac{1}{2}\rangle, \quad |j = \frac{3}{2}, m = -\frac{1}{2}\rangle, \quad |j = \frac{1}{2}, m = -\frac{1}{2}\rangle \\
 m = -\frac{3}{2} & \quad |j = \frac{5}{2}, m = -\frac{3}{2}\rangle, \quad |j = \frac{3}{2}, m = -\frac{3}{2}\rangle \\
 m = -\frac{5}{2} & \quad |j = \frac{5}{2}, m = -\frac{5}{2}\rangle
 \end{aligned} \tag{3.167}$$

Implica

$$\begin{aligned}
 (j_1 = 1) \otimes (j_2 = 3/2) &= (j = 1/2) \oplus (j = 3/2) \oplus (j = 5/2) \\
 3 \times 4 &= 2 + 4 + 6.
 \end{aligned} \tag{3.168}$$

Los estados $|jm\rangle$ forman la **base acoplada** de $V^{j_1} \otimes V^{j_2}$.

Para un grupo cualquiera (compacto) los **coeficientes de Clebsch-Gordan** relacionan las bases acoplada y desacoplada

$$\begin{aligned}
 |\alpha\lambda k\rangle &= \sum_{i,j} (\mu i \nu j | \alpha\lambda k) |\mu, i\rangle \otimes |\nu, j\rangle \\
 |\mu, i\rangle \otimes |\nu, j\rangle &= \sum_{\alpha,\lambda,k} (\alpha\lambda k | \mu i \nu j) |\alpha\lambda k\rangle.
 \end{aligned} \tag{3.169}$$

α distingue entre las $\langle \lambda | \mu, \nu \rangle$ copias de la irrep λ en la serie de C-G de $\mu \otimes \nu$.³⁶

En particular para $SU(2)$ no hace falta la etiqueta α ya que la multiplicidad es cero o uno ($SU(2)$ es un grupo **simplemente reducible**),

$$\begin{aligned} |j_1 j_2; j, m\rangle &= \sum_{m_1, m_2} C(j_1, j_2, j; m_1, m_2, m) |j_1, m_1\rangle \otimes |j_2, m_2\rangle, \\ |j_1, m_1\rangle \otimes |j_2, m_2\rangle &= \sum_{j, m} C(j_1, j_2, j; m_1, m_2, m) |j_1, j_2; j, m\rangle. \end{aligned} \quad (3.170)$$

Los coeficientes de Clebsch-Gordan de $SU(2)$ son reales y ortogonales. Las sumas indicadas toman sus recorridos naturales ya que el coeficiente se anula en otro caso por definición (por ejemplo, si $m \neq m_1 + m_2$).³⁷

Los coeficientes de C-G se pueden calcular explícitamente usando operadores escalera y ortogonalidad. Veamos un ejemplo para $j_1 = j_2 = 1/2$, $j = 0, 1$. Dos partículas de espín $\frac{1}{2}$ acopladas a $j = 0$ están en estado **singlete** de espín, y acopladas a $j = 1$ en estado **triplete**.

		m_2	
	$m = m_1 + m_2$	-1/2	1/2
	1/2	0	1
m_1	-1/2	-1	0

En primer lugar

$$|\frac{1}{2}, \frac{1}{2}\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle = |\frac{1}{2}, \frac{1}{2}; 1, 1\rangle. \quad (3.171)$$

Construimos el estado $|\frac{1}{2}, \frac{1}{2}; 1, 0\rangle$ aplicando los operadores escalera. Usamos

$$J_- |\frac{1}{2}, \frac{1}{2}\rangle = |\frac{1}{2}, -\frac{1}{2}\rangle, \quad J_- |\frac{1}{2}, -\frac{1}{2}\rangle = 0, \quad J_- |1, 1\rangle = \sqrt{2} |1, 0\rangle, \quad J_-^{\text{tot}} = J_{1-} + J_{2-}. \quad (3.172)$$

Entonces

$$\begin{aligned} J_-^{\text{tot}} |\frac{1}{2}, \frac{1}{2}; 1, 1\rangle &= \sqrt{2} |\frac{1}{2}, \frac{1}{2}; 1, 0\rangle \\ &= (J_{1-} + J_{2-}) |\frac{1}{2}, \frac{1}{2}\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle = |\frac{1}{2}, -\frac{1}{2}\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle + |\frac{1}{2}, \frac{1}{2}\rangle \otimes |\frac{1}{2}, -\frac{1}{2}\rangle \end{aligned} \quad (3.173)$$

³⁶Como las representaciones son unitarias, las bases son ortonormales, por tanto la matriz de cambio de base entre bases acoplada y desacoplada es unitaria y $(\mu i \nu j | \alpha \lambda k) = (\alpha \lambda k | \mu i \nu j)^*$.

³⁷Los coeficientes de C-G sirven igual en los dos sentidos (acoplar o desacoplar). Esto es consecuencia de que los coeficientes de C-G definen una matriz ortogonal ya que pasan de una base ortonormal a otra y son reales. Para matriz una ortogonal, $R^{-1} = R^T$, si $e'_i = R_{ia} e_a$, también $e_a = R_{ia} e'_i$.

Es decir,

$$|\frac{1}{2}, \frac{1}{2}; 1, 0\rangle = \frac{1}{\sqrt{2}} (|\frac{1}{2}, \frac{1}{2}\rangle \otimes |\frac{1}{2}, -\frac{1}{2}\rangle + |\frac{1}{2}, -\frac{1}{2}\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle). \quad (3.174)$$

Nótese que el vector sale correctamente normalizado, y ésta es una comprobación del cálculo.

Aplicando de nuevo J_-^{tot} se obtiene

$$|\frac{1}{2}, \frac{1}{2}; 1, -1\rangle = |\frac{1}{2}, -\frac{1}{2}\rangle \otimes |\frac{1}{2}, -\frac{1}{2}\rangle. \quad (3.175)$$

El estado con $j = 0$, $|\frac{1}{2}, \frac{1}{2}, 00\rangle$, se obtiene por ortogonalidad con $|\frac{1}{2}, \frac{1}{2}; 1, 0\rangle$,

$$|\frac{1}{2}, \frac{1}{2}; 0, 0\rangle = \frac{1}{\sqrt{2}} (|\frac{1}{2}, \frac{1}{2}\rangle \otimes |\frac{1}{2}, -\frac{1}{2}\rangle - |\frac{1}{2}, -\frac{1}{2}\rangle \otimes |\frac{1}{2}, \frac{1}{2}\rangle). \quad (3.176)$$

Al reducir el espacio, por cada nuevo valor de j hay una ambigüedad de fase que generalmente se fija con el convenio

$$C(j_1, j_2, j; j_1, j - j_1, j) > 0. \quad (3.177)$$

Es decir, en el desarrollo de $|j_1, j_2; j, j\rangle$ en la base desacoplada, el coeficiente del estado $|j_1, j_1\rangle \otimes |j_2, j - j_1\rangle$ debe ser real y positivo. Siguiendo este convenio en el caso anterior el signo de $|\frac{1}{2}, \frac{1}{2}; 0, 0\rangle$ se ha tomado de modo que la componente de $|\frac{1}{2}, \frac{1}{2}\rangle \otimes |\frac{1}{2}, -\frac{1}{2}\rangle$ sea positiva. La fase de $|\frac{1}{2}, \frac{1}{2}; 1, 1\rangle$ en ec. (3.171) se ha fijado con este convenio. Nótese que la necesidad de introducir convenios de fases hace que en general se obtengan distintos coeficientes de Clebsch-Gordan al acoplar $V_{j_1} \otimes V_{j_2}$ y $V_{j_2} \otimes V_{j_1}$ cuando $j_1 \neq j_2$.

Usando la notación frecuente $|\uparrow\rangle \equiv |\frac{1}{2}, \frac{1}{2}\rangle$, $|\downarrow\rangle \equiv |\frac{1}{2}, -\frac{1}{2}\rangle$, las relaciones anteriores quedarían

$$\begin{aligned} |1, 1\rangle &= |\uparrow\uparrow\rangle, & |1, 0\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle), & |1, -1\rangle &= |\downarrow\downarrow\rangle, & \text{(triplete)} \\ |0, 0\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle), & & & & & \text{(singlete)} \end{aligned} \quad (3.178)$$

Este resultado es consistente con la teoría general de reducción de tensores bajo $GL(n, \mathbb{C})$. En este caso, $j_1 = j_2 = \frac{1}{2}$, tenemos biespinores $n = 2$. $V_2 \otimes V_2$ se reduce en dos espacios, el simétrico [2] y el antisimétrico [1²], el primero con dimensión 3 es el triplete $j = 1$, y el segundo con dimensión 1 es el singlete, $j = 0$.

4. Grupo de Poincaré

4.1. Transformaciones de Lorentz

Cada evento espacio-temporal tiene una coordenada $(t, \mathbf{x}) \in \mathbb{R}^4$ (**espacio-tiempo**) para un observador inercial dado A . Para un observador B que se mueva con una velocidad relativa \mathbf{v} , en la teoría **no relativista** las coordenadas (t', \mathbf{x}') estarán relacionadas mediante

$$\mathbf{x} = R\mathbf{x}' + \mathbf{v}t' + \mathbf{a}, \quad t = t' + \tau \quad (\text{transformación de Galileo}) \quad (4.1)$$

(R es una rotación y también permitimos una traslación en tiempo y espacio) de modo que la posición de B , $(t', \mathbf{x}' = 0)$ en su propio sistema, vista en A es

$$(t', \mathbf{x}' = 0), \quad \mathbf{x} = \mathbf{v}(t - \tau) + \mathbf{a} \quad (4.2)$$

que expresa que B se mueve con velocidad \mathbf{v} ($d\mathbf{x}/dt = \mathbf{v}$).

La transformación indicada en ec. (4.1) es una **transformación de Galileo**. Forman el **grupo de Galileo** y la ley de composición de velocidades es simplemente

$$\mathbf{v}_{12} = \mathbf{v}_1 + \mathbf{v}_2. \quad (4.3)$$

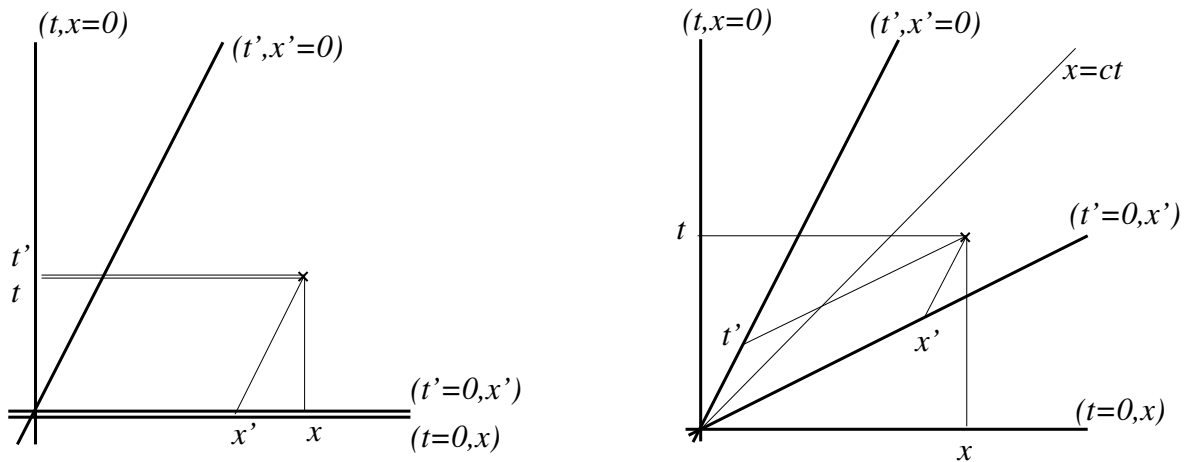


Figura 6: Transformaciones galileanas (izquierda) y relativistas (derecha).

Para simplificar, vamos a considerar 1 + 1 dimensiones y $\alpha = \tau = 0$, es decir, elegimos el origen de coordenadas de modo que coinciden en los dos sistemas (ver Fig. 6)

$$x = x' + vt', \quad t = t'. \quad (4.4)$$

Esta relación no es compatible con el postulado de la invariancia de la velocidad de la luz en cualquier sistema inercial. Para ello hay que permitir que el tiempo no sea absoluto, $t \neq t'$:

$$x = ax' + bt', \quad t = ex' + ft'. \quad (4.5)$$

Podemos suponer $a > 0$ (los dos observadores eligen igualmente orientadas la coordenada x) y $f > 0$ (los dos relojes avanzan en el tiempo). Ahora podemos imponer la condición de que una señal luminosa se mueva con velocidad c en ambos sistemas hacia la derecha o hacia la izquierda, y también que A vea a B moverse con velocidad v

$$x = ct \Leftrightarrow x' = ct', \quad x = -ct \Leftrightarrow x' = -ct', \quad x' = 0 \Leftrightarrow x = vt. \quad (4.6)$$

Al imponer estas tres condiciones sobre los parámetros a, b, e, f en (4.5) sólo queda un parámetro libre y la transformación se puede escribir

$$\begin{cases} x = \lambda \gamma (x' + vt') \\ t = \lambda \gamma (t' + \frac{v}{c^2} x') \end{cases} \quad \begin{cases} x' = \lambda^{-1} \gamma (x - vt) \\ t' = \lambda^{-1} \gamma (t - \frac{v}{c^2} x) \end{cases} \quad \gamma \equiv (1 - v^2/c^2)^{-1/2}, \quad \lambda > 0. \quad (4.7)$$

Implica $|v| < c$ ($\gamma > 1$). Según estas relaciones el observador A ve el reloj de B ($x' = 0$) cambiando a un ritmo $dt'/dt = \lambda^{-1} \gamma^{-1}$, B el de A ($x = 0$) a un ritmo $dt/dt' = \lambda \gamma^{-1}$. Para que ambos observadores sean equivalentes (no haya observadores inerciales privilegiados) se requiere $\lambda = 1$. Así se obtienen las **transformaciones de Lorentz**:

$$\begin{cases} x = \gamma (x' + vt') \\ t = \gamma (t' + \frac{v}{c^2} x') \end{cases} \quad \begin{cases} x' = \gamma (x - vt) \\ t' = \gamma (t - \frac{v}{c^2} x) \end{cases} \quad \gamma \equiv (1 - v^2/c^2)^{-1/2}. \quad (4.8)$$

Es conveniente usar coordenadas homogéneas para espacio y tiempo, por lo que se usa ct en vez de t , así, en forma matricial

$$\begin{pmatrix} ct \\ x \end{pmatrix} = \begin{pmatrix} \gamma & \gamma v \\ \gamma v & \gamma \end{pmatrix} \begin{pmatrix} ct' \\ x' \end{pmatrix}, \quad \begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma v \\ -\gamma v & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix}. \quad (4.9)$$

Nótese que estas matrices no son unitarias. Componiendo dos transformaciones sucesivas se obtiene la ley relativista de suma de velocidades

$$\begin{pmatrix} \gamma_{12} & \gamma_{12} v_{12} \\ \gamma_{12} v_{12} & \gamma_{12} \end{pmatrix} = \begin{pmatrix} \gamma_1 & \gamma_1 v_1 \\ \gamma_1 v_1 & \gamma_1 \end{pmatrix} \begin{pmatrix} \gamma_2 & \gamma_2 v_2 \\ \gamma_2 v_2 & \gamma_2 \end{pmatrix}, \quad v_{12} = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}}. \quad (4.10)$$

Se deduce que v no es una coordenada canónica, la coordenada canónica es $\xi = \tanh^{-1}(v/c)$, de modo que $\xi_{12} = \xi_1 + \xi_2$.

Más generalmente, en 3 + 1 dimensiones, descomponiendo \mathbf{x} y \mathbf{x}' según la dirección de \mathbf{v} ,

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_{\parallel} + \mathbf{x}_{\perp}, & \mathbf{x}' &= \mathbf{x}'_{\parallel} + \mathbf{x}'_{\perp}, \\ \mathbf{x}_{\parallel} &= \gamma(\mathbf{x}'_{\parallel} + \mathbf{v}t'), & \mathbf{x}_{\perp} &= \mathbf{x}'_{\perp}, & t &= \gamma(t' + \frac{\mathbf{v}}{c^2} \cdot \mathbf{x}'). \end{aligned} \quad (4.11)$$

Esta transformación es un **boost** de velocidad \mathbf{v} (o **transformación de Lorentz pura**).

Los boosts conservan el **intervalo**

$$s^2 \equiv \mathbf{x}^2 - c^2 t^2 = \gamma^2(\mathbf{x}'_{\parallel} + \mathbf{v}t')^2 + \mathbf{x}'_{\perp}{}^2 - c^2 \gamma^2(t' + \frac{\mathbf{v}}{c^2} \cdot \mathbf{x}')^2 = \mathbf{x}'^2 - c^2 t'^2. \quad (4.12)$$

El intervalo también es conservado por las rotaciones

$$(t, \mathbf{x}) \mapsto (t, R\mathbf{x}). \quad (4.13)$$

4.2. Grupo de Lorentz

Usamos la notación

$$x \in \mathbb{R}^4, \quad x^{\mu} = (ct, \mathbf{x}), \quad \mu = 0, 1, 2, 3. \quad (4.14)$$

El intervalo puede escribirse como

$$\begin{aligned} s^2 &= -(x^0)^2 + \mathbf{x}^2 = -(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2 \equiv g_{\mu\nu} x^{\mu} x^{\nu}, \\ g_{\mu\nu} &\equiv \text{diag}(-1, +1, +1, +1). \end{aligned} \quad (4.15)$$

El tensor $g_{\mu\nu}$ es la **métrica de Minkowski**. En la literatura también se usa con mucha frecuencia la signatura $(+, -, -, -)$. \mathbb{R}^4 con la métrica de Minkowski es el **espacio de Minkowski**,

$$x \cdot y = g_{\mu\nu} x^{\mu} y^{\nu} = \mathbf{x} \cdot \mathbf{y} - x^0 y^0, \quad \|x\|^2 \equiv x^2 = g_{\mu\nu} x^{\mu} x^{\nu}. \quad (4.16)$$

Por definición el **grupo de Lorentz**, L , son las transformaciones lineales (en realidad no hay otras) que dejan invariante el intervalo, o equivalentemente el producto escalar, o la métrica.³⁸ Si representamos una transformación lineal en el espacio de Minkowski mediante

$$x' = \Lambda x, \quad x'^{\mu} = \Lambda^{\mu}_{\nu} x^{\nu}, \quad (4.17)$$

³⁸La conservación de la norma equivale a la conservación del producto escalar, por $\|x+y\|^2 - \|x-y\|^2 = 4x \cdot y$.

la conservación del producto escalar implica

$$x \cdot y = g_{\mu\nu} x^\mu y^\nu = x' \cdot y' = g_{\alpha\beta} x'^\alpha y'^\beta = g_{\alpha\beta} \Lambda^\alpha_\mu x^\mu \Lambda^\beta_\nu y^\nu, \quad (4.18)$$

es decir,

$$\Lambda \in L \Leftrightarrow g_{\mu\nu} = g_{\alpha\beta} \Lambda^\alpha_\mu \Lambda^\beta_\nu, \quad (4.19)$$

o en notación matricial

$$(G)_{\mu\nu} \equiv g_{\mu\nu} \quad (\Lambda)_{\mu\nu} \equiv \Lambda^\mu_\nu \quad G = \Lambda^T G \Lambda. \quad (4.20)$$

Así

$$L = O(3,1) = \{\Lambda, \text{matrices reales } 4 \times 4, G = \Lambda^T G \Lambda\} \quad (\text{grupo de Lorentz}) \quad (4.21)$$

sus elementos se denominan **transformaciones de Lorentz**. Es inmediato comprobar partiendo de su definición que este conjunto forma un grupo: si Λ_1, Λ_2 son de Lorentz, $\Lambda_1 \Lambda_2$ y Λ_1^{-1} también.³⁹ El conjunto $O(3,1)$ es un grupo de matrices pseudo ortogonales 4×4 y tiene dimensión 6. $O(3,1)$ es una extensión del grupo de rotaciones $SO(3)$ (para $SO(3)$ la métrica es $(G_3)_{ij} = \delta_{ij}$, $i, j = 1, 2, 3$, de modo que la condición $G_3 = R^T G_3 R$, equivale a $R^T R = 1$).

4.3. Métricas y aplicaciones lineales

Hay que notar que en un espacio V de dimensión n , las aplicaciones lineales, A^i_j , y las métricas g_{ij} se pueden representar mediante matrices $n \times n$, pero son objetos geoméricamente distintos y se transforman de modo distinto bajo cambios de base. Sea U un cambio de base

$$e'_i = U^j_i e_j, \quad x = x'^i e'_i = x'^i U^j_i e_j = x^j e_j, \quad x^j = U^j_i x'^i, \quad x'^i = (U^{-1})^i_j x^j \quad (4.22)$$

matricialmente

$$x = U x', \quad x' = U^{-1} x \quad (4.23)$$

siendo U una matriz $n \times n$ y x, x' matrices $n \times 1$ (**matrices columna**). Los índices que se transforman como i en e_i se denominan **índices covariantes**, los que se transforman como i en x^i se denominan **índices contravariantes**. Se suelen poner abajo y arriba, respectivamente.

Si A es una aplicación lineal, $Ax = y$,

$$Ae_i = A^j_i e_j, \quad y = A(x^i e_i) = x^i A^j_i e_j, \quad y^j = A^j_i x^i, \quad y = Ax, \quad (4.24)$$

³⁹Equivalentemente, $\Lambda_1 \Lambda_2^{-1} \in L$. En general, un subconjunto H de un grupo G es subgrupo sii $HH^{-1} \subseteq H$.

al cambiar de base

$$y' = U^{-1}y = U^{-1}Ax = U^{-1}AUx' \equiv A'x', \quad A' = U^{-1}AU, \quad A'^i_j = (U^{-1})^i_k A^k_l U^l_j. \quad (4.25)$$

Si G es una métrica, $\mathbf{x} \cdot \mathbf{y} = g_{ij}x^i y^j \equiv x^T G y$, al cambiar de base

$$\begin{aligned} \mathbf{x}' \cdot \mathbf{y}' &= x'^T G' y' = \mathbf{x} \cdot \mathbf{y} = x^T G y = (Ux')^T G U y', \\ \Rightarrow G' &= U^T G U, \quad g'_{ij} = (U^T)^i_k g_{kl} U^l_j = U^k_i g_{kl} U^l_j. \end{aligned} \quad (4.26)$$

Con una métrica no singular se puede asociar un vector covariante (una 1-forma del espacio dual V^*) a cada vector de V (vectores contravariantes):

$$g^{ij} \equiv (G^{-1})_{ij}, \quad g^{ij} g_{jk} = \delta^i_k, \quad x_i \equiv g_{ij} x^j, \quad x^i = g^{ij} x_j. \quad (4.27)$$

Bajo un cambio de base x_i se transforma covariantemente

$$x'_i = g'_{ij} x'^j = U^k_j g_{kl} U^l_j (U^{-1})^j_m x^m = U^k_j g_{kl} x^l = U^k_j x_k. \quad (4.28)$$

g^{ij} y g_{ij} se pueden usar para subir y bajar índices en tensores, en particular, la propia métrica

$$g^{ij} = g^{ik} g_{kl} g^{lj}, \quad g^i_j = g^{ik} g_{kj} = \delta^i_j. \quad (4.29)$$

Nótese que δ^i_j (aplicación lineal) es un tensor invariante, en cambio δ_{ij} (métrica) no es invariante bajo cambios de base arbitrarios. Igualmente si en una base $T_{ij} = \delta_{ij}$ ello no implica que $T^i_j = \delta^i_j$.

Para su identificación como matrices, cuál es el primer índice y cuál es el segundo índice es siempre importante (exceptuando el caso de matrices simétricas o antisimétricas). Si además la métrica no es δ_{ij} , también importa si el índice es covariante o contravariante, de cara a su transformación bajo cambios de base. Así T^{ij} , T^{ji} , T^i_j , etc, son en general distintos objetos.

Un índice covariante se puede contraer con uno contravariante y los papeles se pueden intercambiar:

$$T^{ijk} S_{jk} = T^i_j{}^k S^j_k = T^i_{jk} S^{jk}. \quad (4.30)$$

Y también se pueden pasar de contravariante a covariante (y viceversa) a los dos lados de una ecuación tensorial:

$$V^i = T^i_j S^j, \quad \Leftrightarrow \quad V_i = T_{ij} S^j. \quad (4.31)$$

4.4. Estructura del grupo de Lorentz

En relatividad, en un sistema inercial, la métrica es $g_{\mu\nu} = \text{diag}(-, +, +, +) = g^{\mu\nu}$, de modo que

$$x_\mu = (-x^0, \mathbf{x}), \quad x \cdot y = g_{\mu\nu} x^\mu y^\nu = x^\mu y_\mu = x_\nu y^\nu, \quad (4.32)$$

y bajo una transformación de Lorentz

$$x_\mu \mapsto x'_\mu = g_{\mu\nu} \Lambda^\nu_\alpha x^\alpha = (g_{\mu\nu} \Lambda^\nu_\alpha g^{\alpha\beta}) x_\beta = (\Lambda^{-1})^\beta_\mu x_\beta \equiv \Lambda_\mu^\beta x_\beta, \quad G \Lambda G^{-1} = \Lambda^{T-1}. \quad (4.33)$$

Por definición de transformación de Lorentz $g_{\mu\nu}$ es invariante

$$g'_{\mu\nu} = g_{\alpha\beta} \Lambda^\alpha_\mu \Lambda^\beta_\nu = g_{\mu\nu}. \quad (4.34)$$

Los elementos del grupo de Lorentz se dividen en dos clases disconexas, L_\pm , según $\det \Lambda$ sea positivo o negativo:

$$G = \Lambda^T G \Lambda \Rightarrow \det(G) = \det(G) \det(\Lambda)^2 \Rightarrow \det \Lambda = \pm 1, \quad (4.35)$$

y también en dos clases $L^{\uparrow, \downarrow}$ según se conserve o no el sentido del tiempo (**transformaciones ortócronas o antiortócronas**):

$$g_{00} = g_{\alpha\beta} \Lambda^\alpha_0 \Lambda^\beta_0 \Rightarrow -1 = -(\Lambda^0_0)^2 + \sum_{i=1}^3 (\Lambda^i_0)^2 \Rightarrow \Lambda^0_0 \geq 1 \quad \text{ó} \quad \Lambda^0_0 \leq -1. \quad (4.36)$$

En total $O(3, 1)$ tiene cuatro componentes conexas

$$L = L_+^\uparrow \cup L_-^\uparrow \cup L_+^\downarrow \cup L_-^\downarrow. \quad (4.37)$$

L_-^\uparrow contiene las transformaciones ortócronas con inversión espacial, en particular la **transformación de paridad**

$$(x^0, \mathbf{x}) \mapsto (x^0, -\mathbf{x}), \quad P = \text{diag}(1, -1, -1, -1) \in L_-^\uparrow. \quad (4.38)$$

L_-^\downarrow contiene las transformaciones antiortócronas sin inversión espacial, en particular **inversión temporal**

$$(x^0, \mathbf{x}) \mapsto (-x^0, \mathbf{x}), \quad T = \text{diag}(-1, +1, +1, +1) \in L_-^\downarrow. \quad (4.39)$$

L_+^\uparrow es el **grupo propio ortócrono de Lorentz** y es el que vamos a considerar en lo que sigue. Este grupo es doblemente conexo (al igual que su subgrupo $SO(3)$), no compacto (las transformaciones de Lorentz puras no son matrices unitarias, ni equivalentes a unitarias en \mathbb{R}^4). Su recubridor universal es $SL(2, \mathbb{C})$.

El grupo L_+^\uparrow contiene rotaciones (parametrizadas por ϕ) y boosts (parametrizados por v) en total $3 + 3 = 6$ parámetros, así como productos de ambos.

Los **boosts** son de la forma dada en ec. (4.11) y están caracterizados por su acción sobre $(1, \mathbf{0})$,

$$B(v) \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \gamma \\ \gamma \mathbf{v}/c \end{pmatrix}, \quad \gamma = (1 - \mathbf{v}^2/c^2)^{-1/2}. \quad (4.40)$$

Las **rotaciones** (relativas a un observador) se definen como las transformaciones que dejan $(1, \mathbf{0})$ invariante, y forman un subgrupo isomorfo a $SO(3)$

$$\left\{ R \in L_+^\uparrow, R \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \right\} \cong SO(3). \quad (4.41)$$

Las rotaciones son de la forma $\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & R \end{pmatrix}$, de modo que $(x^0, \mathbf{x}) \mapsto (x^0, R\mathbf{x})$. Las rotaciones no son un subgrupo invariante, esto se debe a que el vector $(1, \mathbf{0})$ es distinto para cada observador.

Los boosts relativistas (a diferencia de los galileanos) no forman un subgrupo: el producto de dos boosts contiene una rotación, a menos que las velocidades de los boosts sean paralelas. Los boosts se pueden considerar como representantes canónicos del espacio cociente $L_+^\uparrow/SO(3)$ (no es un grupo cociente porque las rotaciones no forman un subgrupo invariante).

Todas las transformaciones de L_+^\uparrow se pueden escribir unívocamente en la forma

$$\Lambda = BR, \quad (4.42)$$

donde R es una rotación y B un boost. En efecto, aplicando Λ sobre $(1, \mathbf{0})$

$$\Lambda \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} x^0 \\ \mathbf{x} \end{pmatrix} \equiv \begin{pmatrix} \gamma \\ \gamma \mathbf{v}/c \end{pmatrix}, \quad \frac{\mathbf{v}}{c} = \frac{\mathbf{x}}{x^0}, \quad x^0 = \gamma \quad \text{por} \quad -1 = \mathbf{x}^2 - (x^0)^2, \quad x^0 > 0. \quad (4.43)$$

Esta v define unívocamente el boost B . Ahora

$$B^{-1} \Lambda \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} = B^{-1} \begin{pmatrix} \gamma \\ \gamma \mathbf{v}/c \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \quad (4.44)$$

implica que $B^{-1}\Lambda = R$ es efectivamente una rotación. También es inmediato que

$$RB(\mathbf{v})R^{-1} = B(R\mathbf{v}) \Rightarrow \Lambda = B(\mathbf{v})R = RB(R^{-1}\mathbf{v}). \quad (4.45)$$

4.5. Grupo de Poincaré

Si se consideran transformaciones que dejen invariante el intervalo entre dos sucesos x_1 y x_2

$$s^2 = (x_1 - x_2)^2, \quad (4.46)$$

se obtiene el grupo inhomogéneo de Lorentz o **grupo de Poincaré**, $P = IO(3, 1)$:

$$(\Lambda, a) \in P \quad \Lambda \in L, \quad a \in \mathbb{R}^4, \quad x \mapsto x' = \Lambda x + a, \quad x'^\mu = \Lambda^\mu_\nu x^\nu + a^\mu. \quad (4.47)$$

Incluye **traslaciones espaciales** y **temporales** además de transformaciones de Lorentz. La ley de composición se obtiene inmediatamente haciendo dos **transformaciones de Poincaré** sucesivas

$$(\Lambda_{12}, a_{12}) = (\Lambda_1 \Lambda_2, a_1 + \Lambda_1 a_2). \quad (4.48)$$

Matemáticamente, el grupo de Poincaré tiene estructura de producto semidirecto de traslaciones y Lorentz, $IO(3, 1) = T^4 \otimes_s O(3, 1)$.⁴⁰ El grupo de Poincaré tiene $6 + 4 = 10$ parámetros.

4.5.1. Tipos de intervalo

Si $s^2 < 0$, el intervalo es de **tipo tiempo**. Existe un sistema de referencia en el que $x_1 = x_2$ (los dos eventos están en el mismo sitio pero a distintos tiempos) y $c\tau = \sqrt{-s^2} = |x_1^0 - x_2^0|$, τ es el **tiempo propio**, el tiempo medido por un observador en reposo en ese sistema. Así en la paradoja de los gemelos (ver Fig 7)

$$c\tau_1 = \sqrt{(2ct)^2} = 2ct, \quad c\tau_2 = \sqrt{(ct)^2 - (vt)^2} + \sqrt{(ct)^2 - (-vt)^2} = 2ct\sqrt{1 - v^2/c^2} < 2ct. \quad (4.49)$$

Si $s^2 > 0$, el intervalo es **tipo espacio**. Existe un sistema de referencia en el que ambos sucesos son simultáneos, $t_1 = t_2$, y $d = \sqrt{s^2} = |\mathbf{x}_1 - \mathbf{x}_2|$ es la distancia entre ambos.

Si $s^2 = 0$, el intervalo es de **tipo luz**. En este caso x_1 y x_2 se pueden conectar por una señal luminosa que parta de x_1 y llegue a x_2 (si $t_1 < t_2$) o al revés.

⁴⁰Que un grupo G tenga estructura de **producto semidirecto** $N \otimes_s H$ quiere decir que N es un subgrupo invariante de G , H un subgrupo y $G = NH$ con $N \cap H = \{e\}$ (implica $G/N \cong H$).

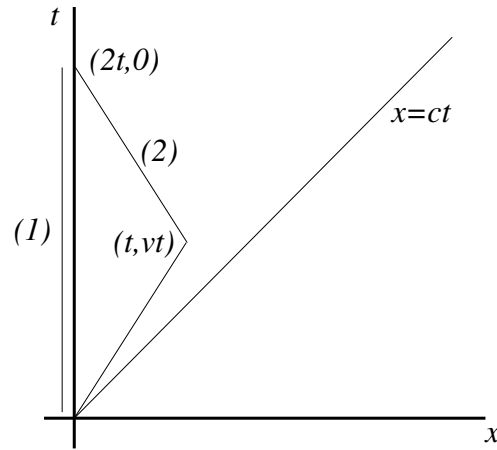


Figura 7: Paradoja de los gemelos: el tiempo propio a lo largo del camino 2 es menor que siguiendo el camino 1.

Para intervalos de tipo tiempo o luz, la ordenación temporal, es decir, el signo de $x_1^0 - x_2^0$, es un invariante bajo transformaciones ortócronas, L^\uparrow , en cambio si el intervalo es tipo espacio el signo de $x_1^0 - x_2^0$ depende del sistema de referencia.

Es interesante notar que el conjunto de transformaciones que deja invariante la condición $(x_1 - x_2)^2 = 0$ (pero no necesariamente $(x_1 - x_2)^2$ cuando este intervalo no es cero) es bastante mayor que el grupo de Poincaré, estas transformaciones forman el **grupo conforme**, de dimensión 15, e incluye dilataciones ($\lambda \neq 1$ en ec. (4.7)) así como transformaciones conformes especiales que son no lineales [1].

4.6. Álgebra de Lie del grupo de Poincaré

4.6.1. Álgebra de Lorentz

Consideremos una transformación de Lorentz infinitesimal

$$\Lambda^\mu{}_\nu = g^\mu{}_\nu - \delta\omega^\mu{}_\nu. \quad (4.50)$$

Para que sea de Lorentz debe conservar la métrica y eso impone condiciones sobre los 16 parámetros

infinitesimales $\delta\omega^\mu{}_\nu$,

$$\begin{aligned} g_{\mu\nu} &= (g^\alpha{}_\mu - \delta\omega^\alpha{}_\mu)g_{\alpha\beta}(g^\beta{}_\nu - \delta\omega^\beta{}_\nu) = g_{\mu\nu} - g_{\beta\mu}\delta\omega^\beta{}_\nu - \delta\omega^\alpha{}_\mu g_{\alpha\nu} \\ \Rightarrow \delta\omega_{\mu\nu} &= -\delta\omega_{\nu\mu}, \quad \text{ó} \quad \delta\omega^\mu{}_\nu = -\delta\omega_\nu{}^\mu. \end{aligned} \quad (4.51)$$

Como $\delta\omega_{\mu\nu}$ es antisimétrico el número de parámetros independientes es 6, que es la dimensión del grupo. Se puede proceder a identificar los 6 parámetros independientes (ϕ y v), lo cual rompe invariancia Lorentz explícita, o bien trabajar con $\omega_{\mu\nu}$ (coordenadas canónicas asociadas a $\delta\omega_{\mu\nu}$) como tensor antisimétrico para mantener la invariancia Lorentz en forma manifiesta. De momento lo hacemos así.

El convenio usual en la definición de los generadores de Lorentz es (unidades $\hbar = 1$)

$$U(\Lambda) = e^{-\frac{i}{2}\omega_{\mu\nu}J^{\mu\nu}}, \quad J^{\mu\nu} = -J^{\nu\mu}. \quad (4.52)$$

Hay sólo 6 generadores independientes.

Para obtener el **álgebra de Lorentz** se puede usar la propia representación matricial Λ que define el grupo. En este caso $J^{\mu\nu}$ son 16 matrices 4×4 (una matriz por cada elección de μ y ν).

$$\begin{aligned} \Lambda^\alpha{}_\beta &= g^\alpha{}_\beta - \delta\omega^\alpha{}_\beta = (U)^\alpha{}_\beta = (1)^\alpha{}_\beta - \frac{i}{2}\delta\omega_{\mu\nu}(J^{\mu\nu})^\alpha{}_\beta, \\ \delta\omega^\alpha{}_\beta &= g^{\alpha\mu}g_\beta{}^\nu\delta\omega_{\mu\nu} = \frac{1}{2}\delta\omega_{\mu\nu}(g^{\alpha\mu}g_\beta{}^\nu - g^{\alpha\nu}g_\beta{}^\mu) \\ \Rightarrow (J^{\mu\nu})^\alpha{}_\beta &= -i(g^{\alpha\mu}g_\beta{}^\nu - g^{\alpha\nu}g_\beta{}^\mu). \end{aligned} \quad (4.53)$$

De aquí se obtienen las **relaciones de conmutación** (que no dependen de la representación usada)

$$[J^{\mu\nu}, J^{\alpha\beta}] = -i(g^{\nu\alpha}J^{\mu\beta} - g^{\mu\alpha}J^{\nu\beta} - g^{\nu\beta}J^{\mu\alpha} + g^{\mu\beta}J^{\nu\alpha}). \quad (4.54)$$

Equivalentemente, si $\mu\nu\alpha\beta$ representan índices distintos, las relaciones son

$$\begin{aligned} [J^{\mu\nu}, J_\mu{}^\beta] &= iJ^{\nu\beta} \quad (\text{no hay suma sobre } \mu) \\ [J^{\mu\nu}, J^{\alpha\beta}] &= 0. \end{aligned} \quad \mu, \nu, \alpha, \beta \text{ distintos} \quad (4.55)$$

En realidad las relaciones de conmutación (4.54) son válidas para cualquier grupo $O(n)$ u $O(n, m)$ ya que no se han usado propiedades especiales de la métrica.

4.6.2. Álgebra de Poincaré

Para el grupo de Poincaré el convenio que define los generadores de las traslaciones espacio-temporales P^μ es

$$U(\Lambda, a) = e^{-ia_\mu P^\mu} e^{-\frac{i}{2}\omega_{\mu\nu} J^{\mu\nu}}, \quad J^{\mu\nu} = -J^{\nu\mu}. \quad (4.56)$$

Las relaciones de conmutación entre J ya las tenemos, nos falta $[P, J]$ y $[P, P]$. El método de antes no se puede usar de forma directa porque la acción de (Λ, a) en \mathbb{R}^4 no es lineal sino afín:

$$(\Lambda, a)x = \Lambda x + a. \quad (4.57)$$

Una opción es reducir esto a una acción lineal en \mathbb{R}^5

$$\begin{pmatrix} \Lambda & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} = \begin{pmatrix} \Lambda x + a \\ 1 \end{pmatrix} \quad (\text{matrices } 5 \times 5). \quad (4.58)$$

Los elementos de matriz de $(J^{\mu\nu})^\alpha_\beta$ ($\alpha, \beta = 0, 1, 2, 3$) son los mismos de antes, $(P^\mu)^\alpha_4 = ig^{\alpha\mu}$, y los demás elementos de matriz se anulan.

Alternativamente, podemos usar la representación de funciones escalares $\psi(x)$ en $\mathcal{C}(\mathbb{R}^4)$ (ec. (2.64)):

$$\psi(x) = \psi(\Lambda^{-1}(x - a)). \quad (4.59)$$

Para una transformación infinitesimal

$$\begin{aligned} \psi'(x) &= \psi(x + \delta\omega x - \delta a) = (1 - \delta a^\mu \partial_\mu + \delta\omega^\mu_{\nu} x^\nu \partial_\mu) \psi(x) \\ \delta\psi &= (-i\delta a_\mu P^\mu - \frac{i}{2}\delta\omega_{\mu\nu} J^{\mu\nu}) \psi. \end{aligned} \quad (4.60)$$

Simetrizando e identificando P^μ y $J^{\mu\nu}$ con los operadores diferenciales, se obtiene, en esta representación

$$P^\mu = -i\partial^\mu, \quad J^{\mu\nu} = -i(x^\mu \partial^\nu - x^\nu \partial^\mu) = x^\mu P^\nu - x^\nu P^\mu. \quad (4.61)$$

Usando las propiedades

$$[\partial_\mu, x^\nu] = g_\mu^\nu, \quad [x^\mu, x^\nu] = [\partial_\mu, \partial_\nu] = 0 \quad (4.62)$$

así como ec. (3.61), para $[J, J]$ se obtiene el resultado ya conocido (álgebra de Lorentz) y

$$[J^{\mu\nu}, P^\alpha] = -i(g^{\nu\alpha}P^\mu - g^{\mu\alpha}P^\nu), \quad [P^\mu, P^\nu] = 0. \quad (4.63)$$

Observaciones:

1) El álgebra de no depende de la representación usada.

2) La representación en $\mathcal{C}(\mathbb{R}^4)$ (o $L^2(\mathbb{R}^4)$) se ha usado aquí únicamente para obtener una representación fiel del grupo de Poincaré. Esta representación *no* es la que aparecería en mecánica cuántica de una partícula relativista sin espín, por ejemplo. Ahí el espacio de Hilbert sería $L^2(\mathbb{R}^3)$. En mecánica cuántica $\psi(\mathbf{x}, t)$ puede estar localizada en una zona del espacio \mathbb{R}^3 pero no localizada en un intervalo temporal ya que $\int d^3x |\psi(\mathbf{x}, t)|^2 = 1$ todo el tiempo, en cambio una función de $L^2(\mathbb{R}^4)$ estaría localizada también temporalmente, por $\int d^4x |\psi(x)|^2 < +\infty$. O también, se ve que en la representación obtenida $[P^0, \mathbf{x}] = 0$, lo cual no es cierto en el caso cuántico (\mathbf{x} no es una constante de movimiento).

3) El operador x^μ existe en la representación en $L^2(\mathbb{R}^4)$ pero no pertenece al álgebra del grupo, y por tanto no tiene que estar definido en otras representaciones del grupo. Por ejemplo, en una representación matricial (como la definida en ec. (4.58), de dimensión 5) nunca pueden hallarse operadores que satisfagan $[x^\mu, P^\nu] = ig^{\mu\nu}$, ya que, por ejemplo para $\mu = \nu = 0$, la traza de la matriz de la derecha no es cero y la de la izquierda sí, por ser un conmutador de matrices.

Por inspección del álgebra de Poincaré se sigue que $J^{\mu\nu}$ forma una subálgebra y genera un subgrupo (el grupo de Lorentz) y P^μ forma un ideal y en consecuencia genera un subgrupo invariante abeliano, el de las traslaciones. Que sea invariante quiere decir que una traslación sigue siendo una traslación para cualquier otro observador transformado Poincaré (en cambio un transformación de Lorentz se verá como Lorentz más traslación para otros observadores).

Por las relaciones de conmutación, bajo una transformación de Lorentz infinitesimal

$$\begin{aligned} P^\mu &\mapsto U(\Lambda)P^\mu U^{-1}(\Lambda) = P^\mu - \frac{i}{2}[\delta\omega_{\alpha\beta}J^{\alpha\beta}, P^\mu] \\ \delta P^\mu &= -i\delta\omega_{\alpha\beta}(-i)g^{\beta\mu}P^\alpha = -\delta\omega_{\alpha}^{\mu}P^\alpha = \delta\omega^{\mu}_{\alpha}P^\alpha. \end{aligned} \quad (4.64)$$

Esta transformación corresponde a un **cuadrivector Lorentz** (a nivel infinitesimal), es decir,

$$P^\mu \mapsto P^\mu + \delta\omega^{\mu}_{\nu}P^\nu = (g^{\mu}_{\nu} + \delta\omega^{\mu}_{\nu})P^\nu = (\Lambda^{-1})^{\mu}_{\nu}P^\nu. \quad (4.65)$$

Aplicando la transformación infinitesimal repetidamente se obtiene el resultado finito

$$P^\mu \mapsto U(\Lambda)P^\mu U^{-1}(\Lambda) = (\Lambda^{-1})^\mu{}_\nu P^\nu. \quad (4.66)$$

Igualmente, se deduce

$$J^{\mu\nu} \mapsto U(\Lambda)J^{\mu\nu}U^{-1}(\Lambda) = (\Lambda^{-1})^\mu{}_\alpha (\Lambda^{-1})^\nu{}_\beta J^{\alpha\beta}, \quad (4.67)$$

que es la ley de transformación de un tensor Lorentz (dos veces contravariante).

Como ya se observó para rotaciones, los operadores se transforman al revés que las coordenadas, $P^\mu \mapsto (\Lambda^{-1})^\mu{}_\nu P^\nu$ frente $x^\mu \mapsto \Lambda^\mu{}_\nu x^\nu$. Ambas transformaciones son consistentes (y no lo serían al revés):

$$\begin{aligned} x &\xrightarrow{2} x' = \Lambda_2 x \xrightarrow{1} x'' = \Lambda_1 x' = \Lambda_1 \Lambda_2 x = \Lambda_{12} x, \\ P &\xrightarrow{2} U_2 P U_2^{-1} = \Lambda_2^{-1} P \xrightarrow{1} U_1 (\Lambda_2^{-1} P) U_1^{-1} = \Lambda_2^{-1} (U_1 P U_1^{-1}) = \Lambda_2^{-1} \Lambda_1^{-1} P = (\Lambda_{12})^{-1} P. \end{aligned} \quad (4.68)$$

4.7. Álgebra de Poincaré en la base espacio-temporal

El álgebra de Poincaré también puede escribirse en la base adaptada a las coordenadas ϕ (rotaciones), \mathbf{v} (boosts), τ (traslaciones temporales) y \mathbf{a} (traslaciones espaciales), con generadores asociados \mathbf{J} (momento angular), \mathbf{K} (generador de los boosts), H (hamiltoniano) y \mathbf{P} (momento lineal). Para una transformación de Poincaré infinitesimal,

$$\begin{aligned} U &= e^{-i\delta X}, & \delta X &= \delta a_\mu P^\mu + \frac{1}{2} \delta \omega_{\mu\nu} J^{\mu\nu} \\ & & &= -\delta\tau H + \delta\mathbf{a} \cdot \mathbf{P} + \delta\mathbf{v} \cdot \mathbf{K} + \delta\phi \cdot \mathbf{J}. \end{aligned} \quad (4.69)$$

Veamos primero la relación entre los dos conjuntos de coordenadas. Usamos

$$\delta\mathbf{r} = \delta\mathbf{a} + \delta\mathbf{v}t + \delta\phi \times \mathbf{r}, \quad \delta t = \delta\tau + \frac{1}{c^2} \delta\mathbf{v} \cdot \mathbf{r}, \quad (4.70)$$

a comparar con

$$\delta x^\mu = \delta a^\mu - \delta \omega^\mu{}_\nu x^\nu, \quad (4.71)$$

teniendo en cuenta que $x^\mu = (ct, \mathbf{r})$. Particularizando para $\mu = i$ se obtiene

$$\delta x^i = \delta a^i - \delta \omega^i_{0c} ct - \delta \omega^i_{jx^j}, \quad (4.72)$$

que implica

$$\begin{aligned} (\delta \mathbf{a})_i &= \delta a^i, & (\delta \mathbf{v})_i &= -\delta \omega^i_{0c} c = c \delta \omega_{0i}, \\ \varepsilon_{ikj} \delta \phi_k x^j &= -\delta \omega^i_{jx^j} \Rightarrow \delta \omega_{ij} = \varepsilon_{ijk} \delta \phi_k, & \delta \phi_k &= \frac{1}{2} \varepsilon_{ijk} \delta \omega_{ij}. \end{aligned} \quad (4.73)$$

Si se toma $\mu = 0$ se obtiene

$$c \delta t = \delta a^0 - \delta \omega^0_{ix^i} \Rightarrow \delta a^0 = c \delta \tau. \quad (4.74)$$

Para los generadores

$$\delta X = \delta a_\mu P^\mu + \frac{1}{2} \delta \omega_{\mu\nu} J^{\mu\nu} = -c \delta \tau P^0 + \delta a^i P^i + \frac{1}{c} \delta v^i J^{0i} + \frac{1}{2} \varepsilon_{ijk} \delta \phi_k J^{jk}, \quad (4.75)$$

que implica

$$P^0 = \frac{1}{c} H, \quad P^i = (\mathbf{P})_i, \quad J^{0i} = c K^i, \quad J^k = \frac{1}{2} \varepsilon_{ijk} J^{ij}, \quad J^{ij} = \varepsilon_{ijk} J^k. \quad (4.76)$$

Cambiando de variables podemos reexpresar el álgebra de Poincaré, ecs. (4.54) y (4.63), en la base $H, \mathbf{P}, \mathbf{K}$ y \mathbf{J} ,

$$\begin{aligned} [J^i, J^j] &= i \varepsilon_{ijk} J^k, & [J^i, K^j] &= i \varepsilon_{ijk} K^k, & [K^i, K^j] &= -\frac{i}{c^2} \varepsilon_{ijk} J^k, \\ [P^i, P^j] &= [H, P^i] = 0, \\ [H, J^i] &= 0, & [H, K^i] &= i P^i, & [J^i, P^j] &= i \varepsilon_{ijk} P^k, & [P^i, K^j] &= i \delta_{ij} \frac{H}{c^2}. \end{aligned} \quad (4.77)$$

La primera línea es el álgebra de Lorentz y contiene a las rotaciones como subgrupo. En el límite $c \rightarrow \infty$ (y suponiendo que los generadores son finitos en ese límite) se recupera el álgebra del grupo de Galileo, en el que los boosts conmutan y forman un subgrupo abeliano.

Por la relación de conmutación con \mathbf{J} , se deduce que \mathbf{P}, \mathbf{K} y \mathbf{J} son operadores vectoriales, $\mathbf{P} \mapsto R^{-1} \mathbf{P}$, etc. A su vez la relación $[J, P] \sim P$ dice que \mathbf{J} tiene una componente extrínseca, que cambia bajo traslaciones. En efecto, para una traslación infinitesimal

$$U(\delta \mathbf{a}) \mathbf{J} U(\delta \mathbf{a})^{-1} = \mathbf{J} - i[\delta \mathbf{a} \cdot \mathbf{P}, \mathbf{J}] = \mathbf{J} - \delta \mathbf{a} \times \mathbf{P}, \quad (4.78)$$

y para una transformación finita

$$U(\mathbf{a})\mathbf{J}U(\mathbf{a})^{-1} = \mathbf{J} - \mathbf{a} \times \mathbf{P}. \quad (4.79)$$

Esto es consistente con $\mathbf{J} = \mathbf{L} + \mathbf{S}$ donde \mathbf{S} (el espín) es intrínseco (invariante bajo traslaciones) y $\mathbf{L} = \mathbf{x} \times \mathbf{P}$ es la parte orbital, teniendo en cuenta que $\mathbf{x} \mapsto \mathbf{x} - \mathbf{a}$ bajo traslaciones (consecuencia de $[x^i, P^j] = i\delta_{ij}$).⁴¹ Nótese otra vez que \mathbf{x} no forma parte del álgebra de Poincaré.

Las relaciones de conmutación indican que H , \mathbf{P} y \mathbf{J} con constantes de movimiento (conmutan con H) y al mismo tiempo que H es invariante bajo traslaciones y rotaciones.

4.8. Representaciones irreducibles del grupo de Lorentz

Veamos la irreps de dimensión finita del grupo de Lorentz. Estas irreps no son unitarias: como el grupo es simple y no compacto sus representaciones unitarias son de dimensión infinita (excepto la trivial).

Definimos los nuevos operadores

$$\mathbf{J}_L := \frac{1}{2}(\mathbf{J} + ic\mathbf{K}), \quad \mathbf{J}_R := \frac{1}{2}(\mathbf{J} - ic\mathbf{K}), \quad \mathbf{J} = \mathbf{J}_L + \mathbf{J}_R, \quad \mathbf{K} = -\frac{i}{c}(\mathbf{J}_L - \mathbf{J}_R). \quad (4.80)$$

Es inmediato comprobar que el álgebra de Lorentz se puede reescribir en la forma

$$[J_L^i, J_L^j] = i\varepsilon_{ijk}J_L^k, \quad [J_R^i, J_R^j] = i\varepsilon_{ijk}J_R^k, \quad [J_L^i, J_R^j] = 0, \quad (4.81)$$

y $\mathbf{J}_{L,R}$ forman dos álgebras de $SU(2)$ independientes. Una consecuencia inmediata es que \mathbf{J}_L^2 y \mathbf{J}_R^2 son operadores de Casimir del álgebra de Lorentz. Están relacionados con los invariantes $J^{\mu\nu}J_{\mu\nu}$ y $\varepsilon_{\mu\nu\alpha\beta}J^{\mu\nu}J^{\alpha\beta}$, expresados en la base $J^{\mu\nu}$.

Las matrices que representan a \mathbf{J} y \mathbf{K} son irreducibles si y sólo si $\mathbf{J}_{L,R}$ lo son, entonces el problema se reduce a encontrar irreps de $su(2)$ que como sabemos son necesariamente equivalentes a unitarias

$$\mathbf{J}_{L,R} = \mathbf{J}_{L,R}^\dagger, \quad \mathbf{J} = \mathbf{J}^\dagger, \quad \mathbf{K} = -\mathbf{K}^\dagger. \quad (4.82)$$

Estas irreps son conocidas y se caracterizan por etiquetas $j_{L,R} = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots$. Si usamos $\hat{\mathbf{J}}_L$ para denotar las matrices de dimensión $2j_L + 1$, de la irrep j_L de $SU(2)$, y lo mismo para $\hat{\mathbf{J}}_R$, lo que se

⁴¹Alternativamente, $U(\mathbf{a})\mathbf{x}U(\mathbf{a})^{-1}|\mathbf{r}\rangle = U(\mathbf{a})\mathbf{x}|\mathbf{r} - \mathbf{a}\rangle = (\mathbf{r} - \mathbf{a})U(\mathbf{a})|\mathbf{r} - \mathbf{a}\rangle = (\mathbf{r} - \mathbf{a})|\mathbf{r}\rangle = (\mathbf{x} - \mathbf{a})|\mathbf{r}\rangle$.

tiene para el álgebra de Lorentz es una irrep de dimensión $(2j_L + 1)(2j_R + 1)$ que actúa en $V^{j_L} \otimes V^{j_R}$,

$$\mathbf{J}_L = \hat{\mathbf{J}}_L \otimes 1, \quad \mathbf{J}_R = 1 \otimes \hat{\mathbf{J}}_R, \quad \mathbf{J} = \hat{\mathbf{J}}_L \otimes 1 + 1 \otimes \hat{\mathbf{J}}_R, \quad \mathbf{K} = -\frac{i}{c}(\hat{\mathbf{J}}_L \otimes 1 - 1 \otimes \hat{\mathbf{J}}_R). \quad (4.83)$$

Nótese que el factor 1 es la identidad en $(2j_R + 1)$ y en $(2j_L + 1)$ dimensiones, respectivamente.

Cada irrep de Lorentz está caracterizada por los valores de j_L y j_R y la denotamos $[j_L, j_R]$. En particular,

$[0, 0]$, dimensión 1, son los escalares Lorentz (la representación trivial).

$[\frac{1}{2}, \frac{1}{2}]$, dimensión 4, son los cuadvectores, $A^\mu = (A^0, \mathbf{A})$, $A^\mu \mapsto (\Lambda^{-1})^\mu_\nu A^\nu$. A^0 es un escalar bajo rotaciones ($j = 0$) y \mathbf{A} un vector ($j = 1$). Como $\mathbf{J} = \mathbf{J}_L + \mathbf{J}_R$ al acoplar $j_L = 1/2$ con $j_R = 1/2$ se obtiene $j = 0, 1$.

Las irreps básicas son $[\frac{1}{2}, 0]$ y $[0, \frac{1}{2}]$:

$$\begin{aligned} [\frac{1}{2}, 0], \quad \mathbf{J}_L = \frac{\boldsymbol{\sigma}}{2}, \quad \mathbf{J}_R = 0, \quad \mathbf{J} = \frac{\boldsymbol{\sigma}}{2}, \quad \mathbf{K} = -\frac{i}{c} \frac{\boldsymbol{\sigma}}{2} \\ [0, \frac{1}{2}], \quad \mathbf{J}_L = 0, \quad \mathbf{J}_R = \frac{\boldsymbol{\sigma}}{2}, \quad \mathbf{J} = \frac{\boldsymbol{\sigma}}{2}, \quad \mathbf{K} = +\frac{i}{c} \frac{\boldsymbol{\sigma}}{2}. \end{aligned} \quad (4.84)$$

Puesto que en estas representaciones $\mathbf{K} = \pm i\mathbf{J}/c$, la relación $\mathbf{K} \times \mathbf{K} = -i\mathbf{J}/c^2$ es consecuencia inmediata de $\mathbf{J} \times \mathbf{J} = i\mathbf{J}$.

Las representaciones básicas tienen dimensión 2 y son conjugadas una de otra. Exponenciando por ejemplo $[\frac{1}{2}, 0]$ se obtiene

$$D(\Lambda) = e^{-i\boldsymbol{\phi} \cdot \mathbf{J} - i\boldsymbol{\xi} \cdot \mathbf{K}} = e^{-\frac{i}{2}(\boldsymbol{\phi} - \frac{i}{c}\boldsymbol{\xi}) \cdot \boldsymbol{\sigma}} \equiv e^{-i\boldsymbol{\alpha} \cdot \boldsymbol{\sigma}/2}, \quad \boldsymbol{\phi}, \boldsymbol{\xi} \in \mathbb{R}^3, \quad \boldsymbol{\alpha} \in \mathbb{C}^3, \quad (4.85)$$

que es el conjunto de matrices complejas 2×2 con determinante unidad, el grupo $\text{SL}(2, \mathbb{C})$. Este grupo es el recubridor universal de L_+^\uparrow . El grupo de Lorentz es doblemente conexo por $\text{SL}(2, \mathbb{C})/\mathbb{Z}_2 \cong \text{L}_+^\uparrow$, y tiene representaciones bivaluadas igual que $\text{SO}(3)$.

4.9. Representaciones irreducibles del grupo de Poincaré

Nos referimos al grupo conexo P_+^\uparrow . Las irreps de Poincaré se clasifican por operadores invariantes. Hay básicamente dos operadores invariantes.

El primero es la **masa invariante** (al cuadrado) del sistema

$$M^2 := -\frac{1}{c^2}P^\mu P_\mu = \frac{1}{c^4}H^2 - \frac{1}{c^2}\mathbf{P}^2, \quad H^2 = (Mc^2)^2 + (c\mathbf{P})^2. \quad (4.86)$$

Puesto que P^μ es un cuadrivector, M^2 es un escalar Lorentz, y conmuta con $J^{\mu\nu}$, y también conmuta con P^μ . Por el lema de Schur, en una irrep del grupo M^2 toma un valor constante. Como los P^μ conmutan, los estados de la base de la irrep se pueden elegir propios de \mathbf{P} , $|\alpha, \mathbf{p}\rangle$ (α son otros posibles números cuánticos) y P^0 se obtiene con M^2 (salvo signo).

Hay cuatro tipos de irreps.

La representación trivial. Tiene dimensión 1 y el estado correspondiente suele denotarse $|0\rangle$, denominado **estado vacío**.⁴² En esta representación

$$U(\Lambda, a)|0\rangle = |0\rangle, \quad P^\mu|0\rangle = J^{\mu\nu}|0\rangle = 0. \quad (4.87)$$

El estado vacío no tiene momento, energía ni momento angular y representa el vacío físico en las teorías cuánticas relativistas, donde se postula que éste es el estado fundamental (es decir, el de menor energía) y que está no degenerado. Esto implica que para todos los demás estados $H > 0$. El grupo de Poincaré contiene irreps con $H < 0$ pero no aparecen en teorías admisibles y no las consideramos.

Representaciones masivas. Para éstas $M^2 > 0$ y $H > 0$. Representan partículas con masa, o sistemas de partículas. P^μ es de tipo tiempo, y haciendo un boost de velocidad

$$\mathbf{v} = -c^2 \frac{\mathbf{P}}{H}, \quad (4.88)$$

se puede llevar el sistema al reposo (**sistema del centro de masas**)

$$\mathbf{P} = 0, \quad H = Mc^2, \quad M := +\sqrt{M^2} \quad (4.89)$$

Mc^2 es la energía total del sistema en el sistema del centro de masas. En otro sistema

$$H = +\sqrt{M^2c^4 + c^2\mathbf{P}^2} = \gamma Mc^2, \quad \gamma = (1 - \mathbf{v}^2/c^2)^{-1/2}. \quad (4.90)$$

M es la **masa invariante** del sistema.

⁴²Nótese que $|0\rangle$ no es el vector 0 del espacio de Hilbert. De hecho es un estado normalizado a uno.

Representaciones sin masa. Para éstas $M^2 = 0$ y $H > 0$. Representan partículas sin masa, P^μ es de tipo luz y no admite un sistema centro de masas en el que $\mathbf{P} = 0$, más bien $H = c|\mathbf{P}| > 0$. Aunque la partícula va siempre a la velocidad c , sí cambia su energía y momento al aplicar un boost (P^μ se transforma como un cuadrivector Lorentz en todos los casos).

Representaciones taquiónicas. En estas representaciones $M^2 < 0$ y P^μ es de tipo espacio. En este caso el signo de P^0 depende del sistema de referencia (se puede cambiar mediante transformaciones de Lorentz). Las partículas en estas representaciones serían taquiones, moviéndose a velocidad superior a c . Esto lleva a paradojas (Fig. 8) y de hecho no se ha encontrado aplicación de estas irreps en la naturaleza. El hecho de que H se pueda hacer arbitrariamente negativo indica que no hay estado fundamental (si en una teoría aparecen taquiones quiere decir que el supuesto vacío es en realidad un estado metaestable, un falso vacío).

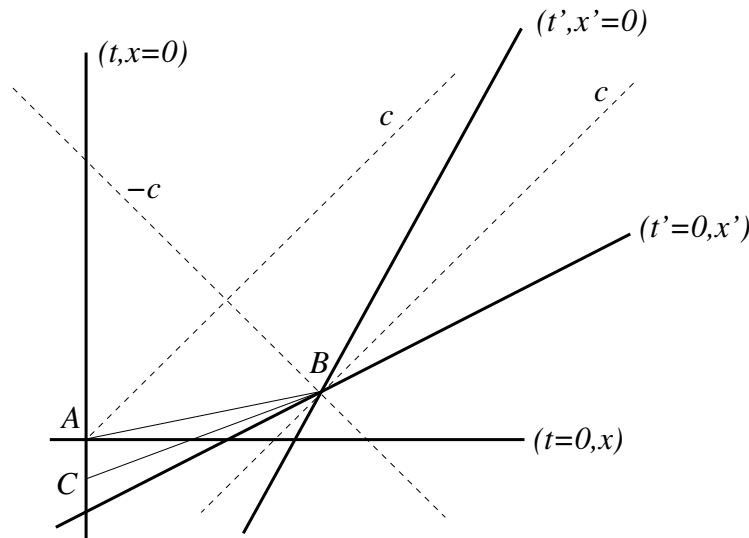


Figura 8: La señal supralumínica $A \rightarrow B$ viaja hacia adelante en el tiempo para A , la señal supralumínica $B \rightarrow C$ también viaja hacia adelante en el tiempo para B , sin embargo C está en el pasado causal de A .

Hay que observar que M incluye toda la energía en reposo del sistema físico, incluida la debida a interacciones. Así por ejemplo, un termo con agua caliente tiene más masa inercial y gravitatoria (ambas coinciden por el principio de equivalencia) que con agua fría. Si la ec. (4.90) se desarrolla en serie en potencias de $1/c$

$$H = Mc^2 + \frac{\mathbf{P}^2}{2M} + O\left(\frac{1}{c^2}\right), \quad (4.91)$$

parecería que sólo hay energía cinética, sin embargo, separando

$$Mc^2 = M_0c^2 + V \quad (4.92)$$

donde V indica un potencial de interacción, por ejemplo,

$$H = M_0c^2 + V + \frac{\mathbf{P}^2}{2(M_0 + \frac{1}{c^2}V)} + O(\frac{1}{c^2}) = M_0c^2 + V + \frac{\mathbf{P}^2}{2M_0} + O(\frac{1}{c^2}), \quad (4.93)$$

que es consistente con la fórmula no relativista con un término de interacción. M contiene toda la energía. En el límite no relativista, se incluyen en V aquellas energías (gravitatorias terrestres, químicas, etc) cuya variación y transformación en energía cinética no supone cambios de velocidades de las partículas comparables a c y en M_0 las que sí (nucleares, etc). M_0 se puede considerar inerte si y sólo si el tratamiento no relativista es adecuado.

El segundo operador invariante Poincaré que permite clasificar las irreps del grupo está relacionado con el espín. Hacemos un tratamiento cualitativo. Un tratamiento sistemático se basa en el operador de Pauli-Lubanski, $W_\mu = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}P^\nu J^{\alpha\beta}$.⁴³ Como se puede comprobar W^2 es un invariante Poincaré [5].

Para *partículas con masa* (representaciones $M^2 > 0$), el sistema físico se puede llevar al reposo, $\mathbf{P} = 0$, $P^0 = Mc$. El subgrupo del grupo de Lorentz que deja invariante $P^\mu = (Mc, \mathbf{0})$ es el grupo de rotaciones (en el sistema centro de masas) y se puede usar para reducir el espacio de Hilbert según el valor de \mathbf{J}^2 de los estados. Una vez que hemos fijado el sistema de referencia (el centro de masas) el valor de \mathbf{J}^2 es un invariante Lorentz. Además, por $\mathbf{P} = 0$, el momento angular no tiene parte orbital ($\mathbf{L} = \mathbf{x} \times \mathbf{P} = 0$) sólo hay espín, $\mathbf{J} = \mathbf{L} + \mathbf{S} = \mathbf{S}$, y por tanto \mathbf{J} es invariante bajo traslaciones espaciales, y también temporales por ser \mathbf{J} conservado. En definitiva el observable \mathbf{S}^2 así definido es un invariante Poincaré. Su espectro es el usual de un momento angular, $\mathbf{S}^2 = s(s+1)$, $s = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots$ y s se denomina el **espín** de la partícula.

Las representaciones masivas son del tipo $|M, s; \mathbf{p}, \lambda\rangle$ donde M (la masa) y s (el espín) son fijos y caracterizan la irrep. \mathbf{p} es el momento (el valor propio de \mathbf{P}) y λ es la **helicidad**, que se define como el momento angular en la dirección de \mathbf{P} , $\lambda = -s, -s+1, \dots, s$.

Para *partículas sin masa* ($M = 0$), el análisis es más complicado. El subgrupo que deja invariante $P^\mu = (|\mathbf{P}|, \mathbf{P})$ es el grupo euclídeo bidimensional, también de dimensión 3 [5]. Las únicas irreps a las que se ha encontrado aplicación física son aquellas con $W^2 = 0$ (esencialmente corresponde a

⁴³Hay dos convenios para el tensor de Levi-Civita, aquí adoptamos el convenio $\epsilon^{0ijk} = \epsilon_{ijk}$.

estados de espín finito). Para estas irreps $W^\mu = \lambda P^\mu$. Lo que se encuentra es que la helicidad

$$\lambda := \frac{\mathbf{P} \cdot \mathbf{J}}{|\mathbf{P}|}, \quad (4.94)$$

es un invariante Lorentz y Poincaré. Para una partícula masiva no lo es: la partícula se puede llevar al reposo, rotarla, y luego deshacer el boost, cambiando λ . Pero las partículas sin masa no se pueden llevar al reposo y λ no se puede cambiar. Por tanto en este caso $M = 0$ y λ son los operadores invariantes, y los valores permitidos para λ son $0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \dots$

Las representaciones sin masa son del tipo $|M = 0, \lambda; \mathbf{p}\rangle$ donde $M = 0$ (la masa) y λ (la helicidad) son fijos y caracterizan la irrep. Nótese que una partícula sin masa sólo tiene un estado espín, en vez de los $2s + 1$ posibles estados de espín de una partícula masiva.⁴⁴

⁴⁴El fotón admite dos valores, $\lambda = \pm 1$, pero cada uno define una representación irreducible de P_\pm^\dagger . Los dos estados se mezclan bajo paridad, que es realizable por ser una simetría de la interacción electromagnética. No hay fotones con helicidad nula.

5. Representaciones de $SU(n)$

5.1. Representación tensorial de $GL(n, \mathbb{C})$

Como se vio, al acoplar dos espines $\frac{1}{2}$ se podía obtener un estado triplete ($j = 1$) o singlete ($j = 0$)

$$\begin{aligned} |1, 1\rangle &= |\uparrow\uparrow\rangle, & |1, 0\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle), & |1, -1\rangle &= |\downarrow\downarrow\rangle \\ |0, 0\rangle &= \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle). \end{aligned} \quad (5.1)$$

Estos estados se pueden escribir en la forma

$$|\psi\rangle = \psi^{i_1 i_2} |i_1, i_2\rangle \quad i_k = 1, 2 \quad (\text{ó } \uparrow, \downarrow). \quad (5.2)$$

Los estados $j = 1$ son simétricos, $\psi^{i_1 i_2} = +\psi^{i_2 i_1}$ y el estado $j = 0$ es antisimétrico, $\psi^{i_1 i_2} = -\psi^{i_2 i_1}$.

Más generalmente, si $\{|i\rangle, i = 1, \dots, n\}$ es una base de $V_n \cong \mathbb{C}^n$,

$$|i_1, i_2, \dots, i_r\rangle = |i_1\rangle \otimes \dots \otimes |i_r\rangle \quad \text{base de} \quad V_n^r \equiv V_n \otimes \dots \otimes V_n, \quad (5.3)$$

un vector cualquiera es de la forma (suma implícita)

$$|\psi\rangle = \psi^{i_1 i_2 \dots i_r} |i_1, i_2, \dots, i_r\rangle. \quad (5.4)$$

$|\psi\rangle \in V_n^r$, o equivalentemente $\psi^{i_1 i_2 \dots i_r}$, es un **tensor contravariante** de rango r . En general, un tensor es un objeto caracterizado por su transformación bajo un grupo que actúe en V_n . En nuestro caso el grupo es $G_n \equiv GL(n, \mathbb{C})$ (el grupo de cambios de base en \mathbb{C}^n). En el punto de vista activo, $g \in G_n$ actúa sobre los vectores de la base de V_n así

$$|i\rangle \mapsto g^j_i |j\rangle, \quad (5.5)$$

y sobre las componentes del tensor actúa según

$$|\psi\rangle \xrightarrow{g} |\psi\rangle' = \psi^{i_1 \dots i_r} g^{j_1}_{i_1} \dots g^{j_r}_{i_r} |j_1, \dots, j_r\rangle \equiv \psi'^{j_1 \dots j_r} |j_1, \dots, j_r\rangle, \quad (5.6)$$

de donde se lee la ley de transformación de las componentes del tensor

$$\psi'^{j_1 \dots j_r} = g^{i_1}_{j_1} \dots g^{i_r}_{j_r} \psi^{i_1 \dots i_r}. \quad (5.7)$$

Esto define la representación tensorial $g \otimes \cdots \otimes g$ de $\text{GL}(n, \mathbb{C})$ sobre V_n^r . Como luego van a aparecer representaciones más generales, denotamos este tipo de representaciones por T^r . Todas las irreps inequivalentes de $\text{SU}(n)$ aparecen al reducir los espacios V_n^r , $r = 0, 1, 2, \dots$

5.2. Reducción de V_n^r bajo S_r y $\text{GL}(n, \mathbb{C})$

Para reducir V_n^r buscamos subespacios invariantes bajo $\text{GL}(n, \mathbb{C})$. El grupo de permutaciones S_r también actúa en el espacio de tensores V_n^r ,

$$p \in S_r, \quad |i_1, \dots, i_r\rangle = |i_1\rangle_1 \cdots |i_r\rangle_r \xrightarrow{p} |i_1\rangle_{p1} \cdots |i_r\rangle_{pr} = |i_{p^{-1}1}\rangle_1 \cdots |i_{p^{-1}r}\rangle_r = |i_{p^{-1}1}, \dots, i_{p^{-1}r}\rangle, \quad (5.8)$$

y para las componentes

$$\psi^{i_1, \dots, i_r} \xrightarrow{p} \psi^{i_{p1}, \dots, i_{pr}}. \quad (5.9)$$

Así por ejemplo

$$\begin{aligned} |ijk\rangle &\xrightarrow{(12)} |jik\rangle \xrightarrow{(23)} |jki\rangle, & (23)(12) &= (132) \\ \psi^{ijk}|ijk\rangle &\xrightarrow{(12)} \psi^{ijk}|jik\rangle = \psi^{jik}|ijk\rangle \xrightarrow{(23)} \psi^{jik}|ikj\rangle = \psi^{kij}|ijk\rangle. \end{aligned} \quad (5.10)$$

Las acciones de los grupos S_r y $\text{GL}(n, \mathbb{C})$ en V_n conmutan, en consecuencia, aplicando proyectores P_λ sobre irreps de S_r (λ es un diagrama de Young con r casillas)

$$V_n^r = \bigoplus_{\lambda} (V_n^r)_{\lambda} \quad (5.11)$$

donde $(V_n^r)_{\lambda}$ son espacios invariantes bajo los dos grupos.

Generalmente $(V_n^r)_{\lambda}$ es todavía reducible: cada irrep λ de S_r aparece con una multiplicidad d_λ . Tomando una base estándar en cada uno de los d_λ espacios irreducibles (mínimos) se tiene una base de $(V_n^r)_{\lambda}$,⁴⁵

$$\{|\lambda \alpha a\rangle, a = 1, \dots, n_\lambda, \alpha = 1, \dots, d_\lambda\}, \quad (5.12)$$

donde n_λ denota la dimensión de la irrep λ de S_r . Por construcción $\{|\lambda \alpha a\rangle, a = 1, \dots, n_\lambda\}$, para λ, α fijos es una base estándar de un espacio irreducible de S_r de tipo λ . Se demuestra que a su

⁴⁵Nótese que la reducción de $(V_n^r)_{\lambda}$ en espacios irreducibles de S_r no es única si hay más de uno, $d_\lambda > 1$.

vez $\{|\lambda \alpha a\rangle, \alpha = 1, \dots, d_\lambda\}$ con λ, a fijos, subtiende un espacio irreducible respecto de $GL(n, \mathbb{C})$, y además dos irreps de este tipo son equivalentes si y sólo si sus diagramas de Young, λ , son iguales.

$$\begin{aligned}
 V_n^r &= \bigoplus_{\lambda} \bigoplus_{\alpha=1}^{d_\lambda} V_{\lambda, \alpha} \equiv \bigoplus_{\lambda} d_\lambda V_\lambda \quad (\text{reducción bajo } S_r) \\
 &= \bigoplus_{\lambda} \bigoplus_{a=1}^{n_\lambda} V'_{\lambda, a} \equiv \bigoplus_{\lambda} n_\lambda V'_\lambda \quad (\text{reducción bajo } G_n).
 \end{aligned} \tag{5.13}$$

En particular para $r = 1$ el único diagrama es $[1]$ (\square) y se tiene que la propia representación que define el grupo, g^i_j , es irreducible (evidente, ya que con transformaciones invertibles arbitrarias se puede llevar cualquier vector de V_n a cualquier otro, no hay espacios invariantes propios).

En el siguiente caso más simple $r = 2$

$$\begin{aligned}
 r = 2, \quad n = 1, 2, \dots, \quad V_n^2 &= V_n \otimes V_n \\
 \psi^{ij} \quad i, j = 1, \dots, n \quad \lambda &= \begin{matrix} [2] & , & [1^2] \\ \square \square & & \square \\ & & \square \end{matrix}
 \end{aligned} \tag{5.14}$$

$$V_n^2 = (V_n^2)_{\square\square} \oplus (V_n^2)_{\square} = \text{lin} \{ \psi, \psi^{ij} = +\psi^{ji} \} \oplus \text{lin} \{ \psi, \psi^{ij} = -\psi^{ji} \}$$

$\text{dim} = n(n+1)/2 \qquad \text{dim} = n(n-1)/2$

$$\psi^{ij} = \psi_S^{ij} + \psi_A^{ij}, \quad \psi_{S,A}^{ij} = \frac{1}{2}(\psi^{ij} \pm \psi^{ji}).$$

$$\begin{aligned}
 V_n^2 &= \frac{n(n+1)}{2} \times V_{\square\square} \oplus \frac{n(n-1)}{2} \times V_{\square} \quad (\text{Reducción bajo } S_2) \\
 &= 1 \times V'_{\square\square} \oplus 1 \times V'_{\square} \quad (\text{Reducción bajo } GL(n, \mathbb{C}))
 \end{aligned} \tag{5.16}$$

Como se verá las irreps tensoriales T^r de $GL(n, \mathbb{C})$ son también irreducibles bajo $SU(n)$. Así aplicando el resultado anterior a $SU(2)$

$$V_{j_1=1/2} \otimes V_{j_2=1/2} = V_{j=0} \oplus V_{j=1}, \quad V_{j=1} = V'_{\square\square} \quad (\text{dim} = 3), \quad V_{j=0} = V'_{\square} \quad (\text{dim} = 1) \tag{5.17}$$

Tensores de rango 3. Hay n^3 estados y vamos a separarlos por tipos.

Tipo $|ijk\rangle$. Si aplicamos S_3 sobre $|ijk\rangle$ para ijk distintos, se obtienen $3!$ estados que forman la representación regular de S_3 . Esta representación se puede reducir aplicando idempotentes asociados a los tableros estándar $e_{123}, e_{12}, e_{13}, e_1$. Puesto que los índices se van a permutar, se puede elegir $i < j < k$. Esto produce⁴⁶

$$\begin{aligned}
 & |ijk\rangle \quad 1 \leq i < j < k \leq n \\
 & \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n = \frac{n(n-1)(n-2)}{6} \text{ casos} \\
 & 1 \times 1 + 2 \times 2 + 1 \times 1 = 6
 \end{aligned}
 \begin{array}{c}
 \begin{array}{|c|c|c|} \hline i & j & k \\ \hline \end{array} \\
 \begin{array}{|c|} \hline i \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array} \begin{array}{|c|} \hline i \\ \hline \end{array} \\
 \begin{array}{|c|} \hline k \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array} \begin{array}{|c|} \hline k \\ \hline \end{array} \\
 \begin{array}{|c|} \hline i & k \\ \hline \end{array} \\
 \begin{array}{|c|} \hline j \\ \hline \end{array}
 \end{array}
 \tag{5.18}$$

Tipo $|ijj\rangle$ con $1 \leq i < j \leq n$. Ahora al aplicar permutaciones sobre $|ijj\rangle$ (para i, j dados) se obtienen $3!/(1!2!) = 3$ estados (que por construcción forman un espacio invariante bajo permutaciones). Los idempotentes actúan igual que antes y se obtiene lo mismo poniendo $k = j$ en ec. (5.18):

$$\begin{aligned}
 & |ijj\rangle \quad 1 \leq i < j \leq n \\
 & \sum_{i=1}^n \sum_{j=i+1}^n = \frac{n(n-1)}{2} \text{ casos} \\
 & 1 \times 1 + 1 \times 2 + 0 \times 1 = 3
 \end{aligned}
 \begin{array}{c}
 \begin{array}{|c|c|c|} \hline i & j & j \\ \hline \end{array} \\
 \begin{array}{|c|} \hline i \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array} \begin{array}{|c|} \hline i \\ \hline \end{array} \\
 \begin{array}{|c|} \hline j \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array} \\
 \begin{array}{|c|} \hline i & j \\ \hline \end{array} \\
 \begin{array}{|c|} \hline j \\ \hline \end{array}
 \end{array}
 \tag{5.19}$$

El segundo $\begin{array}{|c|} \hline i \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array}$ se va porque está repetido y $\begin{array}{|c|} \hline i \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array}$ se va por tener índices repetidos en la misma columna (se anula al antisimetrizar). En total se tienen 3 estados para i, j dados.

Tipo $|iij\rangle$, $1 \leq i < j \leq n$. En este caso se tiene

$$\begin{aligned}
 & |iij\rangle \quad 1 \leq i < j \leq n \\
 & \sum_{i=1}^n \sum_{j=i+1}^n = \frac{n(n-1)}{2} \text{ casos} \\
 & 1 \times 1 + 1 \times 2 + 0 \times 1 = 3
 \end{aligned}
 \begin{array}{c}
 \begin{array}{|c|c|c|} \hline i & i & j \\ \hline \end{array} \\
 \begin{array}{|c|} \hline i \\ \hline \end{array} \begin{array}{|c|} \hline i \\ \hline \end{array} \begin{array}{|c|} \hline i \\ \hline \end{array} \\
 \begin{array}{|c|} \hline j \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array} \begin{array}{|c|} \hline j \\ \hline \end{array} \\
 \begin{array}{|c|} \hline i & j \\ \hline \end{array} \\
 \begin{array}{|c|} \hline i \\ \hline \end{array}
 \end{array}
 \tag{5.20}$$

⁴⁶ Por ejemplo, el espacio $\begin{array}{|c|} \hline i & j \\ \hline \end{array} \begin{array}{|c|} \hline k \\ \hline \end{array}$, está generado por el vector $e_{12}|ijk\rangle = s_{12}a_{13}|ijk\rangle = |ijk\rangle - |kji\rangle + |jik\rangle - |jki\rangle$. Este espacio tiene dimensión 2 ya que al aplicar permutaciones arbitrarias se obtienen dos vectores linealmente independientes.

Como era de esperar se obtiene lo mismo que en el caso $|ijj\rangle$ ya que el etiquetado de los estados no puede afectar al resultado (aunque superficialmente la justificación para eliminar algunos tableros sea distinta).

Finalmente, tipo $|iii\rangle$. Este estado ya es invariante bajo permutaciones (espacio invariante unidimensional). Tomando $i = j = k$ en ec. (5.18):

$$\begin{aligned}
 & |iii\rangle \quad 1 \leq i \leq n \\
 & \sum_{i=1}^n 1 = n \text{ casos} \\
 & 1 \times 1 + 0 \times 2 + 0 \times 1 = 1
 \end{aligned}
 \begin{array}{ccc}
 \begin{array}{|c|c|c|} \hline i & i & i \\ \hline \end{array} & \begin{array}{|c|c|} \hline i & i \\ \hline \end{array} & \begin{array}{|c|} \hline i \\ \hline \end{array} \\
 & \begin{array}{|c|c|} \hline i & i \\ \hline \end{array} & \begin{array}{|c|} \hline i \\ \hline \end{array} \\
 & \begin{array}{|c|c|} \hline i & i \\ \hline \end{array} & \begin{array}{|c|} \hline i \\ \hline \end{array}
 \end{array}
 \tag{5.21}$$

Reuniendo todos los casos se obtiene

$$V_n^3 = d_{\square\square\square} \times \begin{array}{|c|c|c|} \hline & & \\ \hline \end{array} + d_{\square\square} \times \begin{array}{|c|c|} \hline & \\ \hline \end{array} + d_{\square} \times \begin{array}{|c|} \hline \\ \hline \end{array} \quad (\text{Reducción bajo } S_3) \tag{5.22}$$

$$\begin{aligned}
 d_{\square\square\square} &= \frac{n(n-1)(n-2)}{6} + \frac{n(n-1)}{2} + \frac{n(n-1)}{2} + n = \frac{n(n+1)(n+2)}{6}, \\
 d_{\square\square} &= 2 \times \frac{n(n-1)(n-2)}{6} + \frac{n(n-1)}{2} + \frac{n(n-1)}{2} + 0 \times n = \frac{n(n+1)(n-1)}{3}, \\
 d_{\square} &= 1 \times \frac{n(n-1)(n-2)}{6} = \frac{n(n-1)(n-2)}{6}, \\
 n^3 &= d_{\square\square\square} \times 1 + d_{\square\square} \times 2 + d_{\square} \times 1.
 \end{aligned}
 \tag{5.23}$$

Al mismo tiempo la reducción bajo G_n es

$$V_n^3 = 1 \times \begin{array}{|c|c|c|} \hline & & \\ \hline \end{array} + 2 \times \begin{array}{|c|c|} \hline & \\ \hline \end{array} + 1 \times \begin{array}{|c|} \hline \\ \hline \end{array} \quad (\text{Reducción bajo } G_n) \tag{5.24}$$

Los d_λ (multiplicidad de λ en la reducción bajo S_3) son la dimensión de la irrep λ de G_n . Una de las dos irreps $\square\square$ de G_n corresponde a los estados $\{|[2, 1], \alpha, a = 1\rangle\}$ y la otra a $\{|[2, 1], \alpha, a = 2\rangle\}$. Por otra parte, todos los tensores completamente simétricos de rango 3 forman una sola irrep de G_n y lo mismo los completamente antisimétricos.

Más generalmente, por cada irrep de S_r de tipo λ hay un estado de la base de la irrep λ de G_n y por ello esta irrep tiene tantos estados como **tableros estándar de G_n** : estos son diagramas de Young de tipo λ llenados con r etiquetas $i_1, \dots, i_r = 1, \dots, n$ (con repetición) tales que la etiqueta no decrezca al moverse a la derecha por una fila y crezca estrictamente al moverse hacia abajo por una columna.

Así, por ejemplo, los tableros estándar de $GL(3, \mathbb{C})$ para $\lambda = [2^2]$ son

$$\begin{array}{|c|c|} \hline 1 & 1 \\ \hline 2 & 2 \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 2 & 3 \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 3 & 3 \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 2 & 3 \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 3 \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline 2 & 2 \\ \hline 3 & 3 \\ \hline \end{array} \quad (5.25)$$

y se concluye que la dimensión de $[2^2]$ en G_3 es 6.

Obviamente, en $GL(n, \mathbb{C})$ la dimensión es 0 (la irrep no existe) para λ 's con columnas de longitud superior a n .

Una fórmula para la dimensión es

$$d_\lambda = \frac{\prod_{i=1}^r c_i}{\prod_{i=1}^r l_i}, \quad (5.26)$$

donde l_i es la longitud del gancho de la casilla i -ésima y c_i es n para la primera casilla (izquierda arriba) aumentando en uno al moverse a la derecha y disminuyendo en uno al moverse hacia abajo. Por ejemplo

$$\dim_{G_n} [2^2] = \frac{\begin{array}{|c|c|} \hline n & n+1 \\ \hline n-1 & n \\ \hline \end{array}}{\begin{array}{|c|c|} \hline 3 & 2 \\ \hline 2 & 1 \\ \hline \end{array}} = \frac{n(n+1)(n-1)n}{3 \cdot 2 \cdot 2 \cdot 1} = \frac{n^2(n^2-1)}{12}. \quad (5.27)$$

La fórmula análoga para la dimensión de la irrep λ de S_r es

$$n_\lambda = \frac{r!}{\prod_{i=1}^r l_i}, \quad (5.28)$$

Por ejemplo

$$\dim_{S_4} [2^2] = \frac{4!}{\begin{array}{|c|c|} \hline 3 & 2 \\ \hline 2 & 1 \\ \hline \end{array}} = \frac{4!}{3 \cdot 2 \cdot 2 \cdot 1} = 2, \quad (5.29)$$

correspondiente a los dos tableros estándar de S_4 de $[2^2]$, $\begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline \end{array}$ y $\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 2 & 4 \\ \hline \end{array}$.

Ejemplo. Consideremos $n = 2$ y $r = 3$. La reducción bajo $GL(2, \mathbb{C})$ es

$$V_2^3 = 1 \times \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array} \oplus 2 \times \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \oplus 1 \times \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array}, \quad 2^3 = 1 \times 4 + 2 \times 2 + 1 \times 0. \quad (5.30)$$

Esto es completamente consistente con lo que sabemos de la serie de Clebsch-Gordan de $SU(2)$:

$$V_{j=\frac{1}{2}} \otimes V_{j=\frac{1}{2}} \otimes V_{j=\frac{1}{2}} = (V_{j=0} \oplus V_{j=1}) \otimes V_{j=\frac{1}{2}} = V_{j=\frac{3}{2}} \oplus V_{j=\frac{1}{2}} \oplus V_{j=\frac{1}{2}}, \quad (5.31)$$

con dimensiones $4 + 2 + 2 = 8$.

Aparte de la irrep $[1]$, que es la propia g^i_j que define el grupo $GL(n, \mathbb{C})$, otra irrep interesante es $[1^n]$ (una columna con n casillas). Tiene dimensión 1 y corresponde a los tensores completamente antisimétricos de rango n , y no es más que el determinante de g :

$$\begin{aligned} \psi_A^{i_1 \dots i_n} &= \varepsilon^{i_1 \dots i_n} \psi \quad \text{con} \quad \psi \equiv \psi_A^{1 \dots n}, \\ \varepsilon^{i_1 \dots i_n} &\mapsto g^{i_1}_{j_1} \dots g^{i_n}_{j_n} \varepsilon^{j_1 \dots j_n} = \varepsilon^{i_1 \dots i_n} g^{1}_{j_1} \dots g^n_{j_n} \varepsilon^{j_1 \dots j_n} = \det(g) \varepsilon^{i_1 \dots i_n}. \end{aligned} \quad (5.32)$$

Aquí $\varepsilon^{i_1 \dots i_n}$ es el **tensor de Levi-Civita** que está completamente definido por ser totalmente antisimétrico y $\varepsilon^{1 \dots n} = 1$. En efecto, el determinante define una representación ya que $\det(g_1 g_2) = \det(g_1) \det(g_2)$.

Más generalmente, para una irrep tensorial de $GL(n, \mathbb{C})$ dada por un tablero λ , cada columna de longitud n produce un factor $\det(g)$. Por ejemplo,

$$\begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} = \det(g)^2 \times \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \quad \text{en } GL(3, \mathbb{C}). \quad (5.33)$$

5.3. Serie de Clebsch-Gordan para irreps tensoriales de $GL(n, \mathbb{C})$

Como se ha visto, cada irrep tensorial T^r de $GL(n, \mathbb{C})$ viene dada por un diagrama de Young y la serie de C-G puede escribirse como **producto de tableros**. Este producto se puede hacer a nivel de tableros, sin referirse a un n concreto. Para multiplicar dos diagramas λ y μ , en primer lugar etiquetamos el segundo diagrama con 1's en la primera fila, 2's en la segunda, y así sucesivamente. Así por ejemplo, si queremos multiplicar $[2]$ con $[2, 1^2]$,

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array} \times \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 2 & \\ \hline 3 & \\ \hline \end{array}. \quad (5.34)$$

A continuación se van añadiendo casillas etiquetadas del segundo tablero al primero de todas las formas posibles tales que se obtenga un **diagrama admisible** y un **tablero admisible** distinto.⁴⁷ Como diagrama, la construcción es admisible cuando la longitud de una fila superior es mayor o igual que la longitud de otra fila inferior. Como tablero, éste es admisible si 1) no hay dos etiquetas iguales en una misma columna, y 2) leído de derecha a izquierda fila por fila de arriba a abajo, el número de etiquetas i en ningún momento supera el número de etiquetas $i - 1$, y esto para cualquier i . Se guardan todos los tableros admisibles distintos, y se quitan la etiquetas. Al hacer esto pueden quedar diagramas repetidos, lo cual indica que esas irreps aparecen con multiplicidad mayor que uno en la reducción de $\lambda \otimes \mu$.

En el ejemplo de antes, después de añadir las casillas con etiquetas 1':

(5.35)

Después de añadir las casillas con etiquetas 2:

(5.36)

Finalmente añadiendo las casillas con etiqueta 3 (omitimos ya los cuatro tableros inadmisibles que se obtienen poniendo un 3 en la primera fila):

(5.37)

⁴⁷En sentido estricto usamos *diagrama* para la estructura de casillas (una partición de r) y un *tablero* para un diagrama etiquetado. En sentido más vago se usa tablero para un diagrama de Young con o sin etiquetas.

Finalmente

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \times \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} + \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}. \quad (5.38)$$

En realidad, puesto que el producto de tableros es conmutativo es más conveniente calcular $[2, 1^2] \times [2]$

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \times \begin{array}{|c|c|} \hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline \square & \square & 1 & 1 \\ \hline \square & \square & 1 & 1 \\ \hline \square & \square & 1 & 1 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline \square & \square & 1 \\ \hline \square & \square & 1 \\ \hline \square & \square & 1 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline \square & \square & 1 \\ \hline \square & \square & 1 \\ \hline \square & \square & 1 \\ \hline \end{array} + \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} + \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} + \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}. \quad (5.39)$$

Nótese que cuando el producto de tableros se aplica a un n concreto, los tableros con columnas con longitud mayor que n tienen dimensión cero y sobran. Así, por ejemplo

$$(n = 2) \quad \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \times \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} = \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} + \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \quad 1 \times 2 = 2 + 0. \quad (5.40)$$

En $SU(2)$ $\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}$ y $\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}$ tienen dimensión 2 y corresponden a $j = \frac{1}{2}$ y $\begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array}$ tiene dimensión 1 y corresponde a $j = 0$.

Al calcular series de C-G una comprobación extraordinariamente útil es verificar que las dimensiones a ambos lados sean iguales. (Compruébese ec. (5.38), bien para n genérico o para algún n no totalmente trivial.)

5.4. Reducibilidad de representaciones tensoriales bajo $SU(n)$

Veamos que las representaciones tensoriales que son irreducibles en $GL(n, \mathbb{C})$ también lo son en $SU(n)$, y por tanto en $U(n)$ y $SL(n, \mathbb{C})$. Nótese que es trivial que si una representación es irreducible para un subgrupo lo es para el grupo, pero no al revés.

La primera observación es que las representaciones tensoriales de $GL(n, \mathbb{C})$ son **analíticas**, es decir, las matrices $D(g)$ dependen analíticamente de los elementos de matriz g^i_j . Por otro lado, la extensión analítica de $SU(n)$ es el grupo $SL(n, \mathbb{C})$ (matrices complejas de determinante 1). Esto se deduce de que

$$SU(n) \ni g = e^{-ia^i X_i}, \quad X_i = X_i^\dagger, \quad \text{tr}(X_i) = 0, \quad a \in \mathbb{R}^n, \quad (5.41)$$

y si se toma $a \in \mathbb{C}^n$ se obtiene un elemento arbitrario de $SL(n, \mathbb{C})$, $g = e^A$, $\text{tr}(A) = 0$. Igualmente, la extensión analítica de $U(n)$ es $GL(n, \mathbb{C})$.

Supongamos que $D(g)$ es una representación tensorial de $GL(n, \mathbb{C})$ tal que es reducible en $SU(n)$, es decir, cuando $g \in SU(n)$ $D(g) = \begin{pmatrix} A(g) & 0 \\ B(g) & C(g) \end{pmatrix}$. Entonces, por extensión analítica tendrá la misma forma en $SL(n, \mathbb{C})$. Y también será reducible en $GL(n, \mathbb{C})$: en efecto, todo $g \in GL(n, \mathbb{C})$ puede escribirse como $g = zg'$ con $z \in \mathbb{C}$ y $g' \in SL(n, \mathbb{C})$ (de hecho $z^n = \det g$). Por ser $D(g)$ una función **homogénea** de grado r en g (ver ec. (5.7)) se tiene $D(g) = z^r D(g') = \begin{pmatrix} z^r A(g') & 0 \\ z^r B(g') & z^r C(g') \end{pmatrix}$, que es reducible. Esto demuestra que las representaciones irreducibles de $GL(n, \mathbb{C})$ lo siguen siendo cuando se restringen a $SU(n)$.

5.5. Otras representaciones de $GL(n, \mathbb{C})$

Hasta ahora hemos visto las representaciones de tipo $g \otimes \dots \otimes g$ de $GL(n, \mathbb{C})$ pero hay más. Hay cuatro representaciones básicas, a saber, g , g^{-1T} , g^* y $g^{-1\dagger}$. Así por ejemplo,

$$(g_1 g_2)^{-1\dagger} = g_1^{-1\dagger} g_2^{-1\dagger}. \quad (5.42)$$

En componentes (por conveniencia usamos índices $a, b, \dots = 1, \dots, n$ en vez de i, j, \dots)

$$\begin{aligned} g : \psi^a &\mapsto g^a_b \psi^b, & g^{-1T} : \psi_a &\mapsto (g^{-1})^b_a \psi_b \equiv (g^{-1T})^b_a \psi_b, \\ g^* : \psi_{\dot{a}} &\mapsto (g^{\dot{a}}_{\dot{b}})^* \psi_{\dot{b}} \equiv (g^*)^{\dot{a}}_{\dot{b}} \psi_{\dot{b}}, & g^{-1\dagger} : \psi^{\dot{a}} &\mapsto ((g^{-1})^{\dot{b}}_{\dot{a}})^* \psi^{\dot{b}} \equiv (g^{-1\dagger})^{\dot{a}}_{\dot{b}} \psi^{\dot{b}}. \end{aligned} \quad (5.43)$$

Los índices con punto recorren los mismos valores, $\dot{a}, \dot{b} = 1, \dots, n$, el punto sirve para indicar que esos índices no se transforman con g sino con g^* .

Igualmente, para las bases de los espacios correspondientes V_n , \tilde{V}_n , V_n^* y \tilde{V}_n^*

$$\begin{aligned} g|e_a\rangle &= g^b_a |e_b\rangle, & g^{-1T}|e^a\rangle &= (g^{-1T})^b_a |e^b\rangle, \\ g^*|e_{\dot{a}}\rangle &= (g^{\dot{a}}_{\dot{b}})^* |e_{\dot{b}}\rangle, & g^{-1\dagger}|e^{\dot{a}}\rangle &= ((g^{-1})^{\dot{b}}_{\dot{a}})^* |e_{\dot{b}}\rangle. \end{aligned} \quad (5.44)$$

Estas cuatro representaciones son irreducibles e inequivalentes para $GL(n, \mathbb{C})$. Dentro del subgrupo $U(n)$ (matrices unitarias) sólo dos de ellas son inequivalentes, ya que cuando g es unitaria $g^{-1\dagger} = g$ y $g^{-1T} = g^*$ (y siguen siendo irreducibles).

Todas las representaciones de dimensión finita de $GL(n, \mathbb{C})$ se obtienen como subespacios del producto tensorial de las cuatro irreps básicas:

$$V_n^r \otimes \tilde{V}_n^s \otimes \tilde{V}_n^{*p} \otimes V_n^{*q} \quad r, s, p, q = 0, 1, 2, \dots \quad (5.45)$$

Los correspondientes elementos son tensores de tipo $T_{s,q}^{r,p}$ con componentes

$$\psi_{b_1 \dots b_s, \dot{b}_1 \dots \dot{b}_q}^{a_1 \dots a_r, \dot{a}_1 \dots \dot{a}_p}. \quad (5.46)$$

Los tensores V_n^r considerados hasta ahora son de tipo $T^r \equiv T_{0,0}^{r,0}$.

Es inmediato comprobar que si las variables $\psi_{b_1 \dots b_s, \dot{b}_1 \dots \dot{b}_q}^{a_1 \dots a_r, \dot{a}_1 \dots \dot{a}_p}$ se transforman como las componentes de un tensor de tipo $T_{s,q}^{r,p}$, entonces $(\psi_{b_1 \dots b_s, \dot{b}_1 \dots \dot{b}_q}^{a_1 \dots a_r, \dot{a}_1 \dots \dot{a}_p})^*$ (**representación conjugada**) se transforman como las componentes de un tensor de tipo $T_{p,r}^{q,s}$.

Las representaciones $T_s^r \equiv T_{s,0}^{r,0}$ son **analíticas** (su transformación depende analíticamente de g) y las $T_{0,q}^{0,p}$ son **antianalíticas** (funciones conjugadas de analíticas). Todas las irreps (de dimensión finita) de $GL(n, \mathbb{C})$ son separables, de la forma analítica por antianalítica

$$\psi_{b_1 \dots b_s}^{a_1 \dots a_r} (\phi_{d_1 \dots d_p}^{c_1 \dots c_q})^*. \quad (5.47)$$

Este resultado se deduce notando que las irreps de un producto directo de grupos se obtienen como el producto tensorial de irreps, $D^\mu(g_1)^i_j D^\nu(g_2)^k_l$. En el presente caso lo que se tiene es $D^\mu(g)^i_j D^\nu(g^*)^k_l$ y a efectos prácticos las variables g y g^* se pueden considerar como variables independientes por lo que es irreducible.

La consecuencia es que sin pérdida de generalidad nos podemos restringir a estudiar las irreps analíticas, T_s^r , de $GL(n, \mathbb{C})$.

5.6. Representaciones de tipo T_s^r de $GL(n, \mathbb{C})$ y $U(n)$

Cuando nos restringimos a $U(n)$ (subgrupo de matrices unitarias) $g^{-1\dagger} = g$ y $g^* = g^{-1T}$, en consecuencia en este subgrupo no hay diferencia entre índices con y sin punto. Todas las irreps de $U(n)$ se encuentran al reducir $V_n^r \otimes \check{V}_n^s$. La demostración que se hizo anteriormente de que las irreps de $GL(n, \mathbb{C})$ de tipo T^r se mantienen irreducibles en $SU(n)$ (Sec. 5.4) también se aplica al caso más general T_s^r ya que sólo se usó que las representaciones eran analíticas y homogéneas en g lo cual también vale para T_s^r .⁴⁸ Además, representaciones inequivalentes de $GL(n, \mathbb{C})$ siguen siendo inequivalentes al restringirlas a $U(n)$, por extensión analítica.⁴⁹

⁴⁸No se aplica al caso general $T_{s,q}^{r,p}$. Por ejemplo, los tensores $\psi^{a,b}$ forman una representación irreducible de $GL(n, \mathbb{C})$ en cambio en $U(n)$ es del tipo $\psi^{a,b}$ que es reducible ya que a y b no están simetrizados/antisimetrizados.

⁴⁹Esto ya no es cierto para $SU(n)$. Como se verá, dos irreps inequivalentes de $U(n)$ pueden pasar a ser equivalentes al restringirlas a $SU(n)$.

Nótese que para $U(n)$ $(T_s^r)^* = T_r^s$. Aparte de esto, por extensión analítica, todas las propiedades lineales de T_s^r bajo $GL(n, \mathbb{C})$ valen para $U(n)$ y viceversa.

Sobre los tensores $\psi_{b_1 \dots b_s}^{a_1 \dots a_r}$ actúan los grupos S_r y S_s de permutaciones de los índices contravariantes y covariantes por separado, y de nuevo su acción conmuta con $GL(n, \mathbb{C})$. Entonces, por la misma construcción que para T^r , se obtienen subespacios invariantes de tensores asociados a tipos de simetría de permutaciones (diagramas de Young) λ (r casillas) y μ (s casillas) para los índices contravariantes (a_1, \dots, a_r) y covariantes (b_1, \dots, b_s) por separado. Sin embargo, a diferencia del caso $s = 0$, estos subespacios no son irreducibles en general.

El motivo es que el tensor δ_b^a es invariante

$$\delta_b^a \mapsto g^{a' a} (g^{-1T})_b^{b'} \delta_{b'}^{a'} = g^{a' a} (g^{-1})^{a'}_b = \delta_b^a. \quad (5.48)$$

Más generalmente son invariantes los tensores formados por productos de factores $\delta_{b_j}^{a_i}$ y sumas de éstos. Aparte de éstos no hay otros tensores invariantes.⁵⁰

Así, por ejemplo, el espacio de tensores de la forma ψ_b^a , de dimensión n^2 , contiene un subespacio invariante de dimensión 1: el dado por el tensor con componentes δ_b^a

$$|e_a\rangle \otimes |e^a\rangle \mapsto |e_a\rangle \otimes |e^a\rangle. \quad (5.49)$$

La invariancia de δ_b^a equivale a decir que la traza ψ_a^a , obtenida por contracción del índice contravariante con el covariante, es un invariante bajo el grupo.⁵¹ El espacio ψ_b^a se descompone en dos espacios invariantes irreducibles, con dimensiones $n^2 - 1$ (tensores sin traza) y 1 (tensores proporcionales a δ_b^a):

$$\psi_b^a = \hat{\psi}_b^a + \psi \delta_b^a \quad \text{con} \quad \psi = \frac{1}{n} \psi_a^a, \quad \hat{\psi}_a^a = 0. \quad (5.50)$$

En general, tomando traza respecto de dos índices cualesquiera (uno contravariante y otro covariante) de un tensor de tipo T_s^r se obtiene de nuevo un tensor, de tipo T_{s-1}^{r-1} . Esto permite reducir T_s^r en espacios invariantes, a saber, como suma de **tensores sin traza** respecto de ningún par de índices de tipo T_{s-k}^{r-k} ($k = 0, 1, \dots$) multiplicados por tensores invariantes tipo $\delta \otimes \dots \otimes \delta$. Por ejemplo,

$$\psi_c^{ab} = \hat{\psi}_c^{ab} + \psi^a \delta_c^b + \phi^b \delta_c^a \quad (5.51)$$

⁵⁰Por el lema de Schur, cualquier matriz que conmute con todos los g^{ij} debe ser múltiplo de la identidad, es decir, δ^i_j .

⁵¹Como es sabido, la traza de una aplicación lineal no depende de la base: es un invariante bajo el grupo de cambios de base $GL(n, \mathbb{C})$.

donde $\hat{\psi}_c^{ab}$ no tiene trazas: $\hat{\psi}_a^{ab} = \hat{\psi}_b^{ab} = 0$. Basta tomar trazas a ambos lados de la ecuación y resolver en ψ^a y ϕ^b

$$\begin{aligned} \psi_c^{cb} &= \psi^b + n\phi^b, & \psi_c^{ac} &= n\psi^a + \phi^a, \\ \psi^a &= \frac{1}{n^2 - 1}(n\psi_c^{ac} - \psi_c^{ca}), & \phi^a &= \frac{1}{n^2 - 1}(n\psi_c^{ca} - \psi_c^{ac}). \end{aligned} \quad (5.52)$$

$\hat{\psi}_c^{ab}$, $\psi^a \delta_c^b$ y $\phi^b \delta_c^a$ subtienden tres espacios invariantes, los dos últimos con dimensión n y el primero con dimensión $n^3 - 2n$.

Obviamente esta reducción es simétrica respecto de todos los índices contravariantes y respecto de todos los índices covariantes, y en consecuencia es compatible con la reducción por tipo de simetría de permutaciones, que se puede aplicar antes o después. Así si ψ_c^{ab} era un tensor simétrico/antisimétrico en ab se tendrá

$$\psi_c^{ab} = \hat{\psi}_c^{ab} + \psi^a \delta_c^b \pm \psi^b \delta_c^a \quad (5.53)$$

con $\hat{\psi}_c^{ab}$ simétrico/antisimétrico y sin trazas. En este caso hay dos subespacios invariantes $\psi^a \delta_c^b \pm \psi^b \delta_c^a$, de dimensión n , y $\hat{\psi}_c^{ab}$, de dimensión $n^2(n \pm 1)/2 - n$. Estos espacios ya son irreducibles.

Aplicando las dos reducciones mencionadas (permutaciones y trazas) se obtienen las representaciones irreducibles de tipo T_s^r de $\text{GL}(n, \mathbb{C})$ y $\text{U}(n)$, a saber, tensores sin traza con simetría bajo permutaciones caracterizada por dos diagramas de Young (λ, μ) (uno para los índices contravariantes y otro para los covariantes). Diagramas distintos corresponden a irreps inequivalentes. Sin embargo no todos los pares (λ, μ) pueden aparecer. El motivo es que al imponer un tipo de simetría y al mismo tiempo que el tensor no tenga traza el sistema de ecuaciones puede quedar sobredeterminado y la solución reducirse al espacio nulo. La regla es que el número de filas de λ más el número de filas de μ debe ser menor o igual que n . Equivalentemente, cada irrep (analítica) de $\text{GL}(n, \mathbb{C})$ y cada irrep de $\text{U}(n)$, está unívocamente caracterizada por cada conjunto de n números enteros ordenados

$$[m_1, \dots, m_n] \quad m_1 \geq m_2 \geq \dots \geq m_n \quad m_i \in \mathbb{Z}. \quad (5.54)$$

Los m_i positivos definen el diagrama λ y los negativos el μ . Por ejemplo

$$n = 7 \quad [3, 3, 1, 0, 0, -1, -2], \quad \lambda = [3, 3, 1] = \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & & \\ \hline \end{array}, \quad \mu = [2, 1] = \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}. \quad (5.55)$$

Tal y como se ha visto en los ejemplos anteriores la dimensión de una irrep (λ, μ) no es directamente el producto de dimensiones de λ y μ . En general es menor por la ligadura impuesta de ser tensores sin traza. La dimensión es la misma que da $\text{SU}(n)$ y se obtendrá más adelante.

Para $U(n)$, la representación conjugada de (λ, μ) es $(\lambda, \mu)^* = (\mu, \lambda)$, o equivalentemente $[-m_n, \dots, -m_1]$.

Con la notación de ec. (5.54) también es muy fácil obtener la **ley de ramificación** de $U(n)$, es decir, cómo se reducen sus irreps con respecto al subgrupo $U(n-1)$ definido por $g = \begin{pmatrix} g' & 0 \\ 0 & 1 \end{pmatrix}$ donde $g' \in U(n-1)$. Las irreps de $U(n-1)$ que aparecen al reducir $[m_1, \dots, m_n]$ son $[k_1, \dots, k_{n-1}]$ con $m_i \geq k_i \geq m_{i+1}$ y cada una aparece exactamente una vez. La misma regla se aplicará a $SU(n)$.

Así, por ejemplo, la representación ψ_j^i de $U(3)$ ($i, j = 1, 2, 3$), es $([1], [1]) = [1, 0, -1]$, de dimensión $3^2 - 1 = 8$. Su ramificación (reducción bajo $U(2)$) produce $[1, 0] = ([1], [])$, $[1, -1] = ([1], [1])$, $[0, -1] = ([], [1])$, y $[0, 0] = ([], [])$, es decir, los tensores de $U(2)$ ψ^a , ψ_b^a , ψ_a y ψ , ($a, b = 1, 2$) con dimensiones $2 + 3 + 2 + 1 = 8$,

$$(\square, \square)_{U(3)} = (\square, \bullet) \oplus (\square, \square) \oplus (\bullet, \square) \oplus (\bullet, \bullet)_{U(2)}. \quad (5.56)$$

(El símbolo \bullet indica la representación trivial, con tablero vacío $[\]$.)

5.7. Representaciones irreducibles de $SU(n)$

Para los subgrupos $SU(n)$ y su extensión analítica $SL(n, \mathbb{C})$, la condición $\det(g) = 1$ introduce un nuevo tensor invariante, el tensor de Levi-Civita. En efecto, por la ec. (5.32)

$$\varepsilon^{i_1 \dots i_n} \mapsto g^{i_1}_{j_1} \dots g^{i_n}_{j_n} \varepsilon^{j_1 \dots j_n} = \det(g) \varepsilon^{i_1 \dots i_n} = \varepsilon^{i_1 \dots i_n}, \quad g \in SL(n, \mathbb{C}) \quad (5.57)$$

y lo mismo $\varepsilon_{i_1 \dots i_n}$.

Este tensor invariante hace que representaciones irreducibles **inequivalentes** de $U(n)$ pasen a ser **equivalentes** en $SU(n)$. Un ejemplo es la representación definida por el propio determinante, $g \mapsto \det(g)$ que corresponde a $([1^n], [])$ en $U(n)$ y equivale a $([], [])$ (la representación trivial $g \mapsto 1$) cuando nos restringimos a $SU(n)$.

Más generalmente en $SU(n)$ los tensores antisimétricos contravariantes de rango r , $([1^r], [])$, son equivalentes a los tensores antisimétricos covariantes de rango $n-r$, $([], [1^{n-r}])$.

Para ver esto, sea $\psi^{i_1 \dots i_r}$ un tensor completamente antisimétrico ($\binom{n}{r}$ variables)

$$\psi^{i_1 \dots i_r} \mapsto \psi^{j_1 \dots j_r} = g^{i_1}_{j_1} \dots g^{i_r}_{j_r} \psi^{j_1 \dots j_r} \quad (5.58)$$

e introducimos su **dual** $\phi_{i_{r+1}, \dots, i_n}$, también completamente antisimétrico (igual número de variables $\binom{n}{n-r} = \binom{n}{r}$), mediante

$$\psi^{i_1 \dots i_r} = \frac{1}{(n-r)!} \varepsilon^{i_1 \dots i_n} \phi_{i_{r+1}, \dots, i_n}, \quad \phi_{i_{r+1}, \dots, i_n} = \frac{1}{r!} \varepsilon_{i_1 \dots i_n} \psi^{i_1 \dots i_r}. \quad (5.59)$$

Se trata de demostrar que $\phi_{i_{r+1}, \dots, i_n}$ así definido es a su vez un tensor covariante si $g \in \text{SU}(n)$:

$$\phi_{i_{r+1}, \dots, i_n} \mapsto \phi'_{i_{r+1}, \dots, i_n} = (g^{-1})^{j_{r+1} i_{r+1}} \dots (g^{-1})^{j_n i_n} \phi_{j_{r+1}, \dots, j_n}. \quad (5.60)$$

Multiplicando la ecuación intermedia en ec. (5.57) por $n-r$ factores g^{-1} , se obtiene la identidad

$$g^{i_1 j_1} \dots g^{i_r j_r} \varepsilon^{j_1, \dots, j_n} = \det(g) (g^{-1})^{j_{r+1} i_{r+1}} \dots (g^{-1})^{j_n i_n} \varepsilon^{i_1, \dots, i_n} \quad 0 \leq r \leq n. \quad (5.61)$$

Ahora podemos ver cómo se transforma ϕ :

$$\begin{aligned} \psi'^{i_1 \dots i_r} &= g^{i_1 j_1} \dots g^{i_r j_r} \psi^{j_1 \dots j_r} = g^{i_1 j_1} \dots g^{i_r j_r} \frac{1}{(n-r)!} \varepsilon^{j_1, \dots, j_n} \phi_{j_{r+1}, \dots, j_n} \\ &= \frac{1}{(n-r)!} \varepsilon^{i_1, \dots, i_n} \det(g) (g^{-1})^{j_{r+1} i_{r+1}} \dots (g^{-1})^{j_n i_n} \phi_{j_{r+1}, \dots, j_n} \\ &\equiv \frac{1}{(n-r)!} \varepsilon^{i_1, \dots, i_n} \phi'_{i_{r+1}, \dots, i_n}. \end{aligned} \quad (5.62)$$

Esto implica que $\phi_{i_{r+1}, \dots, i_n}$ es un tensor cuando $\det(g) = 1$. La misma comprobación se puede hacer en sentido contrario, suponiendo que ϕ es un tensor covariante y comprobando que su dual ψ es también un tensor.

Más generalmente, si se tiene un tensor de tipo (λ, μ) , cada columna de longitud r de λ corresponde a r índices contravariantes antisimetrizados. Contrayendo con el tensor de Levi-Civita se transforman en $n-r$ índices covariantes antisimetrizados, es decir, en una columna de longitud $n-r$ a añadir en μ , y viceversa. Por este procedimiento de mover columnas entre λ y μ se obtienen nuevos pares (λ', μ') que corresponden a irreps que son equivalentes en $\text{SU}(n)$ (ídem $\text{SL}(n, \mathbb{C})$). Así por ejemplo, en $\text{SU}(3)$ se tienen las siguientes equivalencias

$$\begin{aligned} (\begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array}, \bullet) &\equiv (\square, \square) \equiv (\bullet, \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array}), \\ (\begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array}, \bullet) &\equiv (\begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array}, \square) \equiv (\square, \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array}) \equiv (\bullet, \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array}), \end{aligned} \quad \text{en } \text{SU}(3). \quad (5.63)$$

Podemos comprobar que la equivalencia conserva correctamente la dimensión. Por ejemplo, en $SU(n)$ las irreps $([1], [1])$ (tensores ψ_j^i) deben tener la misma dimensión que $([2, 1^{n-2}], [])$ (tensores ψ^{i_1, \dots, i_n} simétricos en los dos primeros índices y “antisimétricos” en los $n - 1$ últimos). Aplicando la ec. (5.26), la dimensión del tablero $([2, 1^{n-2}], [])$ es $n^2 - 1$. La dimensión de ψ_j^i sería n^2 por los dos índices independientes, pero se pierde un grado de libertad al imponer que la traza se anule, quedando una dimensión $n^2 - 1$.

Mediante estas equivalencias, dada una irrep (λ, μ) siempre se puede elegir otra equivalente dentro de $SU(n)$, que podemos indicar como $(\lambda \tilde{\mu}, \bullet)$ (cada columna de longitud r de μ se ha cambiado por una columna de longitud $n - r$ y se ha añadido a λ). Este es un tensor de tipo T^r , que únicamente tiene índices contravariantes. Se concluye entonces que reduciendo las representaciones tensoriales T^r se tienen todas las irreps de $SU(n)$.

Nótese que no todos los tableros λ producen representaciones inequivalentes en $SU(n)$, ya que n índices antisimetrizados (que producirían $\det(g)$) equivalen a la representación trivial por $\det(g) = 1$. Así

$$\begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} \equiv \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \square & \square & \square \\ \hline \end{array} \equiv \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \quad \text{en } SU(3). \quad (5.64)$$

En $SU(n)$ las irreps inequivalentes corresponden a tensores contravariantes (o covariantes) con tableros λ con a lo sumo $n - 1$ filas.

La representación conjugada de una dada λ viene dada por el **tablero dual** $\tilde{\lambda}$ (que depende de n), obtenido cambiando cada columna de longitud r de λ por una columna de longitud $n - r$ en $\tilde{\lambda}$. Esto se deduce de $(\lambda, \bullet)^* = (\bullet, \lambda) \equiv (\tilde{\lambda}, \bullet)$. Así, en $SU(3)$ $\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}$ es autoconjugada, y la conjugada de $\begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array}$ es $\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}$:

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \quad \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \square \\ \hline \end{array} \quad \text{tableros duales en } SU(3). \quad (5.65)$$

Una representación y su conjugada tienen la misma dimensión.

A menudo, junto con la notación $\lambda = [m_1, \dots, m_{n-1}, 0]$ ($n - 1$ filas a lo sumo) se utiliza la notación alternativa $(l_1, \dots, l_{n-1}) \equiv (m_1 - m_2, \dots, m_{n-2} - m_{n-1}, m_{n-1} - 0)$ de modo que el dual se obtiene simplemente como (l_{n-1}, \dots, l_1) .

Para tensores T^r no hay ligaduras de tipo traza, por lo cual la dimensión de una representación de tipo λ es la dada en la ec. (5.26). Para una representación irreducible de tipo T_s^r con tableros

(λ, μ) , simplemente se obtiene la irrep puramente contravariante equivalente bajo $SU(n)$, $(\lambda \tilde{\mu}, \bullet)$, y se calcula su dimensión. Esto vale para todos los grupos $GL(n, \mathbb{C})$, $SL(n, \mathbb{C})$, $U(n)$ y $SU(n)$ ya que la irrep no se reduce más al restringirla a $SU(n)$.

Por el mismo motivo, la ley de ramificación de $U(n)$ también se aplica a $SU(n)$.

Ejemplo. Consideremos la irrep $[2, 1]$ de $SU(3)$, de dimensión 8. Para ver cómo se reduce bajo $SU(2)$ la reescribimos como $[2, 1, 0]$ y aplicamos la prescripción $m_i \geq k_i \geq m_{i+1}$. Esto nos da $[2, 1] \equiv [1]$, $[1, 1] \equiv \bullet$, $[2, 0] \equiv [2]$ y $[1, 0] \equiv [1]$,

$$\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}_{SU(3)} = \square \oplus \bullet \oplus \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}_{SU(2)} \oplus \square \quad 8 = 2 + 1 + 3 + 2. \quad (5.66)$$

Puesto que $[2, 1]$ es equivalente a $[1, 0, -1]$ en $SU(3)$ (ec. (5.63)) esta reducción es equivalente a la indicada en ec. (5.56). Indica que el octete de $SU(3)$ se descompone en un triplete ($I = 1$), dos dobletes ($I = 1/2$) y un singlete ($I = 0$) de $SU(2)$. Por ejemplo el octete de mesones pseudoscalares contiene un triplete de isospín, el pion, un doblete de kaones, otro doblete de antikaones y la eta que es singlete de isospín. Otro ejemplo es la descomposición del decuplete $\frac{3}{2}^+$ en estados Δ , Σ , Ξ y Ω , con isospines $I = \frac{3}{2}$, 1 , $\frac{1}{2}$ y 0 :

$$\begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & & \\ \hline \square & & \\ \hline \end{array}_{SU(3)} = \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & & \\ \hline \square & & \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \oplus \square \oplus \bullet_{SU(2)} \quad 10 = 4 + 3 + 2 + 1. \quad (5.67)$$

La ley de ramificación permite identificar unívocamente cada estado de una irrep de $SU(n)$, usando la cadena canónica (es decir, cada irrep de un subgrupo aparece a lo sumo una vez)

$$U(n) \supset U(n-1) \supset \dots \supset U(2) \supset U(1). \quad (5.68)$$

Ya lo hemos visto para $SU(2) \supset U(1) \equiv SO(2)$, los estados son $|j, m\rangle$ correspondientes a \mathbf{J}^2 , J_3 . j es la irrep de $SU(2)$ y m la irrep de $U(1) \equiv SO(2)$, el grupo de rotaciones alrededor del eje z .

Para los estados del octete de $SU(3)$, con la notación $|\lambda_{U(3)}, \lambda_{U(2)}, \lambda_{U(1)}\rangle$

$$\begin{array}{|c|c|c|} \hline \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \rangle & \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \end{array} \rangle \\ \hline \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \rangle & \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \rangle \\ \hline \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \end{array} \rangle & & \\ \hline \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \end{array}, \begin{array}{|c|} \hline \square \\ \hline \end{array} \rangle & \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array}, \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \rangle & \\ \hline \end{array} \quad (5.69)$$

La serie de Clebsch-Gordan que vimos para $GL(n, \mathbb{C})$ en T^r , multiplicando tableros, se aplica inmediatamente a $SU(n)$.

En particular, en $SU(2)$, los tableros sólo tienen una fila y son del tipo $[r]$: tensores completamente simétricos con r índices $(i_1, \dots, i_r = 1, 2)$. Aplicando la ec. (5.26), su dimensión es $r + 1 = 2j + 1$, es decir, $j = r/2$. Los estados con momento angular j se representan por tensores completamente simétricos con $2j$ índices. Si se calcula la serie de C-G se obtiene el resultado usual. Por ejemplo

$$\begin{aligned}
 \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array} \otimes \begin{array}{|c|c|} \hline \square & \square \\ \hline \end{array} &= \begin{array}{|c|c|c|c|c|} \hline \square & \square & \square & \square & \square \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & & \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \square \\ \hline \end{array} \\
 &= \begin{array}{|c|c|c|c|c|} \hline \square & \square & \square & \square & \square \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array} \oplus \begin{array}{|c|} \hline \square \\ \hline \end{array} .
 \end{aligned} \tag{5.70}$$

Igualmente, para $SU(3)$ el producto de dos representaciones adjuntas (ψ_j^i , o $[2, 1]$, octetes) produce

$$\begin{aligned}
 \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \otimes \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} &= \bullet \oplus \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \oplus \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array} \oplus \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & & \\ \hline \end{array} \oplus \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \\ \hline \end{array} \\
 \mathbf{8 \otimes 8 = 1 \oplus 8 \oplus 8 \oplus 10 \oplus 10^* \oplus 27.}
 \end{aligned} \tag{5.71}$$

Otro ejemplo en $SU(3)$

$$\begin{aligned}
 \begin{array}{|c|} \hline \square \\ \hline \end{array} \otimes \begin{array}{|c|} \hline \square \\ \hline \square \\ \hline \end{array} &= \bullet \oplus \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \\ \hline \end{array} \\
 \mathbf{3 \otimes 3^* = 1 \oplus 8.}
 \end{aligned} \tag{5.72}$$

Es interesante notar que, para el caso particular de $SU(3)$, además de elegir una forma canónica de tipo T^r para sus irreps, también se puede elegir representarlas canónicamente por tensores T_s^r sin traza, **completamente simétricos** en los índices **contravariantes** y en los índices **covariantes**. En efecto, como en T^r λ sólo tiene dos filas, las columnas de longitud 2 se pueden dualizar a columnas de longitud 1 para los índices covariantes. Por ejemplo

$$\psi^{i_1, \dots, i_5, i_6, \dots, i_8} \sim \left(\begin{array}{|c|c|c|c|c|} \hline \square & \square & \square & \square & \square \\ \hline \square & \square & & & \\ \hline \end{array}, \bullet \right) \equiv \left(\begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \end{array}, \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \end{array} \right) \sim \psi_{j_1 j_2 j_3}^{i_1 i_2} \quad \mathbf{SU(3)} \tag{5.73}$$

En esta forma es fácil verificar que las representaciones de $SU(3)$ así definidas son irreducibles [2]. En efecto, por teoría general de representaciones, se sabe que en la serie de C-G del producto

de dos irreps, μ y ν

$$\mu \otimes \nu = \bigoplus_{\lambda} \langle \lambda | \mu, \nu \rangle \lambda, \quad (5.74)$$

la multiplicidad de la representación trivial es 1 si μ y ν son conjugadas y 0 en otro caso

$$\langle \mathbf{1} | \mu, \nu \rangle = \delta_{\mu, \nu^*} \quad (5.75)$$

Por tanto si se multiplica una representación (en general reducible) por su conjugada

$$\left(\bigoplus_{\mu} n_{\mu} \mu \right) \otimes \left(\bigoplus_{\nu} n_{\nu} \nu \right)^* = \bigoplus_{\mu, \nu, \lambda} n_{\mu} n_{\nu} \langle \lambda | \mu, \nu^* \rangle \lambda = (\sum_{\lambda} n_{\lambda}^2) \mathbf{1} \oplus \dots, \quad (5.76)$$

la multiplicidad de la irrep trivial nos da información de cómo se reduce, y es irreducible si y sólo si dicha multiplicidad es 1.

La irrep trivial corresponde a un tensor invariante. Al multiplicar un tensor simétrico en índices contravariantes y covariantes por separado y sin traza, por su conjugado

$$\psi_{j_1, \dots, j_s}^{i_1, \dots, i_r} (\psi_{j'_1, \dots, j'_s}^{i'_1, \dots, i'_r})^* \equiv \psi_{j_1, \dots, j_s}^{i_1, \dots, i_r} \psi_{i'_1, \dots, i'_r}^{*j'_1, \dots, j'_s}, \quad (5.77)$$

dado que no tiene traza, sólo se puede formar un *único* tensor invariante, contrayendo todos los índices i con los i' , y los j con j' . Esto implica que estas representaciones son irreducibles.

5.8. Matrices de Gell-Mann

Las matrices de Gell-Mann son una generalización de las matrices de Pauli para SU(3),

$$\begin{aligned} \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, & \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, & \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, & \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, & \lambda_8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \end{aligned} \quad (5.78)$$

Aquí se ve que el rango de SU(3) es 2, por ejemplo λ_3 y λ_8 conmutan.

Las matrices de Gell-Mann satisfacen las relaciones

$$\lambda_i^\dagger = \lambda_i, \quad \text{tr}(\lambda_i) = 0, \quad \text{tr}(\lambda_i \lambda_j) = 2\delta_{ij} \quad (5.79)$$

Los generadores del álgebra son $T_i = \frac{1}{2}\lambda_i$ (análogo a $J_i = \frac{1}{2}\sigma_i$ en $SU(2)$) de modo que las relaciones de conmutación son

$$[T_i, T_j] = if_{ijk}T_k, \quad (5.80)$$

y las constantes de estructura f_{ijk} con completamente antisimétricas en esta base. Esto es general para grupos compactos. La antisimetría en ij es inmediata y falta verificar $f_{ijk} = f_{jki}$:

$$4if_{ijk} = \text{tr}([\lambda_i, \lambda_j]\lambda_k) = \text{tr}(\lambda_i[\lambda_j, \lambda_k]) = 4if_{jki}. \quad (5.81)$$

El procedimiento de construcción de las matrices se extiende fácilmente para otros grupos $SU(n)$.

6. Método Monte Carlo

6.1. Introducción

En términos generales se denomina método Monte Carlo (MC) a todo tratamiento que involucra magnitudes aleatorias para resolver un problema. El problema en sí puede tener naturaleza aleatoria o no.

6.1.1. Ejemplo de cálculo Monte Carlo

Por ejemplo, supongamos que en \mathbb{R}^d tenemos una región Ω de bordes bien definidos de la cual queremos calcular el volumen,⁵²

$$V_{\Omega} = \int_{\Omega} d^d x = \int d^d x \Theta(\mathbf{x} \in \Omega). \quad (6.1)$$

Para ello podemos proceder a calcular los límites de integración, que definen el borde de Ω . Sin embargo, supongamos que Ω no es una región fácil de describir analíticamente. Por ejemplo, podría ocurrir que todo lo que sepamos es que la región Ω cabe en el hipercubo $[0, L]^d$,

$$V_{\Omega} = \int_{[0, L]^d} d^d x \Theta(\mathbf{x} \in \Omega), \quad (6.2)$$

y se nos proporciona un programa tal que al introducir las coordenadas de $\mathbf{x} \in [0, L]^d$ nos dé 0 ó 1 según que \mathbf{x} esté o no en Ω . Un método es dividir $[0, L]^d$ en cubitos pequeños, de lado $a \ll L$, y evaluar $\Theta(\mathbf{x} \in \Omega)$ en el centro de cada cubo. Si Ω es suficientemente bien comportada de modo que la función característica sea integrable Riemann, na^d nos estima V_{Ω} , siendo n el número de puntos dentro de Ω , y $na^d \rightarrow V_{\Omega}$ cuando $a \rightarrow 0$.

Una versión MC del mismo cálculo se basa en la observación de que si se hace el experimento de tirar (o generar) un punto al azar (pero uniformemente distribuido en $[0, L]^d$), la probabilidad de que el punto caiga en Ω es $p = V_{\Omega}/V$, siendo $V = L^d$ el volumen total del hipercubo. Por la definición usual (frecuencial) de probabilidad como el número de casos favorables sobre el número de casos

⁵²Por analogía con la función escalón de Heaviside, $\Theta(x) = 1$ si $x > 0$ y 0 si $x \leq 0$, usamos la notación $\Theta(A)$ para indicar la función que vale 1 si la proposición A es cierta y 0 si es falsa. Así, en particular $\Theta(x) = \Theta(x > 0)$. La función $\Theta(\mathbf{x} \in \Omega) = 0, 1$ se denomina función característica de Ω .

posibles, p se puede medir lanzando N puntos, $\mathbf{x}_1, \dots, \mathbf{x}_N$, con $N \rightarrow \infty$ de modo que⁵³

$$p = \lim_{N \rightarrow \infty} \frac{n}{N}, \quad (6.3)$$

siendo n el número de casos favorables, es decir, el número de casos i tal que $\mathbf{x}_i \in \Omega$. Una vez obtenido p , el volumen se obtiene mediante $V_\Omega = pV$, usando el valor conocido de V .

6.1.2. Monte Carlo y valores esperados

Lo que ha hecho el MC en este ejemplo es estimar (y en el límite $N \rightarrow \infty$, calcular) una probabilidad: p es la probabilidad de que el punto \mathbf{x} lanzado caiga o no dentro de Ω . Una probabilidad se puede ver como un caso particular de un valor esperado (el *valor esperado* o *esperanza matemática* de una variable aleatoria X suele denotarse $\langle X \rangle$ o también $E(X)$) y en general, estimar valores esperados es todo lo que puede hacer el método MC. *Para aplicar el método MC a cualquier problema hay que empezar por reducir el problema al cálculo de un valor esperado.* Para ver que la probabilidad es un cierto valor esperado de algo, basta definir la variable aleatoria θ como $\Theta(\mathbf{x} \in \Omega)$, es decir, $\theta = 1$ si al tirar el punto al azar uniformemente en $[0, L]^d$ éste cae en Ω , y $\theta = 0$ si cae fuera. Que θ sea una variable aleatoria simplemente quiere decir que es una magnitud que puede tomar valores distintos (o no) cada vez que se realiza el mismo experimento aleatorio. En nuestro caso el experimento es tirar un punto \mathbf{x} al azar uniformemente en $[0, L]^d$. Con esta definición $p = \langle \theta \rangle$. Igualmente se podría definir otro experimento aleatorio, el que consiste en tirar N puntos, todos ellos de forma independiente unos de otros, y en este caso n (el número de puntos que cae dentro, de esos N) sería una variable aleatoria, de modo que $p = \langle \frac{n}{N} \rangle$.

6.1.3. Fluctuación en estimaciones Monte Carlo

Consideremos el experimento de lanzar N puntos, y definamos la variable aleatoria $\theta_N = \frac{n}{N}$. Puesto que $p = \langle \theta_N \rangle$, cualquiera que sea el valor de N , incluido $N = 1$, podría preguntarse por qué tomamos N lo mayor posible en la práctica. El motivo es que aunque las variables aleatorias θ_1 y θ_N tienen el mismo valor esperado, su *dispersión* sí depende de N , siendo menor cuanto mayor sea N (excepto en los dos casos triviales $V_\Omega = 0$ y $V_\Omega = V$, correspondientes a $p = 0$ y $p = 1$ respectivamente). Para ver esto,⁵⁴ notemos que la probabilidad de un cierto resultado concreto n

⁵³Nótese que estrictamente hablando éste es un límite impropio ya que la sucesión $a_N = n/N$ no está bien definida y no puede usarse para un tratamiento riguroso del tema.

⁵⁴El mismo resultado se obtiene más fácilmente usando las propiedades de la varianza de una suma de variables aleatorias independientes. Aquí lo vemos dando un rodeo supuestamente pedagógico.

(para N dado) viene dado por la *distribución binomial*

$$p_n = \binom{N}{n} p^n (1-p)^{N-n}, \quad \sum_{n=0}^N p_n = 1. \quad (6.4)$$

Aquí p es la probabilidad de que un punto caiga dentro de Ω y p_n la probabilidad de que en total caigan n puntos dentro al tirar N . El factor $\binom{N}{n}$ es el número de modos de elegir n tiradas distintas de entre N , y el factor $p^n (1-p)^{N-n}$ es la probabilidad de que los puntos caigan dentro de Ω exactamente en las n tiradas seleccionadas y sólo en éstas. (Aquí se ha usado que las tiradas son independientes unas de otras.) No es difícil calcular la media y desviación estándar de esta distribución. Un método conveniente es usar la **función generatriz**:

$$g(x) \equiv \sum_n p_n x^n, \quad \langle n^k \rangle = \sum_n p_n n^k = \left(x \frac{d}{dx} \right)^k g(x) \Big|_{x=1}. \quad (6.5)$$

Estas fórmulas son válidas en general. En nuestro caso

$$g(x) = (xp + 1 - p)^N, \quad \langle n \rangle = pN, \quad \langle n^2 \rangle = pN + p^2 N(N-1). \quad (6.6)$$

Por tanto, para la dispersión

$$\sigma_N = \sqrt{\langle n^2 \rangle - \langle n \rangle^2} = \sqrt{Np(1-p)}. \quad (6.7)$$

La variable aleatoria n está centrada en pN con una dispersión que escala como \sqrt{N} . De hecho, si se utiliza la fórmula de Stirling

$$\log x! = x(\log x - 1) + \log \sqrt{2\pi x} + O\left(\frac{1}{x}\right), \quad (6.8)$$

se obtiene la forma asintótica de p_n para N grande manteniendo $(n - pN)/\sqrt{N}$ fijo,⁵⁵

$$p_n = \frac{1}{\sqrt{2\pi\sigma_N}} e^{-(n-pN)^2/2\sigma_N^2} \left(1 + O(N^{-1/2}) \right). \quad (6.9)$$

Es decir, para N grande se obtiene una distribución normal, esta es una ilustración del teorema del límite central.

⁵⁵Para aplicar la fórmula de Stirling lo mejor es trabajar con $\log p_n$, usar que $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ y cambiar de variable n a $x = (n - pN)/\sqrt{N}$.

La consecuencia es que al medir n se obtendrá un valor aleatorio alrededor de pN con una fluctuación de orden $\sqrt{p(1-p)N}$, lo cual suele indicarse así:

$$n = pN \pm \sqrt{p(1-p)N}, \quad (6.10)$$

y para $\theta_N \equiv n/N$

$$\theta_N = p \pm \sqrt{\frac{p(1-p)}{N}}. \quad (6.11)$$

Se deduce que conviene tomar N grande, para que la dispersión sea pequeña y el valor de θ_N que se obtenga al hacer el experimento tenga más probabilidad de parecerse al valor esperado.

Hay que notar que el valor esperado (igual que la probabilidad) es un concepto intuitivo que realmente nunca llega a materializarse.⁵⁶ En un experimento aleatorio, por complicado o extenso que sea, lo único que se obtiene siempre es una *muestra* de una cierta variable aleatoria; en nuestro caso, el vector $\mathbf{X} = (x_1, \dots, x_N)$. La teoría de la probabilidad sólo hace afirmaciones sobre probabilidades y valores esperados de estas variables aleatorias. Por ejemplo, por muy grande que sea N , la teoría no dice que θ_N vaya a acabar coincidiendo con su esperanza matemática, p , o incluso no se puede garantizar que $|\theta_N - p| < \varepsilon$ (para cualquier $\varepsilon > 0$ dado) sólo por aumentar N . Aunque improbable, podrían generarse todos los puntos fuera de Ω por casualidad, o todos dentro, etc. Lo único que puede afirmarse es que la *probabilidad* de que al hacer el experimento salga $|\theta_N - p| > \varepsilon$ puede hacerse arbitrariamente pequeña aumentando N . Si el experimento se repite K veces, puede calcularse la *media de la muestra* de las θ_N así obtenidas, $\overline{\theta_N} = \frac{1}{K} \sum_{j=1}^K \theta_{N,j}$, y de nuevo $\langle \overline{\theta_N} \rangle = p$ pero la situación de fondo es la misma: por muy grande que sea K todo lo que se obtiene es una muestra de otra variable aleatoria, a saber, $\mathbf{X} = (x_1, \dots, x_{KN})$, y $\overline{\theta_N}$ es también una variable aleatoria, ya que puede cambiar de valor cada vez que se hace el KN -experimento. Lo que sí se consigue es reducir la dispersión, que pasa a ser $\sqrt{p(1-p)/(KN)}$, y por tanto disminuir la probabilidad de que $\overline{\theta_N}$ esté lejos del valor esperado de la distribución.

En MC, después de reducir el problema a valores esperados de ciertas variables aleatorias, se construye una muestra o realización de dichas variables mediante un experimento aleatorio, y ello nos proporciona una estimación de los valores esperados buscados. Por lo tanto un tema central en MC es el de *reducir lo más posible la varianza* (el cuadrado de la dispersión), para aumentar la probabilidad de que la estimación se parezca al valor esperado.

En el ejemplo anterior queríamos calcular el volumen de Ω , que se puede estimar por $\hat{V}_\Omega = V \theta_N$

⁵⁶Aunque intuitivo, el valor esperado está sujeto a condiciones básicas, tales como, ser lineal, $\langle 1 \rangle = 1$ y $\langle X \rangle \geq 0$ si la variable $X \geq 0$.

ya que $\langle \hat{V}_\Omega \rangle = V_\Omega$. Teniendo en cuenta la dispersión

$$\hat{V}_\Omega = V \left(p \pm \sqrt{\frac{p(1-p)}{N}} \right) = V_\Omega \pm \sqrt{\frac{V_\Omega(V - V_\Omega)}{N}}. \quad (6.12)$$

(Puesto que no conocemos V_Ω no podemos calcular la dispersión exacta, pero se puede estimar a su vez usando \hat{V}_Ω como aproximación a V_Ω .) El error relativo en la estimación de V_Ω , es por tanto

$$\sqrt{\frac{V - V_\Omega}{NV_\Omega}}. \quad (6.13)$$

Se deduce que, aparte de tomar N grande, conviene que V sea lo más próximo a V_Ω posible, es decir, debemos tomar la caja (la región de \mathbb{R}^d donde tiramos los puntos) lo más ajustada posible a Ω , si tenemos esa opción. De ningún modo conviene tomar una caja mucho más grande de lo necesario: si Ω ocupa sólo una pequeña parte del volumen total, que un punto caiga o no dentro de Ω será muy azaroso (puede variar mucho de un experimento a otro) lo cual aumenta la dispersión en los resultados. Eso habrá que compensarlo aumentando el valor de N . Concretamente, para V grande queremos mantener constante N/V que no es más que la densidad de puntos, de ese modo mantendremos constante la cantidad de puntos que caigan cerca de la zona de interés Ω .⁵⁷

También debe notarse que con MC no se puede calcular el volumen de Ω directamente, sino que es necesario reducirlo primero a un promedio de algo, y esto es lo que requiere introducir una caja (en nuestro caso el hipercubo $[0, L]^d$): MC calcula el volumen de Ω comparado con el volumen total de la caja, a partir de la proporción de puntos que caen dentro. El valor del volumen total de la caja hay que proporcionarlo aparte (no lo da MC). Tomar V lo menor posible permite aumentar la señal frente al ruido, es decir, aumentar el número de puntos que caen en Ω .

6.1.4. Estimación Monte Carlo de integrales

Muy frecuentemente la forma de poder aplicar MC a un problema es reducirlo a integrales, y de ahí a promedios. Podemos generalizar el problema anterior. En vez de Ω , tenemos una función $f(\mathbf{x})$ definida en $[0, L]^d$, y queremos calcular

$$I_f = \int_{[0, L]^d} d^d x f(\mathbf{x}). \quad (6.14)$$

⁵⁷Si V es grande, p será pequeño y el error en su estimación, $\sqrt{p(1-p)/N}$, también lo será, concretamente $O(V^{-1/2})$. Sin embargo, al multiplicar por V para obtener V_Ω , el error pasa a $O(V^{1/2})$ y se magnifica a medida que V aumenta.

De nuevo en este caso lo que se usa es un promedio

$$\langle f(\mathbf{x}) \rangle = \frac{I_f}{V} = \frac{\int_{[0,L]^d} d^d x f(\mathbf{x})}{\int_{[0,L]^d} d^d x 1}, \quad (6.15)$$

de modo que con MC se estima $\langle f(\mathbf{x}) \rangle$, y luego $I_f = \langle f(\mathbf{x}) \rangle V$. Igual que antes, el valor de $\langle f(\mathbf{x}) \rangle$ se puede estimar lanzando N puntos $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ aleatorios independientes y uniformes en $[0, L]^d$, y tomando el promedio

$$\bar{f}_N = \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i). \quad (6.16)$$

Como se verá

$$\bar{f}_N = \langle f(\mathbf{x}) \rangle \pm \frac{\sigma_f}{\sqrt{N}}, \quad \sigma_f^2 = \langle f^2 \rangle - \langle f \rangle^2, \quad (6.17)$$

de modo que, *en probabilidad*, $\bar{f}_N \rightarrow \langle f(\mathbf{x}) \rangle$ cuando $N \rightarrow \infty$. Si aplicamos esta fórmula al caso $f(\mathbf{x}) = \Theta(\mathbf{x} \in \Omega)$, se tiene $\langle f^2 \rangle = \langle f \rangle = V_\Omega/V = p$, es decir, $\sigma_f^2 = p(1-p)$, como antes.

Un rasgo prácticamente universal de MC es que el error disminuye como $1/\sqrt{N}$. Dado que el esfuerzo de cálculo (tiempo de computación) suele crecer como N (al tratarse de tiradas independientes), se tiene que el ritmo de convergencia de un cálculo MC es relativamente lento.

Por ejemplo, si se calcula

$$I_f = \int_a^b dx f(x) \quad (6.18)$$

con N puntos no aleatorios, sino equidistantes, $x_i = a + ih$, $i = 0, \dots, N-1$, y $h = (b-a)/(N-1)$, el error es como mucho $O(h) = O(N^{-1})$. El método trapezoidal ya da $O(h^2) = O(N^{-2})$ si $f''(x)$ existe y es continua. Si se utiliza un método un poco más eficiente, por ejemplo Simpson, el error pasa a ser $O(h^4) = O(N^{-4})$, que es considerablemente más rápido que $O(N^{-1/2})$ de MC. (Por ejemplo, multiplicando por 10 el número de puntos en Monte Carlo el error se divide por 3, en Simpson se divide por 10000.) Claramente, para calcular $\int_a^b dx f(x)$ no es práctico usar MC.

Sin embargo, la utilidad de MC aparece cuando el número de dimensiones crece. En efecto, en d dimensiones, si ponemos K puntos Simpson en cada dirección, el error relativo será $O(h^4) = O(K^{-4})$, en cada una de las d integrales, y la suma de errores relativos para la integral completa, también $O(K^{-4})$. El número total de puntos requeridos (evaluaciones de $f(\mathbf{x})$ requeridas) es $N = K^d$, por tanto, el error escala como $O(N^{-4/d})$. En cambio, en MC el error sigue escalando como $O(N^{-1/2})$, independientemente del problema. Por este motivo MC empieza a ser menos ineficiente que los otros

métodos cuando d crece. Para Simpson, MC empieza a ser preferible cuando $d > 8$. Por ejemplo,⁵⁸ si $d = 20$, el número de puntos usado en Simpson con $K = 10$ sería $N = K^d = 10^{20}$ para una precisión $\varepsilon_S = 1/10^4$, mientras que MC con el mismo esfuerzo daría una precisión $\varepsilon_{MC} = 1/10^{10}$. En realidad 10^{20} es una cantidad de puntos enorme que está más allá de las posibilidades prácticas de computación. En la práctica se usaría MC con menos puntos y menos precisión, pero mucho mayor que Simpson con los mismos puntos.

Como regla general, para problemas con muchas variables, los demás métodos se vuelven inviables y hay que recurrir a MC. Nótese que de nada sirve usar métodos de cuadraturas más sofisticados (por ejemplo Gauss) ya que a efectos prácticos $O(N^{-100/d})$ es lo mismo que $O(1)$ (el error no disminuye) si d es realmente grande.⁵⁹ Por ejemplo, para simular 1000 partículas de una gas (evidentemente poco parecido a un gas realista, con 10^{23} partículas) $d = 6000$. O para estudiar un plasma de gluones (cada gluon tiene cuatro polarizaciones y 8 colores) en una red 16^4 hay que hacer una integral sobre $d = 8 \times 4 \times 16^4 = 2^{21} = 2.1 \times 10^6$ dimensiones.

Por supuesto, que MC sea preferible cuando hay muchas variables es sólo la regla general. En problemas concretos puede haber un método particular no MC que sea mejor. Un caso obvio es el de integración en muchas dimensiones, pero de una función separable de modo que la integral equivale a d integrales unidimensionales. Lo más eficiente será probablemente integrar cada dimensión por separado ya que en este caso $N = dK$ y no K^d .

En resumen, MC no es un método diseñado para obtener resultados muy precisos, pero es capaz de dar buenas estimaciones fiables y con gran facilidad en casos en los que otros métodos fallan.

Aparte, tampoco se debe abusar de MC: generalmente será más eficiente calcular de modo analítico todo lo que se pueda hacer así que dejarlo a MC. Por ejemplo, supongamos que $0 \leq f(\mathbf{x}) \leq K$, $\mathbf{x} \in [0, L]^d$, y queremos calcular $I = \int_0^L d^d x f(\mathbf{x})$ con MC.

Método *a*): $\bar{f} = \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i)$, con \mathbf{x}_i independientes y uniformemente distribuidos en $[0, L]^d$. Puesto que $I = V \langle f \rangle$ ($V = L^d$), se obtiene una estimación con

$$\bar{I}_a = V \bar{f} = I \pm V \frac{\sigma_f}{\sqrt{N}}. \quad (6.19)$$

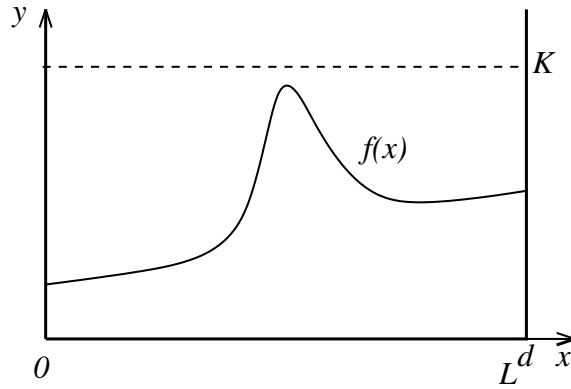


Figura 9: Función a integrar en $[0, L]^d$ acotada entre 0 y K .

Método *b*): En \mathbb{R}^{d+1} tenemos la caja $[0, L]^d \times [0, K]$, que contiene la región $\Omega = \{(\mathbf{x}, y) | f(\mathbf{x}) < y\}$. El volumen de Ω es la integral pedida I , y podemos aplicar el método visto anteriormente: si lanzamos N puntos y n cumplen $y < f(\mathbf{x})$, KVn/N es una estimación de I (ya que ahora KV es el volumen total):

$$\bar{I}_b = KV \frac{n}{N} = I \pm \sqrt{\frac{I(KV - I)}{N}}. \quad (6.20)$$

Comparando las varianzas de los dos métodos, se tiene (usando $I = V\langle f \rangle$)

$$\begin{aligned} \frac{N}{V^2}(\sigma_{\bar{I}_b}^2 - \sigma_{\bar{I}_a}^2) &= (K\langle f \rangle - \langle f \rangle^2) - (\langle f^2 \rangle - \langle f \rangle^2) \\ &= \langle (K - f)f \rangle \geq 0. \end{aligned} \quad (6.21)$$

El método *a* siempre es más eficiente (tiene menor dispersión). De hecho la dispersión de \bar{I}_b puede ser muy grande si K es mucho mayor que el valor típico de f (esto puede ser inevitable, por ejemplo, si f tiene un pico alto y estrecho). El motivo es que en el método *b* se está haciendo una integral extra, en $[0, K]$, mediante MC, para calcular el tamaño del intervalo $[0, f(\mathbf{x})]$, mientras que en el método *a* esa integral se calcula exactamente. Como regla, se reduce la varianza haciendo las integrales analíticas directamente, si es posible, evitando usar MC ahí.

⁵⁸Ésta es una estimación muy burda ya que estas tendencias deben completarse con coeficientes cuyo valor puede introducir una modificación sustancial en la estimación.

⁵⁹No sólo eso, a menos que el problema sea enormemente regular, será inviable calcular los límites de integración anidados de cada variable. Si simplemente se pone a cero la función fuera de su intervalo, dejará de ser suave, perdiéndose las altas precisiones teóricas del método de integración usado.

Otra cosa a tener muy en cuenta es que aunque MC se base en experimentos aleatorios eso no quiere decir que se pueda proceder de modo arbitrario. Elegir los puntos de un modo ad hoc “más o menos aleatorio” inventado por nosotros, producirá resultados completamente incorrectos, especialmente en problemas multidimensionales, en los que, como veremos, la región donde el integrando es relevante puede ser extremadamente pequeña.

Que la arbitrariedad es inaceptable, incluso en casos simples se puede ver en el siguiente ejemplo: Queremos saber cuál es el valor promedio del área de “un cuadrado elegido al azar” de entre los de lado $0 \leq \ell \leq L$.

Método a): El valor del lado está entre 0 y L , así que promediamos sobre ℓ , teniendo en cuenta que el área es $A = \ell^2$,

$$\langle A \rangle_a = \frac{\int_0^L d\ell \ell^2}{\int_0^L d\ell} = \frac{1}{3}L^2. \quad (6.22)$$

Método b): El área está entre 0 y L^2 , por tanto promediamos directamente sobre A

$$\langle A \rangle_b = \frac{\int_0^{L^2} dA A}{\int_0^{L^2} dA} = \frac{1}{2}L^2. \quad (6.23)$$

Se obtienen valores distintos. El primer resultado supone lados equiprobables, mientras que el segundo supone áreas equiprobables, y ambas situaciones no son equivalentes. En cada caso concreto habrá que usar el tratamiento correcto, yendo al origen del problema que se estudia. Tal y como está, “un cuadrado elegido al azar” no es una afirmación suficientemente bien definida para una respuesta precisa. En este caso ambas prescripciones producen una estimación similar (difieren en un factor del orden de la unidad) pero no sería así si en lugar de un cuadrado se tratara de un hipercono en 10^6 dimensiones.

Otro ejemplo, si se quiere calcular el tamaño de una región Ω contenida dentro de una caja y tiramos más puntos en unas zonas que en otras, estaremos sobrestimando V_Ω si Ω está en la zona más visitada y subestimándolo si está en la zona menos visitada. Se está introduciendo un sesgo en la estimación. Supongamos, para concretar, que tenemos el problema de calcular el área de una región $\Omega \subset \mathbb{R}^2$ y en lugar de enmarcarla en una caja rectangular la enmarcamos dentro de una región circular D de radio R . Si tiramos puntos uniformemente en D obtendremos la estimación correcta del área. Sin embargo nos inventamos la siguiente prescripción, primero se genera el radio r uniformemente en $[0, R]$, y a continuación se genera el ángulo ϕ uniformemente en $[0, 2\pi]$, y como antes, se estima V_Ω mediante $\theta_N = n/N$. Esta prescripción introduce un sesgo, ya que pone tantos puntos en $0 < r < R/2$ como en $R/2 < r < R$, cuando en realidad el área de la segunda región es

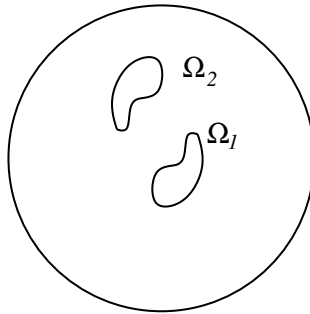


Figura 10: Regiones Ω_1 y Ω_2 en una caja circular. Para que la estimación de las áreas sea correcta los puntos deben generarse de modo que r^2 , no r , tenga una distribución uniforme.

tres veces mayor. Si tenemos dos regiones Ω_1 y Ω_2 , la primera situada más cerca del centro y la otra más lejos, la prescripción sobrestimaré Ω_1 y subestimaré Ω_2 .

6.2. Probabilidad. Variables aleatorias.

6.2.1. Probabilidades

Un experimento aleatorio puede dar lugar a uno cualquiera de los **resultados** posibles, x . El conjunto de resultados es el **espacio muestral** Ω . Los subconjuntos $A \subseteq \Omega$ se denominan **sucesos**.⁶⁰ Una **probabilidad** P definida sobre Ω es una medida que sea positiva y normalizada:

$$P(A) \geq 0, \quad P(\emptyset) = 0, \quad P(A \cup B) = P(A) + P(B) \text{ si } A \cap B = \emptyset, \quad P(\Omega) = 1. \quad (6.24)$$

$P(A)$ quiere decir $P(x \in A)$, es decir, la probabilidad de que al hacer el experimento, el resultado x que ocurra esté en A . (Es equivalente hablar de probabilidades de subconjuntos o probabilidades de proposiciones). $P(x)$ denota $P(\{x\})$ ($A = \{x\}$ en ese caso). Que P es la probabilidad definida sobre Ω , el conjunto de resultados x , se suele indicar con $x \sim P$: “ x sigue o está distribuido según la probabilidad P ”.

Dos sucesos A y B son **incompatibles** si son disjuntos. Otro concepto importante es el de **probabilidad condicionada**, $P(B|A)$ es la probabilidad de que el resultado $x \in B$ cuando $x \in A$

⁶⁰Para simplificar suponemos que todos los subconjuntos son admisibles.

(suponemos que $P(A) \neq 0$), y se puede expresar como

$$P(B|A) = \frac{P(AB)}{P(A)} \quad (6.25)$$

(donde $P(AB) \equiv P(A \cap B)$).⁶¹ También se tiene $P(AB) = P(B|A)P(A) = P(A|B)P(B)$, que relaciona $P(A|B)$ con $P(B|A)$ (Teorema de Bayes). Dos sucesos A y B son **independientes** cuando $P(AB) = P(A)P(B)$, equivalentemente $P(A|B) = P(A)$ o $P(B|A) = P(B)$.

Si A_i , $i = 1, \dots, n$ es una partición de Ω , es decir $\cup_{i=1}^n A_i = \Omega$ y $A_i \cap A_j = \emptyset$ si $i \neq j$, entonces BA_i es una partición de B , y

$$P(B) = \sum_{i=1}^n P(BA_i) = \sum_{i=1}^n P(B|A_i)P(A_i), \quad (6.26)$$

es decir, la probabilidad de un suceso se puede obtener si se conocen su probabilidad condicionada a un conjunto de alternativas, y las probabilidades de éstas.

Si Ω es un conjunto discreto (finito o infinito numerable), $\Omega = \{x_i, i = 1, 2, \dots\}$, y cada x_i tiene una probabilidad que denotamos p_i , con $0 \leq p_i \leq 1$, $\sum_i p_i = 1$. La función $p : i \in \{1, 2, \dots\} \rightarrow p_i \in [0, 1]$, es la **función de distribución de la probabilidad**, y se indica $x \sim p$.

Si Ω es continuo, un subconjunto de \mathbb{R}^d ó una variedad d -dimensional, podemos tomar un sistema de coordenadas,⁶² x , y considerar una partición de Ω formada por elementos infinitesimales de volumen $(x, d^d x)$, cada uno con una probabilidad infinitesimal $d^d x p(x)$, $p(x) \geq 0$, de modo que

$$\int_{\Omega} d^d x p(x) = 1, \quad P(A) = \int_A d^d x p(x) = \int_{\Omega} d^d x p(x) \Theta(x \in A). \quad (6.27)$$

La función $p(x)$ se denomina **densidad de probabilidad**. Con cierto abuso de lenguaje se puede indicar $x \sim p(x)$. Sin embargo es importante tener en cuenta que la densidad de probabilidad $p(x)$ es una función que cambia al cambiar de sistema de coordenadas usado para describir Ω . En efecto, si x' es otro sistema de coordenadas

$$P(A) = \int_A d^d x p(x) = \int_A d^d x' p'(x') = \int_A d^d x \left| \frac{\partial x'}{\partial x} \right| p'(x'). \quad (6.28)$$

Como esta relación vale para A arbitrario, se tiene

$$p(x) = \left| \frac{\partial x'}{\partial x} \right| p'(x'(x)), \quad \left| \frac{\partial x'}{\partial x} \right| \equiv \left| \det \left(\frac{\partial x'^i(x)}{\partial x^j} \right) \right|. \quad (6.29)$$

⁶¹En realidad todas las probabilidades son condicionadas; $P(A)$ es realmente $P(A|\Omega)$, siempre puede suponerse que hay un espacio muestral mayor respecto del cual Ω es un subconjunto.

⁶²A menudo no usaremos negrita para enfatizar que x son d coordenadas.

Por tanto, aunque $P(A)$ y $p(x)d^d x$ (la probabilidad de un elemento de volumen infinitesimal) no dependen del sistema de coordenadas, $d^d x$ y $p(x)$ por separado sí dependen. Por ejemplo, si x tiene dimensiones de longitud, L , $p(x)d^d x$ es adimensional pero $p(x)$ tiene dimensiones L^{-d} (y si se cambia de unidades, cambia su valor).⁶³

En el caso unidimensional

$$P(X < a) = \int_{-\infty}^a dx p(x), \quad \frac{dP(X < a)}{da} = p(a) \geq 0. \quad (6.30)$$

Igualmente $p(a) = -\frac{dP(X > a)}{da}$.

6.2.2. Deltas de Dirac

La delta de Dirac en d dimensiones, $\delta(\mathbf{x})$,⁶⁴ es una distribución o función generalizada cuya propiedad definitoria es

$$\int d^d x \delta(\mathbf{x} - \mathbf{y}) f(\mathbf{x}) = f(\mathbf{y}), \quad (6.31)$$

siendo $f(\mathbf{x})$ cualquier función ordinaria suficientemente bien comportada. Una consecuencia inmediata es $\delta(\mathbf{x})f(\mathbf{x}) = \delta(\mathbf{x})f(\mathbf{0})$. Y también

$$\int_A d^d x \delta(\mathbf{x} - \mathbf{y}) = \Theta(\mathbf{y} \in A). \quad (6.32)$$

La delta d -dimensional no es más que el producto de las d deltas unidimensionales

$$\delta(\mathbf{x} - \mathbf{y}) = \prod_{i=1}^d \delta(x^i - y^i). \quad (6.33)$$

Intuitivamente $\delta(x)$ es $+\infty$ en $x = 0$ y cero si $x \neq 0$, pero no toda familia de funciones $f(x, \lambda)$ tal que $\int dx f(x, \lambda) = 1$ y $f(x, \lambda) \xrightarrow{\lambda \rightarrow 0} 0$ ($x \neq 0$) tiende a $\delta(x)$. Por ejemplo,

$$f(x, \lambda) = \delta(x) + \frac{1}{2\lambda} (\delta(x + \lambda) - \delta(x - \lambda)) \rightarrow \delta(x) + \delta'(x) \neq \delta(x). \quad (6.34)$$

⁶³Las magnitudes que no cambian de valor al cambiar de coordenadas, $F(x) = F'(x'(x))$, se denominan *escalares*, las que cambian con el jacobiano, *densidades*. Para variables continuas, la moda (valor más probable) depende de la variable que se use.

⁶⁴A veces se escribe $\delta^{(d)}(x)$ para indicar que es la delta en d dimensiones.

(La distribución $\delta'(x)$ se define integrando por partes, $\int dx \delta'(x) f(x) = -f'(0)$.) Una forma de obtener $\delta(x)$ es como el límite $h(x/a)/a$ para $a \rightarrow 0^+$, siendo $h(x)$ una función continua con integral 1.

Tampoco es cierto que $\delta(\lambda x) = \delta(x)$, de hecho

$$\delta(\lambda x) = \frac{1}{|\lambda|} \delta(x) \quad (\lambda \neq 0). \quad (6.35)$$

En efecto,

$$\int dx \delta(\lambda x) f(x) = \frac{1}{|\lambda|} \int d(|\lambda|x) \delta(\lambda x) f(x) = \frac{1}{|\lambda|} \int dy \delta(y) f(y/\lambda) = \frac{1}{|\lambda|} f(0). \quad (6.36)$$

Una fórmula útil, que se demuestra de la misma forma, es la siguiente: si $f(x)$ tiene ceros simples en $\{x_i\}_{i=1}^n$, y es derivable ahí,

$$\delta(f(x)) = \sum_{i=1}^n \frac{\delta(x-x_i)}{|f'(x_i)|}. \quad (6.37)$$

Estas expresiones indican que $\delta(x)$ es una *densidad* bajo cambios de coordenadas. En efecto, si $\mathbf{y}(\mathbf{x})$ es invertible y se anula en $\mathbf{x} = \mathbf{0}$

$$\delta(\mathbf{y}(\mathbf{x})) = \left| \det \left(\frac{\partial x^i}{\partial y^j} \right) \right| \delta(\mathbf{x}). \quad (6.38)$$

Equivalentemente, $d^d \mathbf{y} \delta(\mathbf{y}) = d^d \mathbf{x} \delta(\mathbf{x})$. Si \mathbf{x} tiene dimensiones, digamos L , $\delta(\mathbf{x})$ tiene dimensiones L^{-d} .

El caso discreto $\text{Prob}(x_i) = p_i$ se podría incluir en el continuo asignando una densidad de probabilidad

$$p(x) = \sum_{i \in I} p_i \delta(x - x_i), \quad (6.39)$$

donde $\delta(x - x_i)$ es la delta de Dirac d -dimensional centrada en x_i , ya que esta densidad reproduce correctamente

$$P(A) = \sum_{x_i \in A} p_i. \quad (6.40)$$

6.2.3. Variables aleatorias

Por conveniencia, en lo que sigue vamos a usar \mathbf{r} (en vez de \mathbf{x}) para denotar a los resultados. En el caso continuo y en un cierto sistema de coordenadas, \mathbf{r} especifica el resultado y $p(\mathbf{r})$ es la densidad de probabilidad.

Una **variable aleatoria** real es cualquier función X real definida sobre el conjunto de resultados Ω ,

$$X : \Omega \rightarrow \mathbb{R}, \quad r \in \Omega \mapsto X(r) \in \mathbb{R}. \quad (6.41)$$

Al realizar el experimento aleatorio se obtiene un cierto resultado r y la variable aleatoria toma el valor $X(r)$ en ese caso. Por ejemplo, el experimento puede ser tirar dos dados y X el valor de la suma de los puntos de las dos caras. Igualmente se pueden definir variables aleatorias que tomen valores en \mathbb{R}^n , \mathbb{C} , operadores, etc. Generalmente interesa que se trate de un espacio vectorial sobre los reales, para definir valores esperados de la variable. Si no se dice otra cosa se supondrá que las variables son reales o en \mathbb{R}^n . Nótese que los propios valores de las coordenadas r definen una variable aleatoria, que podemos denotar R .

El **valor esperado** o **esperanza matemática** de X se denota $\langle X \rangle$ ó $E(X)$. Si $\Omega = \{r_1, r_2, \dots\}$ es discreto,

$$\langle X \rangle = \sum_{i \in I} p_i X_i, \quad X_i = X(r_i). \quad (6.42)$$

Si Ω es continuo⁶⁵

$$\langle X \rangle = \int_{\Omega} d^d r p(r) X(r). \quad (6.43)$$

Si hay más de una distribución de probabilidad posible lo indicamos con un subíndice:

$$\langle X \rangle_P = \int_{\Omega} d^d r p(r) X(r). \quad (6.44)$$

Observaciones:

1) El valor esperado es una abstracción de la media aritmética de X obtenida en muchos experimentos independientes: por la ley de los grandes números $\bar{X} = \frac{1}{N} \sum_i^N X_i$ tiende a $\langle X \rangle$ con probabilidad uno. El “valor esperado” X es lo que deber esperarse para \bar{X} cuando N es grande, ciertamente no es el valor que debe esperarse para X al hacer un experimento. De hecho en general $\langle X \rangle$ no tiene por qué ser uno de los valores posibles que toma X . Por ejemplo, X puede tomar valores 0 y 1 y $\langle X \rangle$ ser 1/2.

2) El conjunto de variables aleatorias define un espacio vectorial. El valor esperado es lineal (respecto de X), no negativo si la variable aleatoria es no negativa, y el valor esperado de $X = 1$ es

⁶⁵Para no sobrecargar la notación aquí usamos Ω también para denotar el conjunto de valores de r . En realidad este conjunto depende de la elección de coordenadas.

1. Y viceversa, una aplicación $\langle \cdot \rangle$ del conjunto de variables aleatorias en \mathbb{R} , $X \mapsto \langle X \rangle$, con estas tres propiedades define unívocamente una densidad de probabilidad $p(\mathbf{r})$ (añadiendo algunas condiciones de regularidad sobre el espacio de las X admisibles).

También se define el **valor esperado condicionado** (que también se denota $E(X|A)$):

$$\langle X|A \rangle = \frac{\int_A d^d r p(\mathbf{r}) X(\mathbf{r})}{\int_A d^d r p(\mathbf{r})} = \frac{\langle X \Theta(\mathbf{R} \in A) \rangle}{\langle \Theta(\mathbf{R} \in A) \rangle}. \quad (6.45)$$

(Aquí \mathbf{R} denota la variable aleatoria definida por las propias coordenadas \mathbf{r} .) El promedio se toma sólo sobre los casos en que $r \in A$, y se normaliza de modo que el promedio de $X = 1$ sea 1 de nuevo.

La probabilidad se puede recuperar a partir del valor esperado, ya que

$$P(A) = \int_A d^d r p(\mathbf{r}) = \int_{\Omega} d^d r p(\mathbf{r}) \Theta(r \in A) = \langle \Theta(\mathbf{R} \in A) \rangle. \quad (6.46)$$

Del mismo modo se obtiene la densidad de probabilidad:

$$p(\mathbf{r}) = \int_{\Omega} d^d r' p(\mathbf{r}') \delta(\mathbf{r}' - \mathbf{r}) = \langle \delta(\mathbf{R} - \mathbf{r}) \rangle. \quad (6.47)$$

Intuitivamente: $p(\mathbf{r})$ es la “probabilidad” de que \mathbf{R} caiga en \mathbf{r} y estos son los casos seleccionados por $\delta(\mathbf{R} - \mathbf{r})$.

Esto se puede generalizar: al hacer el experimento, una variable aleatoria X puede adoptar distintos valores. Repitiendo el experimento muchas veces se puede obtener un histograma de los valores de X y eso define una densidad de probabilidad. Equivalentemente, se puede definir la densidad de probabilidad asociada a una variable X mediante

$$p_X(x) \equiv \langle \delta(X - x) \rangle. \quad (6.48)$$

Esta densidad de probabilidad permite calcular los valores esperados de variables aleatorias que dependan del resultado sólo a través de X :

$$\begin{aligned} \langle F(X) \rangle &= \int d^d r p(\mathbf{r}) F(X(\mathbf{r})) = \int d^d r p(\mathbf{r}) \int dx F(x) \delta(X(\mathbf{r}) - x) \\ &= \int dx \langle \delta(X - x) \rangle F(x) = \int dx p_X(x) F(x). \end{aligned} \quad (6.49)$$

Si se tiene $p_X(x)$ se puede calcular $\langle F(X) \rangle$ sin necesidad de referirse a las variables subyacentes \mathbf{r} y $p(\mathbf{r})$.

También se define la **función de distribución acumulada** de un variable aleatoria X como

$$P_X(x) \equiv \text{Prob}(X < x) = \int_{-\infty}^x dx' p_X(x'), \quad p_X(x) = \frac{dP_X(x)}{dx} \geq 0. \quad (6.50)$$

La virtud de la probabilidad acumulada es que, a diferencia de la densidad de probabilidad, es un escalar bajo cambios de variable x . En el caso discreto, la probabilidad acumulada es una función escalonada

$$P_X(x) \equiv \sum_{x_i \leq x} p_i. \quad (6.51)$$

Si x_i es estrictamente creciente, $x_i < x_{i+1}$, $p_i = P_X(x_i) - P_X(x_{i-1})$.

La construcción de la densidad de probabilidad se puede generalizar para varias variables aleatorias X_i , $i = 1, \dots, n$, ó $\mathbf{X} \in \mathbb{R}^n$:

$$p_{\mathbf{X}}(\mathbf{x}) \equiv \langle \delta(\mathbf{X} - \mathbf{x}) \rangle = \left\langle \prod_{i=1}^n \delta(X_i - x_i) \right\rangle. \quad (6.52)$$

Si sólo se necesitan variables aleatorias que sean funciones de estas n variables, $p_{\mathbf{X}}(\mathbf{x})$ es la densidad de probabilidad más detallada requerida, pues permite obtener todos los valores esperados:

$$\langle F(\mathbf{X}) \rangle = \int d^n x p_{\mathbf{X}}(\mathbf{x}) F(\mathbf{x}), \quad (6.53)$$

así como la densidad de probabilidad de otra variable construida con las \mathbf{X} , $G = G(\mathbf{X})$:

$$p_G(g) = \langle \delta(G(\mathbf{X}) - g) \rangle = \int d^n x p_{\mathbf{X}}(\mathbf{x}) \delta(G(\mathbf{x}) - g). \quad (6.54)$$

De esta discusión se concluye que no hay una diferencia fundamental entre variables derivadas, \mathbf{X} , y variables subyacentes, \mathbf{R} : en realidad estas \mathbf{R} podrían ser a su vez variables derivadas de otras más detalladas \mathbf{Q} . Se puede trabajar todo el tiempo con un conjunto de variables sin necesidad de especificar en ningún momento si éstas son variables derivadas o no.

Cálculo de densidad de probabilidad de \mathbf{X}

Supongamos que los resultados $\mathbf{R} \in \mathbb{R}^d$ y $\mathbf{X} \in \mathbb{R}^n$ y que las \mathbf{X} son funcionalmente independientes, y en particular $n \leq d$. Las $\mathbf{r} = (r^1, \dots, r^d)$ forman un conjunto de coordenadas, lo que hacemos es completar las \mathbf{X} hasta formar otro sistema de coordenadas $(x^1, \dots, x^n, x^{n+1}, \dots, x^d)$. Partiendo de

$$\langle F(\mathbf{X}) \rangle = \int d^n x p_{\mathbf{X}}(\mathbf{x}) F(\mathbf{x}) = \int d^d r p(\mathbf{r}) F(\mathbf{X}(\mathbf{r})), \quad (6.55)$$

y usando el nuevo sistema de coordenadas

$$\langle F(\mathbf{X}) \rangle = \int d^d x J(\mathbf{x}) p(\mathbf{r}) F(\mathbf{x}), \quad (6.56)$$

donde ahora \mathbf{r} es función de \mathbf{x} y $J(\mathbf{x})$ es el jacobiano del cambio de variable en valor absoluto

$$J(\mathbf{x}) = \left| \det \left(\frac{\partial r^i}{\partial x^j} \right) \right|. \quad (6.57)$$

De aquí se deduce que

$$p_{\mathbf{X}}(\mathbf{x}) = \int d^{n-d} x J(\mathbf{x}) p(\mathbf{r}(\mathbf{x})), \quad (6.58)$$

donde la integral es sólo sobre las variables (x^{n+1}, \dots, x^d) . A menudo las (x^{n+1}, \dots, x^d) se eligen como un subconjunto de las \mathbf{r} .

Si se tiene la distribución de probabilidad de un conjunto de variables aleatorias \mathbf{X} , $p_{\mathbf{X}}(\mathbf{x})$, y se quiere sólo un subconjunto de ellas, por ejemplo X_1, X_2 , se puede obtener su **probabilidad marginal**

$$p_{X_1, X_2}(x_1, x_2) = \int dx_3 \dots dx_n p_{\mathbf{X}}(x_1, x_2, x_3, \dots, x_n) = \langle \delta(X_1 - x_1) \delta(X_2 - x_2) \rangle. \quad (6.59)$$

También se define la **densidad de probabilidad condicionada** de unas variables respecto de otras, por ejemplo

$$p_{X_1, X_2 | X_3}(x_1, x_2 | x_3) = \frac{\langle \delta(X_1 - x_1) \delta(X_2 - x_2) \delta(X_3 - x_3) \rangle}{\langle \delta(X_3 - x_3) \rangle} = \frac{p_{X_1, X_2, X_3}(x_1, x_2, x_3)}{p_{X_3}(x_3)}. \quad (6.60)$$

Esta probabilidad está correctamente normalizada para todo x_3

$$\int dx_1 dx_2 p_{X_1, X_2 | X_3}(x_1, x_2 | x_3) = 1. \quad (6.61)$$

Dos conjuntos de variables, $\mathbf{X} = (X_1, \dots, X_n)$ e $\mathbf{Y} = (Y_1, \dots, Y_m)$ se dicen que son **independientes** (unas de otras) cuando

$$p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x}) p_{\mathbf{Y}}(\mathbf{y}). \quad (6.62)$$

(Y análogamente para tres o más conjuntos de variables.) En este caso, para la probabilidad condicionada

$$p_{\mathbf{X} | \mathbf{Y}}(\mathbf{x} | \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x}), \quad (6.63)$$

y también

$$\langle f(\mathbf{X})g(\mathbf{Y}) \rangle = \langle f(\mathbf{X}) \rangle \langle g(\mathbf{Y}) \rangle. \quad (6.64)$$

Y viceversa, si ec. (6.64) se cumple para funciones arbitrarias f y g , las variables \mathbf{X} y \mathbf{Y} son independientes. (En efecto, basta tomar $f(\mathbf{X}) = \delta(\mathbf{X} - \mathbf{x})$ y $g(\mathbf{Y}) = \delta(\mathbf{Y} - \mathbf{y})$.)

Ejemplo. Si X e Y son dos variables aleatorias con densidad de probabilidad $p_{X,Y}(x,y)$ podemos calcular la densidad de probabilidad de $Z = X + Y$

$$p_{X+Y}(z) = \int dx dy p_{X,Y}(x,y) \delta(x+y-z) = \int dx p_{X,Y}(x,z-x). \quad (6.65)$$

Si X e Y son variables independientes, $p_{X,Y}(x,y) = p_X(x)p_Y(y)$,

$$p_{X+Y}(z) = \int dx p_X(x) p_Y(z-x) \equiv (p_X * p_Y)(z) \quad (\text{convolución de } p_X \text{ e } p_Y). \quad (6.66)$$

Más generalmente, para la suma de N variables independientes, $Z = \sum_{i=1}^N X_i$,

$$p_Z(z) = (p_{X_1} * \dots * p_{X_N})(z). \quad (6.67)$$

(La convolución es asociativa y conmutativa.) Si en lugar de la suma se toma la media aritmética

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i, \quad p_{\bar{X}}(z) = N(p_{X_1} * \dots * p_{X_N})(Nz). \quad (6.68)$$

Una forma práctica de convolucionar es mediante transformada de Fourier, ya que si

$$\begin{aligned} \tilde{f}(k) &= \int_{-\infty}^{+\infty} e^{-ikx} f(x) dx, & f(x) &= \int_{-\infty}^{+\infty} e^{ikx} \tilde{f}(k) \frac{dk}{2\pi}, \\ f(x) &= (f_1 * f_2)(x) & \Leftrightarrow & \tilde{f}(k) = \tilde{f}_1(k) \tilde{f}_2(k). \end{aligned} \quad (6.69)$$

Respecto de la transformada de Fourier, también es interesante la relación

$$\tilde{p}_X(k) = \int dx e^{-ikx} p_X(x) = \langle e^{-ikX} \rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle X^n \rangle, \quad (6.70)$$

es decir, es la función generatriz de los **momentos** de la distribución $p(x)$. Análogamente en d dimensiones ($\mathbf{X} = (X^1, \dots, X^d)$) son d variables cualesquiera

$$\tilde{p}_{\mathbf{X}}(\mathbf{k}) = \langle e^{-i\mathbf{k} \cdot \mathbf{X}} \rangle = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \sum_{i_1=1}^d \dots \sum_{i_n=1}^d k_{i_1} \dots k_{i_n} \langle X^{i_1} \dots X^{i_n} \rangle. \quad (6.71)$$

La **varianza** de una variable real X se define

$$\text{Var}(X) = \sigma_X^2 = \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2. \quad (6.72)$$

La varianza es no negativa, y su raíz cuadrada positiva es la dispersión o **desviación estándar**, σ_X . Es una medida de cuánto puede fluctuar X alrededor de su valor medio. Análogamente la **covarianza** de dos variables X e Y se define⁶⁶

$$\text{Cov}(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle. \quad (6.73)$$

Una covarianza positiva indica que cuando una de las variables está por encima de su media la otra tiende a estar también por encima (ídem por debajo). Una covarianza negativa indica lo contrario: cuando una de las variables está por encima de su media la otra tiende a estar por debajo. Cuando las variables son independientes su covarianza se anula (pero no necesariamente al revés).

La varianza de la suma de variables se puede expresar como

$$\text{Var}\left(\sum_i X_i\right) = \sum_{i,j} \text{Cov}(X_i, X_j) = \sum_i \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j), \quad (6.74)$$

por tanto, cuando las variables X_i son *independientes*

$$\text{Var}\left(\sum_i X_i\right) = \sum_i \text{Var}(X_i) \quad (\text{variables independientes}). \quad (6.75)$$

6.3. Método Monte Carlo

Podemos ahora deducir la fórmula fundamental del cálculo Monte Carlo de integrales. Todo es análogo para el caso discreto.

Sea $f(\mathbf{x})$ una función real definida sobre $\Omega \subseteq \mathbb{R}^d$. Aunque el método funciona igual cuando f toma valores en un espacio vectorial cualquiera, la fórmula de la dispersión es más simple si f es real. Lo que queremos calcular es

$$I_f \equiv \int_{\Omega} d^d x p(\mathbf{x}) f(\mathbf{x}), \quad \text{donde} \quad p(\mathbf{x}) \geq 0, \quad \int_{\Omega} d^d x p(\mathbf{x}) = 1. \quad (6.76)$$

⁶⁶En cálculo numérico, las versiones basadas en $\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle$ son preferibles a las basadas en $\langle XY \rangle - \langle X \rangle \langle Y \rangle$ ya que esta forma tiende a reforzar el error de redondeo. La misma observación se aplica a la estimación numérica de la varianza.

La función no negativa normalizada p puede o no tener la naturaleza de una densidad de probabilidad definida sobre \mathbb{R}^d (u otra variedad d dimensional) pero en todo caso se puede utilizar como si lo fuera a efectos de cálculo. Con esta interpretación

$$I_f = \langle f(\mathbf{X}) \rangle_p \equiv \langle f \rangle. \quad (6.77)$$

Sean \mathbf{X}_i , $i = 1, \dots, N$, N variables aleatorias independientes idénticamente distribuidas (iid.) según p , $\mathbf{X}_i \sim p$. Esto quiere decir que su densidad de probabilidad conjunta es

$$p_{\mathbf{X}_1, \dots, \mathbf{X}_N}(\mathbf{x}_1, \dots, \mathbf{x}_N) = p(\mathbf{x}_1) \cdots p(\mathbf{x}_N). \quad (6.78)$$

El conjunto $\{\mathbf{X}_i\}_{i=1}^N$ forma la **muestra**. Se define la **media de la muestra** (otra variable aleatoria) como

$$\bar{f} \equiv \frac{1}{N} \sum_{i=1}^N f(\mathbf{X}_i) \quad (6.79)$$

que nos proporciona una estimación no sesgada de $\langle f \rangle$

$$\langle \bar{f} \rangle = \frac{1}{N} \sum_{i=1}^N \langle f(\mathbf{X}_i) \rangle = \frac{1}{N} \sum_{i=1}^N \langle f \rangle = \langle f \rangle. \quad (6.80)$$

(Obsérvese que aquí no se ha requerido que las variables sean independientes.)

La **ley de los grandes números** afirma que de hecho $\bar{f} \rightarrow \langle f \rangle$ con *probabilidad uno*, cuando $N \rightarrow \infty$:

$$\forall \varepsilon > 0 \quad \lim_{N \rightarrow \infty} \text{Prob}(|\bar{f} - \langle f \rangle| \leq \varepsilon) = 1. \quad (6.81)$$

Este teorema sólo requiere que exista $\langle f \rangle$ (es decir, que la integral I_f converja).

En cuanto a la dispersión del estimador \bar{f}

$$\text{Var}(\bar{f}) = \frac{1}{N^2} \text{Var}\left(\sum_i f(\mathbf{X}_i)\right) = \frac{1}{N^2} \sum_{i=1}^N \text{Var}(f(\mathbf{X}_i)) = \frac{1}{N} \sigma_f^2, \quad (6.82)$$

En la segunda igualdad se ha usado que las variables son independientes y en la tercera que están distribuidas según p . Aquí

$$\sigma_f^2 = \langle f^2 \rangle - \langle f \rangle^2 = \int_{\Omega} d^d x p(\mathbf{x}) (f(\mathbf{x}) - \langle f \rangle)^2. \quad (6.83)$$

Se deduce que

$$\bar{f} = \langle f \rangle \pm \frac{\sigma_f}{\sqrt{N}}. \quad (6.84)$$

Suponiendo que $\sigma_f < \infty$, el error en la estimación disminuye (en probabilidad) como $1/\sqrt{N}$ al aumentar N .

Cuando, además de converger $\langle f \rangle$, también converge $\langle f^2 \rangle$, o equivalentemente, $\sigma_f < \infty$, se aplica el **teorema del límite central**. Este teorema afirma que, en el límite $N \rightarrow \infty$, la variable \bar{f} sigue una distribución normal centrada en $\langle f \rangle$ y con dispersión σ_f/\sqrt{N} . En otras palabras, en el límite $N \rightarrow \infty$ la distribución de probabilidad de la variable $\varphi \equiv \frac{\bar{f} - \langle f \rangle}{\sigma_f/\sqrt{N}}$ tiende a $\frac{e^{-\varphi^2/2}}{\sqrt{2\pi}}$. Para una distribución normal, la probabilidad de que la variable se aparte de su media menos de 1, 2, 3, 4, 5, ..., desviaciones estándar es 68.2689 %, 95.4500 %, 99.7300 %, 99.9937 %, 99.9999 %, ...

Para N finito la distribución no es normal, pero se pueden aplicar relaciones más generales, como la desigualdad de Chebyshev [7]

$$\forall \xi > 0 \quad \text{Prob}(|X - \langle X \rangle| \geq \xi \sigma_X) \leq \frac{1}{\xi^2}. \quad (6.85)$$

Por ejemplo, la probabilidad de que X diste de su media más de 5 desviaciones estándar ($\xi = 5$) no puede superar el 4%. Esta desigualdad es válida cualquiera que sea la distribución de probabilidad de X . Para distribuciones concretas pueden darse cotas mejores, como se ha visto en el caso de la distribución normal.

En la práctica lo que se tiene es una muestra de las variables \mathbf{X}_i , con la cual se construye la estimación $\bar{f} = \frac{1}{N} \sum_{i=1}^N f(\mathbf{X}_i)$ de $\langle f \rangle$. Es muy importante notar que la dispersión en $f(\mathbf{X}_i)$ es σ_f y no disminuye con N , mientras que la dispersión de \bar{f} es σ_f/\sqrt{N} (si los N \mathbf{X}_i son independientes), que disminuye como $1/\sqrt{N}$ al aumentar N .

El valor de σ_f^2 no se conoce y lo que suele hacerse es estimarlo a su vez a partir de la **varianza de la muestra**:

$$S_f^2 \equiv \frac{1}{N-1} \sum_{i=1}^N (f_i - \bar{f})^2, \quad f_i \equiv f(\mathbf{X}_i). \quad (6.86)$$

S_f^2 es ella misma una variable aleatoria, igual que \bar{f} . (Intuitivamente el $N-1$ se debe a que, de los N valores $f_i - \bar{f}$, sólo $N-1$ independientes, por la relación $\sum_{i=1}^N (f_i - \bar{f}) = 0$.) La varianza de la muestra es un estimador no sesgado de la varianza de la variable:

$$\langle S_f^2 \rangle = \sigma_f^2. \quad (6.87)$$

En efecto, ya que

$$\begin{aligned} S_f^2 &= \frac{N}{N-1}(\overline{f^2} - (\bar{f})^2), \quad \langle \overline{f^2} \rangle = \langle f^2 \rangle, \\ (\bar{f})^2 &= \frac{1}{N^2} \left(\sum_i f_i^2 + \sum_{i \neq j} f_i f_j \right), \quad \langle (\bar{f})^2 \rangle = \frac{1}{N} \langle f^2 \rangle + \frac{N-1}{N} \langle f \rangle^2. \end{aligned} \quad (6.88)$$

Por otro lado, por la ley de los grandes números, aplicada a f y f^2 , se deduce que S_f^2 tiende a σ_f^2 con probabilidad 1 cuando $N \rightarrow \infty$:

$$\langle f \rangle = \bar{f} \pm \frac{S_f}{\sqrt{N}} \quad (\text{con probabilidad 1 para } N \rightarrow \infty). \quad (6.89)$$

Por supuesto, dado que S_f es una variable aleatoria, en un experimento concreto puede ocurrir que S_f no sea una buena estimación de σ_f , o bien N puede no ser lo suficientemente grande.

6.3.1. Promedios pesados

Con mucha frecuencia el problema que se quiere resolver es estimar el valor de

$$\langle f \rangle = \frac{\int_{\Omega} d^d x w(\mathbf{x}) f(\mathbf{x})}{\int_{\Omega} d^d x w(\mathbf{x})}, \quad w(\mathbf{x}) \geq 0, \quad \mathcal{N} \equiv \int_{\Omega} d^d x w(\mathbf{x}) < +\infty, \quad (6.90)$$

donde w es un peso no negativo normalizable pero no normalizado. Este problema se reduce al anterior definiendo

$$p(\mathbf{x}) \equiv \frac{1}{\mathcal{N}} w(\mathbf{x}). \quad (6.91)$$

Sin embargo, a menudo la normalización \mathcal{N} no se conoce y tampoco es fácilmente calculable (cuando $d \gg 1$). En realidad, como se verá, no es necesario conocer \mathcal{N} para hacer un **muestreo** de la distribución $p(\mathbf{x})$, es decir, construir una sucesión de variables aleatorias independientes tales que $\mathbf{X}_i \sim p$. Por abuso de lenguaje, también se suele denotar $\mathbf{X}_i \sim w$, aunque w no esté normalizada.

Ejemplo. Si $\psi(\mathbf{x})$ es la función de onda no normalizada de M partículas, $\mathbf{x} \in \mathbb{R}^d$, $d = 3M$, con hamiltoniano $H = T + V$ (energía cinética más potencial)

$$\langle V(\mathbf{x}) \rangle = \frac{\int d^d x |\psi(\mathbf{x})|^2 V(\mathbf{x})}{\int d^d x |\psi(\mathbf{x})|^2}, \quad (\text{Valor esperado del potencial}). \quad (6.92)$$

Por otro lado si $\psi(\mathbf{x})$ es un autoestado de H , podemos aplicar el teorema del virial

$$\langle T \rangle = \left\langle \frac{1}{2} \mathbf{x} \cdot \nabla V(\mathbf{x}) \right\rangle. \quad (6.93)$$

Siendo autoestado se podría obtener la energía mediante $H\psi = E\psi$, y de ahí $\langle T \rangle = E - \langle V \rangle$, sin embargo, este método no será viable si no se tiene $\psi(\mathbf{x})$ en forma analítica, mientras que la estimación MC puede usarse aunque sólo se tenga una muestra de la densidad de probabilidad $|\psi(\mathbf{x})|^2$ en un conjunto de puntos.

Ejemplo. (Mecánica estadística clásica): Sea ϕ la configuración del sistema. Por ejemplo, en un sistema de espines \uparrow, \downarrow , en una red con nodos $\Omega \subseteq \mathbb{Z}^d$, $\phi : \Omega \rightarrow \{\uparrow, \downarrow\}$, siendo $\phi_n = \uparrow$ ó \downarrow . Cada función ϕ es una configuración.

Cada configuración tiene una energía $E(\phi)$. La función de partición a temperatura absoluta T se define como

$$Z = \sum_{\phi} e^{-E(\phi)/kT}, \quad (6.94)$$

donde k es la constante de Boltzmann. De acuerdo con Boltzmann y Gibbs, la probabilidad de la configuración ϕ a temperatura T es

$$p(\phi) = \frac{1}{Z} e^{-E(\phi)/kT}, \quad \sum_{\phi} p(\phi) = 1. \quad (6.95)$$

A temperaturas frías el sistema está congelado en los estados de más baja energía, a temperaturas mayores la probabilidad de poblar estados más energéticos aumenta.⁶⁷

La energía interna U es el valor esperado de la energía

$$U = \langle E \rangle = \sum_{\phi} p(\phi) E(\phi) = \frac{\sum_{\phi} w(\phi) E(\phi)}{\sum_{\phi} w(\phi)}, \quad w(\phi) = e^{-E(\phi)/kT}. \quad (6.96)$$

No es necesario conocer Z , sino que se puede hacer un muestreo de $w(\phi)$, y estimar la energía interna con un MC

$$\langle E \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N E(\phi_i) \quad (\text{con prob. } 1), \quad \phi_i \sim e^{-E(\phi)/kT}. \quad (6.97)$$

⁶⁷Por otro lado, el número de estados con una energía dada E crece rápidamente con la energía, de modo que si $\rho(E)$ es la densidad de estados por unidad de energía, $Z = \int dE \rho(E) e^{-E/kT}$, el integrando tiene un máximo en una $E(T)$ que crece con la temperatura.

Para calcular Z , que es la normalización del peso $w(\phi)$, no se puede usar simplemente un MC basado en $\phi \sim w(\phi)$ y hay que emplear un método indirecto, por ejemplo

$$\frac{d \log Z}{d\beta} = -U, \quad \beta \equiv \frac{1}{kT}. \quad (6.98)$$

$U(T)$ se puede calcular con MC y luego integrar numéricamente sobre β , con la condición de contorno $Z \underset{T \rightarrow 0}{\sim} N_0 e^{-\beta E_0}$, siendo E_0 el mínimo de $E(\phi)$ (estado fundamental), y N_0 el número de configuraciones con energía E_0 (degeneración del estado fundamental). También se podría usar un método basado en reweighting (ver Sec. 6.4.8) pero en general eso da resultados con grandes barras de error.

Para ver aplicaciones del método MC en integral de caminos puede consultarse [10].

6.4. Métodos de muestreo

Para calcular los promedios citados es necesario saber generar variables \mathbf{X} distribuidas según una densidad de probabilidad $p(\mathbf{x})$ dada definida en un $\Omega \subseteq \mathbb{R}^d$, $\mathbf{X} \sim p(\mathbf{x})$. Esto quiere decir que

$$\text{Prob}(\mathbf{X} \in A) = \int_A d^d x p(\mathbf{x}) \quad \forall A \subseteq \Omega, \quad (6.99)$$

o en el caso discreto

$$\text{Prob}(\mathbf{X} \in A) = \sum_{\mathbf{x}_i \in A} p_i \quad \forall A \subseteq \Omega. \quad (6.100)$$

Un método de **muestreo** de p es un algoritmo que construya variables independientes distribuidas según p .

Ejemplo. Si por ejemplo los resultados del experimento son números reales $x \in [-L, L]$ con pesos $w(x) = \begin{cases} 1 & x < 0 \\ 2 & x > 0 \end{cases}$, la densidad de probabilidad será

$$p(x) = \frac{1}{3L} \Theta(-L < x < 0) + \frac{2}{3L} \Theta(0 < x < L). \quad (6.101)$$

Los puntos X_i deben caer con igual probabilidad en todo el intervalo $-L < x < 0$ y lo mismo en $0 < x < L$, y con el doble de probabilidad en $x > 0$ que en $x < 0$. Eso no quiere decir que al lanzar N puntos, $N/3$ caigan necesariamente a la izquierda y $2N/3$ a la derecha de 0. Más bien, si se generan N puntos independientes, la probabilidad de que n caigan en la parte positiva debe ser

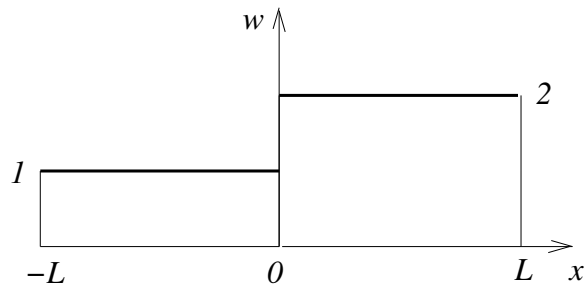


Figura 11: Pesos relativos 1 en la mitad izquierda y 2 en la derecha. En un muestreo correcto los puntos X_i deben caer con el doble de probabilidad a la derecha de 0. Asimismo, deben caer de modo equiprobable a lo largo del intervalo $[-L, 0]$ y también en $[0, L]$.

$\binom{N}{n} \frac{2^n}{3^N}$, de acuerdo con la distribución binomial. A medida que N crece las desviaciones respecto de las proporciones 1 : 2 deben disminuir como $1/\sqrt{N}$.

6.4.1. Números pseudo aleatorios

Para construir los algoritmos de muestreo, generalmente se hace uso de **números pseudo aleatorios**. Éstos son números generados por ordenador, completamente deterministas pero que pasan una amplia gama de tests de aleatoriedad. Por ejemplo, para una generador de bits (los resultados son 0 ó 1 con probabilidad $1/2$), se requiere que 0 y 1 salgan al 50% en un tirada de N bits, cuando N es grande, y que las desviaciones sean consistentes con lo que predice la teoría de la probabilidad. Se requiere que después de 0 siga la cadena 110 una de cada 8 veces, en promedio, que la frecuencia de aparición de 5 unos seguidos sea la correcta, etc, etc. En [11] hay herramientas para medir la aleatoriedad de un generador de números pseudo aleatorios dado.

Los generadores de números pseudo aleatorios suelen producir sucesiones periódicas. Importa que tengan un periodo largo, que sean poco costosos de producir y que sean suficientemente aleatorios. Muchas malas experiencias indican que deben usarse generadores de números pseudo aleatorios que sean bien entendidos y ampliamente contrastados. Un mal generador puede producir resultados directamente incorrectos al usarlos en MC. De ningún modo debe usarse un generador desconocido (y mucho menos uno casero), aunque parezca aleatorio, ni siquiera los que vienen en el hardware de los ordenadores. Fortran 90 viene con un generador, `random_number` (con `random_seed`), que se supone que es fiable (aunque el algoritmo depende del compilador) [12]. El uso de este generador se ilustra en [13].

Que los números pseudo aleatorios sean deterministas tiene la ventaja de que se puede reproducir un cálculo sin tener que guardar la lista de números que ha ido saliendo. También existen generadores de números aleatorios (basados en procesos físicos) pero son más costosos y menos prácticos en el contexto de MC.⁶⁸

Los generadores suelen producir números pseudo aleatorios (aproximadamente) uniformemente distribuidos entre 0 y 1 e independientes. La densidad de probabilidad asociada es $p(u) = \Theta(0 < u < 1)$ ⁶⁹ y esta distribución se suele denotar $U(0,1)$ (**distribución uniforme**), $U \sim U(0,1)$. Para esta distribución

$$\forall a, b \quad 0 \leq a \leq b \leq 1, \quad \text{Prob}(a < U < b) = b - a. \quad (6.102)$$

Todas las demás distribuciones se pueden construir a partir de la uniforme. Es decir, si X es una variable aleatoria con distribución $p_X(x)$ ⁷⁰

$$X = f(U_1, U_2, \dots, U_n) \quad (6.103)$$

donde las U_i son n variables uniformes en $(0,1)$ e independientes (proporcionadas por el generador de números pseudo aleatorios), y f es una función adecuada tal que $X \sim p_X$:

$$p_X(x) = \int_0^1 du_1 \int_0^1 du_2 \cdots \int_0^1 du_n \delta(x - f(u_1, u_2, \dots, u_n)). \quad (6.104)$$

Nótese que lo único aleatorio aquí son las U_i . f es una función fija perfectamente bien definida en cada caso. f equivale a un algoritmo para combinar las U_i y a menudo no es necesario explicitar f como función sino sólo describir (o programar) el algoritmo asociado. A diferencia de la función p_X , el algoritmo f no es ni mucho menos único y suele ser crucial elegir algoritmos que sean eficientes.

6.4.2. Distribuciones de variables discretas

Ejemplo. Generar un bit $X = 0, 1$, con probabilidad $p_0 = 1 - p$ y $p_1 = p$. Aquí p es un parámetro entre 0 y 1 que caracteriza la distribución. Se puede usar el siguiente algoritmo

$$X = \Theta(p - U) = \begin{cases} 1 & U \leq p \\ 0 & U > p \end{cases}, \quad U \sim U(0,1). \quad (6.105)$$

⁶⁸Números aleatorios (frente a pseudo aleatorios) se usan en criptografía y en casos en los cuales es importante que el proceso no se pueda reproducir para no desvelar información que se quiere mantener secreta. Como son muy costosos de producir, principalmente se usan para generar la semilla de una sucesión de números pseudo aleatorios.

⁶⁹Dependiendo del generador 0 y/o 1 pueden estar incluidos o no. Idealmente esto no debería ser relevante para MC, al ser $\{0\} \cup \{1\}$ un conjunto de medida nula.

⁷⁰Esto es una simplificación. Más generalmente, X se construye mediante algún algoritmo bien definido pero que puede involucrar un número indefinido de variables U_i . Este es caso de los algoritmos de aceptación-rechazo.

En efecto, $\text{Prob}(X = 1) = \text{Prob}(U \leq p) = p$. (De nuevo, es irrelevante poner $U \leq p$ o $U < p$, para concretar ponemos lo primero.)

Ejemplo. Queremos generar una variable discreta con valores $X = 1, \dots, n$, cada uno con una probabilidad p_i dada ($0 \leq p_i, \sum_i p_i = 1$). Un algoritmo es

$$X = i \quad \text{sii} \quad \sum_{j < i} p_j < U \leq \sum_{j \leq i} p_j, \quad U \sim U(0, 1). \quad (6.106)$$

Se genera U y se van sumando las p_j ; $X = i$ es el primer valor de i tal que $U \leq p_1 + \dots + p_i$. Este

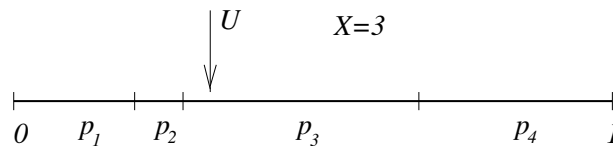


Figura 12: En el ejemplo, $p_1 + p_2 < U < p_1 + p_2 + p_3$, por tanto $X = 3$ en este caso.

método vale igual para n finito o infinito. Si se van a hacer muchas extracciones puede convenir calcular y guardar las sumas parciales, $s_i \equiv \sum_{j \leq i} p_j$. En este caso, dado U se busca el primer i tal que $U \leq s_i$. Alternativamente, se pueden guardar algunas sumas parciales, por ejemplo, las de bloques de tamaño m , s_m, s_{2m} , etc. En este caso se procede a sortear primero el bloque, y luego dentro del bloque, entre los m valores posibles.

Ejemplo. (Casos equiprobables.) Lo anterior es para un caso general, con p_i arbitrarias. Para n casos *equiprobables*, no es necesario hacer una búsqueda sobre los casos, simplemente

$$X = \lceil nU \rceil. \quad (6.107)$$

La función *techo*, $\lceil x \rceil$, quiere decir redondear a entero por arriba (es decir, el menor entero mayor o igual que x). Cuando U recorre uniformemente $(0, 1)$, nU recorre $(0, n)$ y $\lceil nU \rceil$ tiene la misma probabilidad de tomar cualquiera de los valores $1, \dots, n$.

Ejemplo. (Método de alias.) En general, la búsqueda de primer i tal que $U \leq s_i$ puede ser costosa. Si hay que hacerla muchas veces y n es grande, puede ser conveniente usar el método alternativo de **alias**. Veamos cómo funciona con un ejemplo. Supongamos que tenemos $n = 4$ casos, con probabilidades $p_1 = 0.36, p_2 = 0.34, p_3 = 0.10$ y $p_4 = 0.20$. Lo que se hace es construir $2n = 8$ casos organizados en $n = 4$ columnas equiprobables (probabilidad $1/n = 0.25$ cada una). El resultado final está en la tabla de la derecha. Los 4 casos originales están ahora repartidos en 8 casos. Por

1	2	3	4	0.25	0.25	0.25	0.25
0.36	0.34	0.10	0.20	0.10 ₃	0.20 ₄	0.21 ₁	0.25 ₂
0.21	0.34		0.20	0.15 ₁	0.05 ₂	0.04 ₂	0.00 ₂
0.21	0.29						
	0.25						

Cuadro 1: A la izquierda evolución de las probabilidades en las 4 iteraciones. A la derecha resultado final de la distribución en 8 casos. El valor de i aparece como subíndice. Las 4 columnas son equiprobables.

ejemplo, el caso 2 aparece en cuatro sitios, con probabilidad total $0.05 + 0.04 + 0.25 + 0.00 = 0.34$. Para usar la tabla, se sortea entre los 8 casos con las probabilidades indicadas, pero como cada columna es equiprobable, basta sortear primero la columna, lo cual no requiere hacer una búsqueda, y una vez determinada la columna, se sortea entre los dos casos de esa columna. Por ejemplo, para la tercera, habrá una probabilidad $0.21/0.25 = 0.84$ para el caso 1 y una probabilidad $0.04/0.25 = 0.16$ para el caso 2.

Para construir la tabla de la derecha, en la primera iteración se toma el caso menos probable (el 3, con probabilidad 0.10) que ocupara la posición de arriba de la primera columna. La columna se completa con lo que haga falta hasta 0.25 del caso más probable (el 1, del cual tomamos 0.15). Después de descontar esas probabilidades transferidas a la tabla, las nuevas probabilidades son las indicadas en la segunda línea de la izquierda (0.21, 0.34, 0.00 y 0.20). En la segunda iteración se procede del mismo modo: se mueve la probabilidad más pequeña (0.20 del caso 4) a la segunda columna arriba, y se completa la columna con el caso más probable hasta 0.25 (transferimos 0.05 del caso 2). Las probabilidades quedan como se indica en la tercera línea (0.21, 0.29, 0.00 y 0.00). Se procede del mismo modo con las dos columnas restantes.

El método de alias evita hacer una búsqueda pero requiere el trabajo previo de construir y guardar la información de la tabla. No puede aplicarse (al menos tal cual) al caso n infinito.

Ejemplo. (Distribución de Poisson). Una variable discreta $n = 0, 1, \dots$ sigue una distribución de Poisson con media μ si

$$P_n = e^{-\mu} \frac{\mu^n}{n!}, \quad n = 0, 1, 2, \dots \quad (6.108)$$

Es inmediato comprobar que está normalizada, $\sum_{n=0}^{\infty} P_n = 1$, así como que⁷¹

$$\langle n \rangle = \mu, \quad \text{Var}(n) = \mu. \quad (6.109)$$

⁷¹Este resultado se puede obtener usando la función generatriz (pág. 103) $g(x) = e^{(x-1)\mu}$.

La distribución de Poisson es el límite de la binomial cuando N es grande y p pequeña, con $\mu = pN$ y n fijos. Una realización de la distribución de Poisson es una fuente emisora estacionaria (por ejemplo una muestra radioactiva de vida media larga, una cola de llegada de clientes) tal que los lapsos entre dos emisiones sucesivas τ_n sean variables aleatorias independientes con distribución exponencial $\tau_n \sim \Theta(\tau)e^{-\tau}$ (eligiendo las unidades para que el ritmo de emisión sea una emisión por unidad de tiempo en media). Es instructivo ver que en efecto, el número de emisiones, n , después de un tiempo μ sigue una distribución de Poisson. Y además proporciona un método para muestrear esta distribución.

Por hipótesis los lapsos τ_k , $k = 1, 2, \dots$, son independientes y siguen una distribución $\Theta(\tau)e^{-\tau}$. Los tiempos en los que se producen las emisiones $n = 1, 2, \dots$ son

$$T_n = \tau_1 + \tau_2 + \dots + \tau_n, \quad n = 1, 2, \dots \quad (6.110)$$

El algoritmo es generar τ_1 y si $T_1 = \tau_1 > \mu$, $n = 0$. En otro caso, se genera τ_2 . Si $T_2 = \tau_1 + \tau_2 > \mu$, $n = 1$. Y así sucesivamente. La probabilidad de obtener un valor n es

$$P_n(\mu) = \text{Prob}(T_n \leq \mu < T_{n+1}) = \int dt_n dt_{n+1} p_{T_n, T_{n+1}}(t_n, t_{n+1}) \Theta(t_n \leq \mu < t_{n+1}). \quad (6.111)$$

Como las emisiones son independientes

$$p_{T_n, T_{n+1}}(t_n, t_{n+1}) = p_{T_{n+1}|T_n}(t_{n+1}|t_n) p_{T_n}(t_n) = e^{-(t_{n+1}-t_n)} p_{T_n}(t_n) \Theta(t_{n+1} - t_n). \quad (6.112)$$

De aquí,

$$P_n(\mu) = \int_0^\infty dt_n \int_0^\infty d\tau_{n+1} p_{T_n}(t_n) \Theta(\mu - t_n) \Theta(t_n + \tau_{n+1} - \mu) e^{-\tau_{n+1}} = \int_0^\mu dt_n p_{T_n}(t_n) e^{-(\mu - t_n)}. \quad (6.113)$$

Necesitamos p_{T_n} . Como $T_n = \sum_{k=1}^n \tau_k$ y las τ_k son independientes, p_{T_n} es la convolución de las n distribuciones $p_\tau(\tau) = e^{-\tau} \Theta(\tau)$

$$p_{T_n}(t_n) = (p_\tau * \dots * p_\tau)(t_n) \equiv p_\tau^{*n}(t_n). \quad (6.114)$$

Una forma práctica de proceder es tomar transformada de Fourier

$$\tilde{p}_\tau(k) = \int_0^{+\infty} d\tau e^{-\tau} e^{-ik\tau} = \frac{1}{1+ik}, \quad \tilde{p}_{T_n}(k) = \frac{1}{(1+ik)^n}. \quad (6.115)$$

(Comprobamos que $\tilde{p}_\tau(0) = 1$, la normalización es correcta.) Entonces

$$p_{T_n}(t_n) = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} e^{ikt_n} \frac{1}{(1+ik)^n} = 2\pi i \text{Res}_{k=i} \frac{e^{ikt_n}}{(1+ik)^n} = e^{-t_n} \frac{t_n^{n-1}}{(n-1)!}. \quad (6.116)$$

(De nuevo se comprueba que $\int_0^\infty dt_n p_{T_n}(t_n) = 1$.) Podemos ya obtener la distribución de n ,

$$P_n(\mu) = \int_0^\mu dt_n e^{-t_n} \frac{t_n^{n-1}}{(n-1)!} e^{-(\mu-t_n)} = e^{-\mu} \frac{\mu^n}{n!}. \quad (6.117)$$

El cálculo está hecho para $n \geq 1$. El caso $n = 0$ se puede calcular por separado:

$$p_{T_0}(t_0) = \text{Prob}(\mu < T_1) = \int_0^{+\infty} d\tau_1 e^{-\tau_1} \Theta(\tau_1 - \mu) = \int_\mu^{+\infty} d\tau e^{-\tau} = e^{-\mu}. \quad (6.118)$$

6.4.3. Distribuciones de variables continuas

En la construcción de algoritmos de *variables continuas* son útiles las relaciones

$$p_{X+a}(x) = p_X(x-a), \quad p_{\lambda X}(x) = \frac{1}{|\lambda|} p_X(x/\lambda) \quad (\lambda \neq 0). \quad (6.119)$$

Ejemplo. Construcción de $X \sim U(a, b)$ ($a < b$). En este caso $p(x) = \Theta(a < x < b)/(b-a)$, y se obtiene de la distribución uniforme aplicando un factor de escala, de $(0, 1)$ a $(0, b-a)$ y luego una traslación a (a, b) . Por tanto

$$X = a + (b-a)U \quad U \sim U(0, 1). \quad (6.120)$$

Ejemplo. Muestreo de $\mathbf{X} \sim U([a_1, b_1] \times \cdots \times [a_d, b_d])$, $\mathbf{X} \in \mathbb{R}^d$. Simplemente se genera como antes cada una de las componentes de \mathbf{X} , es decir, $X^i \sim U(a_i, b_i)$.

Esto es general: si una distribución es separable, usualmente lo mejor será generar las distintas coordenadas por separado. Si dos (o más) subconjuntos de coordenadas son independientes entre sí, usualmente será más eficiente hacer un muestreo de cada subconjunto por separado, ya que la dificultad de muestreo suele crecer con la dimensión.

6.4.4. Método de inversión

Sea $p(x)$ una densidad de probabilidad con soporte (a, b) , es decir, $p(x) > 0$ sii $a < x < b$, donde a, b pueden ser finitos o infinitos. Se quiere hacer un muestreo de p , es decir, construir una variable aleatoria $X \sim p$.

El método de inversión consiste en definir una función $h(x)$ mediante

$$h(x) \equiv \int_a^x dy p(y) \quad x \in [a, b]. \quad (6.121)$$

Por construcción esta función es estrictamente creciente en $[a, b]$ con $h(a) = 0$ y $h(b) = 1$. Por tanto $u = h(x)$ satisface $0 \leq u \leq 1$. Ahora definimos la función inversa $x = f(u)$ ($f = h^{-1}$) que también es estrictamente creciente. En este caso la variable aleatoria

$$X = f(U) \quad U \sim U(0, 1) \quad (6.122)$$

está distribuida según p , es decir $X \sim p$ ó $p_X = p$.

En efecto, para demostrar esto notemos que

$$\forall x \in [a, b] \quad \text{Prob}(X < x) = \text{Prob}(f(U) < x) = \text{Prob}(U < h(x)) = h(x) = \int_a^x dy p(y) \quad (6.123)$$

por otro lado

$$\text{Prob}(X < x) = \int_a^x dy p_X(y) \quad (6.124)$$

en consecuencia $p_X = p$.

Éste es el denominado **método de inversión**.⁷² Equivalentemente, puede usarse $u(x) = \text{Prob}(X > x) = \int_x^b p_X(x') dx'$.

Ejemplo. Para hacer un muestreo de X con distribución $p_X(x) = e^{-x}\Theta(x)$, podemos usar el método de inversión:

$$u(x) = \text{Prob}(X < x) = \int_{-\infty}^x \Theta(y)e^{-y}dy = \Theta(x) \int_0^x e^{-y}dy = \Theta(x)(1 - e^{-x}). \quad (6.125)$$

Invirtiendo (para x positiva)

$$X = -\log(1 - U). \quad (6.126)$$

También sirve $X = -\log(U)$ ya que U y $1 - U$ están igualmente distribuidos.

Otra forma instructiva de plantearlo es

$$\langle f \rangle = \int_0^\infty dx e^{-x} f(x) = \int_{x=0}^{x=\infty} d(1 - e^{-x}) f(x) = \int_0^1 du f(x(u)), \quad (6.127)$$

⁷²Nótese que el método de inversión $U = \int_{-\infty}^X p_X(x) dx$, es la versión continua del método visto para variables discretas, $s_{i-1} < U \leq s_i$, ec. (6.106).

con el cambio de variable $u = 1 - e^{-x}$, igual que antes. Más generalmente $dxp(x) = du$. El peso 1 en $\int_0^1 du f$ indica que $u \sim U(0, 1)$.

Otra forma de plantearlo: Un método para producir muestreos de nuevas densidades de probabilidad es relacionarlas con otras mediante un cambio de variable. Para un cambio de variable $\mathbf{X}(\mathbf{Y})$, las densidades de probabilidad está relacionadas por

$$p_{\mathbf{X}}(\mathbf{x}(\mathbf{y})) = \left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right| p_{\mathbf{Y}}(\mathbf{y}). \quad (6.128)$$

Si sabemos hacer un muestreo de \mathbf{Y} , la idea es elegir el cambio de variable de modo que $p_{\mathbf{X}}(\mathbf{x})$ sea la densidad de probabilidad deseada. En el caso unidimensional, podemos elegir que $Y = U \sim U(0, 1)$, por tanto $p_U(u) = 1$ para $0 < u < 1$. En consecuencia el cambio de variable $X(U)$ debe cumplir (elegimos por ejemplo que X sea una función creciente de U)

$$p_X(x) = \left| \frac{du}{dx} \right| p_U(u) = \frac{du}{dx}. \quad (6.129)$$

Integrando desde a hasta un x cualquiera⁷³

$$u(x) = \int_a^x p_X(x') dx' = \text{Prob}(X < x). \quad (6.130)$$

Para aplicar este método se genera $U \sim U(0, 1)$, y se obtiene $X \equiv x(U)$ tal que $X \sim p_X$. Aquí $x(u)$ la función inversa de $u(x) \equiv \text{Prob}(X < x)$.

Una generalización del método de inversión para varias variables es la siguiente: Sea $\mathbf{X} \in \mathbb{R}^d$ y $\mathbf{U} = U([0, 1]^d)$, entonces

$$\begin{aligned} u_1(x_1) &= \text{Prob}(X_1 < x_1) = \int_{-\infty}^{x_1} dy_1 \int_{-\infty}^{+\infty} dy_2 \cdots \int_{-\infty}^{+\infty} dy_d p_{\mathbf{X}}(y_1, \dots, y_d) \\ &\vdots \\ u_n(x_1, \dots, x_n) &= \text{Prob}(X_n < x_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1}) \\ &= \frac{\int_{-\infty}^{x_n} dy_n \int_{-\infty}^{+\infty} dy_{n+1} \cdots \int_{-\infty}^{+\infty} dy_d p_{\mathbf{X}}(x_1, \dots, x_{n-1}, y_n, \dots, y_d)}{\int_{-\infty}^{+\infty} dy_n \int_{-\infty}^{+\infty} dy_{n+1} \cdots \int_{-\infty}^{+\infty} dy_d p_{\mathbf{X}}(x_1, \dots, x_{n-1}, y_n, \dots, y_d)} \\ &\vdots \end{aligned} \quad (6.131)$$

Como se puede comprobar $\left| \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right| = p_{\mathbf{X}}(x_1, \dots, x_d)$, y en consecuencia $\mathbf{X}(\mathbf{U})$ (invirtiendo las relaciones) está distribuido según $p_{\mathbf{X}}$. En la práctica esta versión multidimensional no puede aplicarse ya que no es fácil hacer las integrales requeridas ni invertir las ecuaciones.

⁷³También se puede integrar desde $-\infty$ entendiéndose que $p_X(x)$ se anula para $x < a$.

6.4.5. Distribución normal

Cuando la distribución de una variable es de tipo gaussiano

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (6.132)$$

(el prefactor es tal que p_X está normalizada) se dice que X sigue una **distribución normal** centrada en μ y con dispersión σ , ya que en efecto $\langle X \rangle = \mu$ y $\text{Var}(X) = \sigma^2$, $X \sim N(\mu, \sigma^2)$. Evidentemente basta saber muestrear $N(0, 1)$ (y luego $X \mapsto \mu + \sigma X$). Para hacer el muestreo se puede usar el **método de Box-Muller**: se generan dos variables normales independientes X e Y , con distribución

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}, \quad (6.133)$$

y se trabaja en coordenadas polares:

$$\begin{aligned} \langle f \rangle &= \int \frac{dx dy}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} f(x,y) = \int_0^{2\pi} \frac{d\phi}{2\pi} \int_0^\infty dr r e^{-r^2/2} f \\ &= \int_0^{2\pi} \frac{d\phi}{2\pi} \int_{r=0}^{r=\infty} d(1 - e^{-r^2/2}) f = \int_0^1 du_\phi \int_0^1 du_r f, \end{aligned} \quad (6.134)$$

con los cambios de variables

$$u_\phi = \frac{\phi}{2\pi}, \quad u_r = e^{-r^2/2}. \quad (6.135)$$

La última expresión en ec. (6.134) indica que u_ϕ y u_r son independientes y están distribuidas según $U(0, 1)$. Invirtiendo los cambios de variable para pasar de (u_ϕ, u_r) a (ϕ, r) y de ahí a (x, y) , se obtiene

$$\begin{aligned} X &= \cos(2\pi U_\phi) \sqrt{-2 \log U_r} \\ Y &= \sin(2\pi U_\phi) \sqrt{-2 \log U_r}, \quad U_\phi, U_r \sim U(0, 1) \text{ e independientes.} \end{aligned} \quad (6.136)$$

6.4.6. Puntos uniformemente distribuidos en una región

Sea $\Omega \subseteq \mathbb{R}^d$ con volumen V finito, y queremos \mathbf{X} distribuida uniformemente dentro de Ω , $\mathbf{X} \sim U(\Omega)$, es decir, $p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{V} \Theta(\mathbf{x} \in \Omega)$.

Si Ω tiene un límites acotados conocidos, un algoritmo sencillo es el **método de aceptación-rechazo**, que consiste en meter Ω en una caja rectangular, $\Omega \subseteq [a_1, b_1] \times \cdots \times [a_d, b_d]$. Se genera \mathbf{X}

uniformemente en la caja. El valor se rechaza si $\mathbf{X} \notin \Omega$ y se genera uno nuevo hasta que se acepte. Por construcción ese \mathbf{X} estará distribuido correctamente, $\mathbf{X} \sim U(\Omega)$.

El método funciona igual si \mathbf{X} se genera uniformemente en una región cualquiera (no necesariamente rectangular) que contenga a Ω . La eficiencia de este método depende de la probabilidad de aceptación, que es el cociente entre el volumen de Ω y el de la caja. Conviene que la caja sea lo más ajustada posible para aumentar la probabilidad de aceptación (y no desperdiciar recursos generando puntos que no se van a aceptar). Incluso con una caja ajustada el método puede ser ineficiente, como se ve en el siguiente ejemplo.

Ejemplo. (Puntos uniformes dentro o sobre una esfera). Supongamos que la región es una esfera de radio 1 en \mathbb{R}^d y su interior, $B_d = \{\|x\| \leq 1, x \in \mathbb{R}^d\}$. Aplicamos el método de aceptación-rechazo con el cubo $[-1, 1]^d$. Las d coordenadas de la \mathbf{X} propuesta se obtienen fácilmente con $X_i = 2U_i - 1$ (con $U_i \sim U(0, 1)$, $i = 1, \dots, d$ e independientes). La propuesta se acepta sii $\|\mathbf{X}\| \leq 1$. Un \mathbf{X} aceptado está distribuido uniformemente en el *interior* de la esfera B_d .

Si lo que se quiere es que \mathbf{X} esté definido uniformemente *sobre* la esfera $S^d = \{\|x\| = 1, x \in \mathbb{R}^d\}$, el método es construir \mathbf{X} en el interior como antes y luego normalizarlo con $\mathbf{X}/\|\mathbf{X}\|$. (El caso $\mathbf{X} = \mathbf{0}$ es casi imposible, si ocurriera bastaría generar \mathbf{X} de nuevo.)⁷⁴

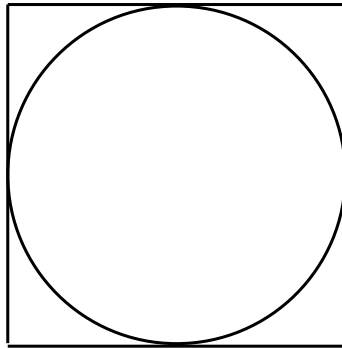


Figura 13: Caja cuadrada ajustada a la esfera para $d = 2$.

En cualquiera de los dos casos se necesita generar un \mathbf{X} dentro de la esfera y la probabilidad de aceptación es el cociente entre el volumen de la esfera V_d y el del cubo, que es 2^d .

Aunque para dimensiones bajas no lo parezca, la probabilidad de aceptación disminuye rápida-

⁷⁴En teoría de la probabilidad, los sucesos que no son \emptyset (suceso imposible) pero tienen probabilidad cero se denominan **casi imposibles**. Igualmente los sucesos con probabilidad 1, excepto Ω mismo, se denominan **casi seguros**.

mente con la dimensión. Esto es fácil de entender notando que las d coordenadas x_i se generan uniformemente en $[-1, 1]$ y la aceptación requiere que $x_1^2 + \dots + x_d^2 \leq 1$. Dado que $\text{Prob}(x_i^2 < 1/d) = 1/\sqrt{d}$, se puede estimar que la probabilidad de aceptación disminuye como $d^{-d/2}$.

Es instructivo ver cómo se puede calcular exactamente V_d y de ahí la probabilidad de aceptación. Consideremos la integral, que es separable en coordenadas cartesianas,

$$\begin{aligned} (2\pi)^{d/2} &= \int d^d x e^{-x^2/2} = S_{d-1} \int_0^\infty dr r^{d-1} e^{-r^2/2} = S_{d-1} 2^{\frac{d}{2}-1} \int_0^\infty dx x^{\frac{d}{2}-1} e^{-x} \\ &= S_{d-1} 2^{\frac{d}{2}-1} \Gamma(d/2). \end{aligned} \quad (6.137)$$

Se han usado coordenadas polares, y S_{d-1} es el ángulo sólido en d dimensiones (el área de la esfera de radio 1 en d dimensiones). Γ es la función gamma de Euler.⁷⁵ Se deduce que

$$S_{d-1} = \frac{2\pi^{d/2}}{\Gamma(d/2)}. \quad (6.138)$$

(Por ejemplo, $S_0 = 2$, $S_1 = 2\pi$, $S_2 = 4\pi$.) Ahora podemos calcular el volumen de la esfera d -dimensional

$$V_d = \int d^d x \Theta(1 - \|x\|) = S_{d-1} \int_0^1 dr r^{d-1} = \frac{1}{d} S_{d-1} = \frac{\pi^{d/2}}{\Gamma(d/2 + 1)}. \quad (6.139)$$

d	p
1	1
2	0.79
3	0.52
10	2.5×10^{-3}
20	2.5×10^{-8}
100	2×10^{-70}

Cuadro 2: Probabilidad de acertar en la esfera en una tirada uniforme en el cubo, en función de la dimensión.

Finalmente, la probabilidad de aceptación será ⁷⁶

$$p = \frac{V_d}{2^d} = \frac{(\pi/4)^{d/2}}{\Gamma(\frac{d}{2} + 1)}. \quad (6.140)$$

⁷⁵ $\Gamma(\alpha) = \int_0^\infty dx x^{\alpha-1} e^{-x}$. Propiedades $\Gamma(x+1) = x\Gamma(x) = x!$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(x) \sim x^x$.

⁷⁶Y en efecto, para d grande $\log p \asymp -\frac{d}{2} \log d$.

Como se ve en la tabla adjunta, la probabilidad de aceptación decrece muy rápidamente a medida que la dimensión crece. Para dimensiones medianas o grandes el método de aceptación-rechazo no es viable y hay que buscar métodos alternativos. Aparte de los métodos markovianos que se verán más adelante, en el caso de la esfera se puede hacer separación de variables en coordenadas polares, generando los $d - 1$ ángulos que permiten reconstruir \mathbf{X} . Por ejemplo para $d = 3$

$$1 = \frac{3}{4\pi R^3} \int_0^R dr r^2 \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin \theta = \int_0^1 du_r \int_0^1 du_\phi \int_0^1 du_\theta, \quad (6.141)$$

$$u_r = \frac{r^3}{R^3}, \quad u_\phi = \frac{\phi}{2\pi}, \quad u_\theta = \frac{\cos \theta}{2} \quad u_r, u_\phi, u_\theta \sim U(0, 1) \quad \text{e independientes.}$$

Es importante notar en este ejemplo que métodos de muestreo que parecen razonables a primera vista pueden ser extremadamente ineficientes cuando vamos a dimensiones grandes (que es cuando se requiere MC). La intuición basada en el caso $d = 2$ (en la Fig. 13) o $d = 3$, el espacio ordinario para el que $p = 0.52$, puede producir impresiones equivocadas respecto de lo que ocurrirá para d grande. Como regla, a medida que d crece, las zonas que son realmente relevantes (en el ejemplo, el volumen ocupado por la esfera) tienden a ser una parte muy pequeña del total y se hace necesario usar métodos eficientes para que el MC sea viable.

6.4.7. Método de aceptación-rechazo

Entre los métodos para obtener muestreos de distribuciones genéricas (en particular no separables) está el método de **aceptación-rechazo**.

Queremos hacer un muestreo $\mathbf{X} \sim w(\mathbf{x})$ donde $w(\mathbf{x})$ es un peso no necesariamente normalizado. Denotamos \mathcal{N} la normalización de $w(\mathbf{x})$ (en general desconocida) y $p(\mathbf{x}) = w(\mathbf{x})/\mathcal{N}$ la probabilidad normalizada.

En la versión mas simple del método se supone que w tiene soporte en cierta región $\Omega \subseteq \mathbb{R}^d$ y que podemos hacer un muestreo uniforme de Ω , esto sólo es posible si Ω tiene un volumen finito, V . Además suponemos conocido un C tal que $w(\mathbf{x}) \leq C \forall \mathbf{x} \in \Omega$. El algoritmo consiste en generar un candidato \mathbf{x} uniformemente en Ω , y este candidato se acepta con probabilidad $w(\mathbf{x})/C$. Es decir, se genera $U \sim U(0, 1)$ y \mathbf{x} se acepta ($\mathbf{X} = \mathbf{x}$) si $UC \leq w(\mathbf{x})$. Si el candidato no se acepta se genera uno nuevo. Los candidatos aceptados están distribuidos según $w(\mathbf{x})$. (Intuitivamente es obvio que \mathbf{x} se acepta proporcionalmente a $w(\mathbf{x})$.) No es necesario conocer la normalización \mathcal{N} de $w(\mathbf{x})$, sólo saber calcular $w(\mathbf{x})$ en cada punto que se pida y conocer el valor de un C válido (que sea una cota superior).

La probabilidad de aceptación es ($V = |\Omega|$)

$$p_{\text{accept}} = \int_{\Omega} d^d x \frac{1}{V} \int_0^1 du \Theta(uC < w(\mathbf{x})) = \int_{\Omega} d^d x \frac{w(\mathbf{x})}{VC} = \frac{\mathcal{N}}{VC}. \quad (6.142)$$

Obviamente, para aumentar la probabilidad de aceptación conviene tomar C lo más ajustada (pequeña) posible.

El gran problema de este método es que la probabilidad de aceptación puede ser extremadamente pequeña en casos prácticos (como se ha visto en el caso de la esfera d -dimensional). De hecho, el factor \mathcal{N}/V no es más que el promedio de $w(\mathbf{x})$ en Ω y lo que se tiene es

$$\langle w \rangle_{\Omega} \leq \max_{\mathbf{x} \in \Omega} w(\mathbf{x}) \leq C, \quad p_{\text{accept}} = \frac{\langle w \rangle_{\Omega}}{C} \leq \frac{\langle w \rangle_{\Omega}}{w_{\text{máx}}}. \quad (6.143)$$

Incluso eligiendo C óptimamente, la probabilidad de aceptación sólo puede ser importante si $\langle w \rangle_{\Omega} \approx w_{\text{máx}}$. y esto sólo es posible si la función $w(\mathbf{x})$ es muy plana en Ω . El método sólo es útil para d no muy grandes y w con poca variación en Ω , de modo que w/C pueda mantenerse próximo a 1. Otra condición que puede ser limitante es que hace falta que Ω sea fácilmente muestreable.

Una forma de mejorar el problema de la probabilidad de aceptación es generar directamente más puntos en la zona relevante de w . Supongamos que 1) tenemos una densidad de probabilidad (normalizada) auxiliar $q(\mathbf{x})$ que sea similar a $p(\mathbf{x})$, 2) q sea fácil de muestrear y 3) además conocemos un K , lo más pequeño posible, tal que $w(\mathbf{x}) \leq Kq(\mathbf{x}) \quad \forall \mathbf{x} \in \Omega$. En ese caso podemos usar el siguiente algoritmo de aceptación-rechazo:

- 1) Se genera un candidato $\mathbf{x} \sim q(\mathbf{x})$.
- 2) Se acepta con probabilidad $\frac{w(\mathbf{x})}{Kq(\mathbf{x})}$. Es decir, se genera $U \sim U(0, 1)$ y se acepta sii

$$UKq(\mathbf{x}) \leq w(\mathbf{x}).$$

En caso contrario se genera un nuevo candidato.

Intuitivamente se ve que se producen más candidatos donde q es mayor, pero se compensa con que se aceptan menos (para w dado). La mejora se debe a que si $q(\mathbf{x})$ es similar a $p(\mathbf{x})$, el cociente w/q variará poco y se puede tomar un K que sea parecido a ese cociente en todo Ω , lo cual aumenta la probabilidad de aceptación. La probabilidad de aceptación es

$$p_{\text{accept}} = \int_{\Omega} d^d y q(\mathbf{y}) \int_0^1 du \Theta(w(\mathbf{y}) - uKq(\mathbf{y})) = \int_{\Omega} d^d y q(\mathbf{y}) \frac{w(\mathbf{y})}{Kq(\mathbf{y})} = \frac{\mathcal{N}}{K}. \quad (6.144)$$

En particular si w está normalizada ($\mathcal{N} = 1$) la probabilidad es $1/K$. En la aplicación de este método debe evitarse que $q(\mathbf{x})$ sea pequeña donde $w(\mathbf{x})$ no lo es, ya que en esas regiones w/q será muy grande lo fuerza a que K también lo sea. La probabilidad la distribución q debe ser, en todo caso, más extendida (mayor soporte, más ancha) que w .

Veamos que el método es correcto, es decir, no introduce un sesgo y un \mathbf{X} aceptado está distribuido según w :

$$\begin{aligned} \mathbf{x}(\text{aceptado}) &\sim \mathbb{E}(\delta(\mathbf{X} - \mathbf{x}) | (\mathbf{X}, u) \text{ aceptado}) \\ &= \frac{\int_{\Omega} d^d \mathbf{y} q(\mathbf{y}) \int_0^1 du \delta(\mathbf{y} - \mathbf{x}) \Theta(w(\mathbf{y}) - uKq(\mathbf{y}))}{\int_{\Omega} d^d \mathbf{y} q(\mathbf{y}) \int_0^1 du \Theta(w(\mathbf{y}) - uKq(\mathbf{y}))} \\ &= \frac{w(\mathbf{x})/K}{\mathcal{N}/K} = \frac{w(\mathbf{x})}{\mathcal{N}} = p(\mathbf{x}). \end{aligned} \quad (6.145)$$

El método de aceptación-rechazo es general y fácil de usar. Su principal inconveniente es que en la práctica, cuando se va a dimensiones grandes, distribuciones q aparentemente parecidas a la w dada no lo son en absoluto y la probabilidad de aceptación se hace extremadamente pequeña. El método funcionará cuando sea posible satisfacer el requerimiento de **muestreo relevante** (*importance sampling*) es decir, que q ponga más puntos donde w es mayor.

6.4.8. Método de reweighting

Relacionado con la idea de usar una probabilidad auxiliar está el **método de reweighting**. Este método no produce un muestreo de la probabilidad dada pero sí permite calcular los valores esperados asociados.

Dado un peso $w(\mathbf{x}) = \mathcal{N} p(\mathbf{x})$ y una densidad de probabilidad $q(\mathbf{x})$ auxiliar, que suponemos normalizada,

$$\langle A(\mathbf{X}) \rangle_w = \frac{\int d^d \mathbf{x} w(\mathbf{x}) A(\mathbf{x})}{\int d^d \mathbf{x} w(\mathbf{x})} = \frac{\int d^d \mathbf{x} q(\mathbf{x}) w(\mathbf{x}) A(\mathbf{x}) / q(\mathbf{x})}{\int d^d \mathbf{x} q(\mathbf{x}) w(\mathbf{x}) / q(\mathbf{x})} = \frac{\langle Aw/q \rangle_q}{\langle w/q \rangle_q}. \quad (6.146)$$

De acuerdo con esta identidad una forma de proceder es calcular los promedios de $A(\mathbf{X})w(\mathbf{X})/q(\mathbf{X})$ y $w(\mathbf{X})/q(\mathbf{X})$ para $\mathbf{X} \sim q$ y el cociente de ambos nos proporcionará el promedio de $A(\mathbf{X})$ para $\mathbf{X} \sim w$. Igual que en el método de aceptación-rechazo, para $\mathbf{X} \sim q$, \mathbf{X} tenderá a producirse más donde q (y no necesariamente w) es importante, pero eso se compensa por el factor w/q que será menor donde q sea grande.

El denominador $\langle w/q \rangle_q$ no es más que \mathcal{N} , y en particular, cuando $w = p$ (w está normalizado) la expresión queda

$$\langle A(\mathbf{X}) \rangle_p = \left\langle \frac{p(\mathbf{X})}{q(\mathbf{X})} A(\mathbf{X}) \right\rangle_q. \quad (6.147)$$

Es importante notar que aunque los promedios coinciden, las varianzas no lo hacen,

$$\text{Var}(pA/q)_q - \text{Var}(A)_p = \left\langle \frac{p^2 A^2}{q^2} \right\rangle_q - \langle A^2 \rangle_p = \left\langle \left(\frac{p}{q} - 1 \right) A^2 \right\rangle_p. \quad (6.148)$$

Por ejemplo, si es posible elegir q proporcional a pA o casi, la variable pA/q será casi constante y su varianza será muy pequeña. Sin embargo ese caso es de interés académico; en la práctica un cálculo MC se adapta a p y no a un observable concreto (el cálculo MC es costoso y se utiliza un mismo muestreo para calcular el valor esperado de múltiples observables a la vez). Generalmente el cálculo con reweighting tiende a empeorar, no mejorar, la dispersión. En un caso claro, si $A = 1$, el cálculo directo de $\langle A \rangle_p$ no tiene ninguna dispersión (cada $\mathbf{X} \sim p$ da el mismo valor $A = 1$), mientras que el cálculo con $A' \equiv pA/q = p/q$ y $\mathbf{X} \sim q$, sí tendrá una varianza que puede ser grande.

Ejemplo. Supongamos que

$$p(\mathbf{x}) = (2\pi)^{-d/2} e^{-\mathbf{x}^2/2}, \quad q(\mathbf{x}) = (2\pi\sigma^2)^{-d/2} e^{-\mathbf{x}^2/2\sigma^2} \quad \text{en } \mathbb{R}^d. \quad (6.149)$$

Por lo que sabemos, $1 = \langle 1 \rangle_p = \langle p/q \rangle_q$, entonces

$$1 + \text{Var}(p/q)_q = \langle p^2/q^2 \rangle_q = \int d^d x \frac{p^2}{q} = \left(\frac{\sigma^4}{2\sigma^2 - 1} \right)^{d/2}, \quad \sigma^2 > \frac{1}{2}. \quad (6.150)$$

Cuando $\sigma^2 \leq 1/2$ (q demasiado poco extendida) la varianza diverge. Por otro lado, si $\sigma = 1$ la varianza se anula, ya que $q = p$. En los demás casos $\sigma^4/(2\sigma^2 - 1) > 1$ y en consecuencia la varianza crece exponencialmente con la dimensión d . Igual que veíamos en el caso de la esfera d -dimensional, a medida que las dimensiones son mayores el parecido entre las dos funciones (para un mismo valor de σ) va disminuyendo y eso hace aumentar la dispersión, tanto en reweighting como en aceptación-rechazo. De nuevo el método sólo será útil si q se satisface el requerimiento de muestreo relevante para w .

En el caso general de peso w no normalizado (es decir, no se conoce su normalización) hay un problema añadido de sesgo. En efecto, aunque el cociente indicado en ec. (6.146) proporciona correctamente $\langle A \rangle_w$, en realidad el valor esperado como tal nunca se calcula, sino sólo resultados de experimentos aleatorios para los que una variable aleatoria toma un valor que se acepta como

estimación del valor esperado. Para estimar el valor esperado en el numerador se genera una muestra $\mathbf{X}_1, \dots, \mathbf{X}_N$ distribuida según q . Para el denominador se puede usar la misma muestra o bien generar otra independiente. En el primer caso la estimación es

$$\bar{A}_1 = \frac{\sum_{i=1}^N w(\mathbf{X}_i)A(\mathbf{X}_i)/q(\mathbf{X}_i)}{\sum_{i=1}^N w(\mathbf{X}_i)/q(\mathbf{X}_i)}, \quad \mathbf{X}_i \sim q \quad (\text{independientes}) \quad (6.151)$$

cuando $N \rightarrow \infty$ $\bar{A}_1 \rightarrow \langle A \rangle_w$ con probabilidad 1 (ya que numerador y denominador lo hacen), sin embargo, para N finito hay un sesgo ya que $E(\bar{A}_1)$ no coincide con $E(A)$. El caso más claro es $N = 1$: en este caso $\bar{A}_1 = A(\mathbf{X}_1)$ con $\mathbf{X}_1 \sim q$, por tanto $E(\bar{A}_1) = \langle A \rangle_q$ y no $\langle A \rangle_w$. Esto es muy distinto del MC usual (se muestrea la distribución p) ya que ahí $E(\bar{A}) = E(A)$ independientemente del valor de N (otra cosa es que para $N = 1$ la dispersión pueda ser grande o no).

En el segundo caso la estimación es

$$\bar{A}_2 = \frac{\sum_{i=1}^N w(\mathbf{X}_i)A(\mathbf{X}_i)/q(\mathbf{X}_i)}{\sum_{i=1}^N w(\mathbf{Y}_i)/q(\mathbf{Y}_i)} \equiv \frac{\overline{wA/q}}{\overline{w/q}}, \quad \mathbf{X}_i, \mathbf{Y}_i \sim q \quad (\text{independientes}). \quad (6.152)$$

De nuevo tiende a $\langle A \rangle_w$ con probabilidad 1, pero

$$E(\bar{A}_2) = \langle wA/q \rangle_q \left\langle \left(\overline{w/q} \right)^{-1} \right\rangle_q. \quad (6.153)$$

Generalmente, $\langle X^{-1} \rangle \neq \langle X \rangle^{-1}$ (que es lo que haría falta, aplicado a $X = \overline{w/q}$ para concluir que $E(\bar{A}_2) = \langle A \rangle_p$). Por otro lado, aunque hay sesgo para N finito, éste no desempeña un papel relevante para N suficientemente grande, ya que el sesgo es menor que la propia dispersión en el estimador. En efecto, para $X = \overline{w/q}$, sea su media μ y su dispersión σ/\sqrt{N} , entonces,

$$\langle X^{-1} \rangle = \left\langle \frac{1}{\mu + X - \mu} \right\rangle = \left\langle \frac{1}{\mu} - \frac{x - \mu}{\mu^2} + \frac{(x - \mu)^2}{\mu^3} + \dots \right\rangle = \frac{1}{\mu} + \frac{1}{\mu^3} \frac{\sigma^2}{N} + \dots \quad (6.154)$$

por tanto el sesgo es

$$\langle X^{-1} \rangle - \langle X \rangle^{-1} = O(N^{-1}), \quad (6.155)$$

mientras que la dispersión es $O(N^{-1/2})$. A medida que N crece el sesgo disminuye más deprisa que las fluctuaciones y eventualmente queda tapado por éstas.

El principal problema del método de reweighting es que la dispersión en numerador y denominador (si lo hay) crece muy rápidamente a medida que el número de dimensiones aumenta: para d grande p y q son muy distintas en general, el muestreo produce la inmensa mayoría de puntos donde q es

grande (no p) y luego le asigna un peso p/q despreciable. Todo el peso viene de los raros casos en los que el punto cae donde p importa. En consecuencia, el número efectivo de puntos que de verdad intervienen en los promedios es muy pequeño por lo que las fluctuaciones son muy grandes. En casos prácticos, el problema crece exponencialmente con el número de variables involucradas.

Puesto que hay métodos eficientes markovianos para hacer un muestreo de cualquier $p(\mathbf{x})$, el interés de usar un $q(\mathbf{x})$ auxiliar, es principalmente poder *reutilizar* configuraciones $\{\mathbf{X}_i\}_{i=1}^N$, $\mathbf{X} \sim q$ que ya se tienen, de modo que si $p(\mathbf{x})$ depende de uno o más parámetros λ , $p(\mathbf{x}, \lambda)$, se pueden calcular estimaciones $\langle f(\mathbf{x}) \rangle_\lambda$ para varios λ a la vez, con un mismo cálculo MC. Según el caso, incluso se pueden obtener curvas enteras en función de λ , lo cual es imposible haciendo una simulación MC para cada λ requerida.

Otra aplicación es al caso de w complejo (o con partes negativas), en lugar de real y positivo en todos los puntos. Aunque los promedios están bien definidos cuando w es complejo, el muestreo $\mathbf{X} \sim w$ tal cual no tiene sentido en este caso.⁷⁷ Sin embargo, el método de reweighting se puede aplicar sin problemas, eligiendo $q(\mathbf{x})$ positivo.

6.5. Métodos markovianos

Para hacer muestreos eficientes de pesos w multidimensionales generales se suelen usar métodos de tipo *markoviano*. El sistema pasa por una secuencia de estados \mathbf{X} parametrizados por un parámetro t que es el **tiempo ficticio** o **de simulación**. La idea es generar un **caminante** que describe un camino $\mathbf{X}(t)$ que recorra la región a muestrear $\Omega \subseteq \mathbb{R}^n$ de manera ergódica, es decir, que a la larga el tiempo que el caminante pasa en una región $A \subseteq \Omega$ cualquiera sea proporcional a $\text{Prob}(A) = \int_A d^n \rho(\mathbf{x})$, donde $\rho(\mathbf{x})$ denota el peso normalizado que se quiere muestrear. Se cambia el promedio sobre \mathbf{x} por un promedio temporal:

$$\langle A \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt A(\mathbf{X}(t)) \quad (\text{con probabilidad } 1). \quad (6.156)$$

En la mayor parte de las implementaciones el tiempo es una variable discreta, k , y el caminante pasa sucesivamente por puntos $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k, \dots$. El adjetivo **markoviano** indica que la probabilidad de $\mathbf{X}_k \rightarrow \mathbf{X}_{k+1}$ no depende de la historia anterior, es decir, depende sólo del valor de \mathbf{X}_k y no de los \mathbf{X}_j , $j < k$:

$$\text{Prob}(\mathbf{X}_{k+1}, \mathbf{X}_{k+2}, \dots | \mathbf{X}_k, \mathbf{X}_{k-1}, \dots) = \text{Prob}(\mathbf{X}_{k+1}, \mathbf{X}_{k+2}, \dots | \mathbf{X}_k). \quad (6.157)$$

⁷⁷Mediante extensión analítica se puede construir un auténtico muestreo MC de $w(\mathbf{x})$ genéricos que proporciona correctamente los valores esperados [14], sin embargo el método es costoso.

Los procesos markovianos son los procesos estocásticos más sencillos y mejor estudiados.

El salto $\mathbf{X}_k \rightarrow \mathbf{X}_{k+1}$ se rige por cierta regla aleatoria prefijada⁷⁸

$$\text{Prob}(\mathbf{X}_{k+1} = \mathbf{y} | \mathbf{X}_k = \mathbf{x}) = W_k(\mathbf{y} | \mathbf{x}). \quad (6.158)$$

$W_k(\mathbf{y} | \mathbf{x})$ es la (densidad de) probabilidad de saltar a \mathbf{y} en tiempo $k+1$ si en tiempo k el caminante estaba en \mathbf{x} y evidentemente W_k debe estar normalizada y ser no negativa

$$\int d^n y W_k(\mathbf{y} | \mathbf{x}) = 1, \quad W_k(\mathbf{y} | \mathbf{x}) \geq 0. \quad (6.159)$$

A menudo la función W_k no depende de k .

En cada momento k la distribución de caminantes vendrá dada por una cierta densidad $\rho_k(\mathbf{x})$. Al aplicar el salto markoviano, esta probabilidad se actualiza a $\rho_{k+1}(\mathbf{x})$:

$$\rho_{k+1}(\mathbf{y}) = \int W_k(\mathbf{y} | \mathbf{x}) \rho_k(\mathbf{x}) d^n x. \quad (6.160)$$

Eligiendo W_k adecuadamente se consigue que a la larga (esto es $k \rightarrow \infty$), $\mathbf{X}_k \sim \rho$. Es decir,

$$\lim_{k \rightarrow \infty} \rho_k(\mathbf{x}) = \rho(\mathbf{x}) \quad (6.161)$$

en el sentido de valores esperados (convergencia débil).

Generalmente, se empieza con un caminante generado con una distribución ρ_1 que no se parece a ρ , por lo que hay que esperar un cierto número de pasos τ_0 hasta que la distribución se termalice a la de equilibrio, ρ . Después de este periodo transitorio se tiene $\mathbf{X}_{\tau_0} \sim \rho$. Los siguientes puntos también están distribuidos según ρ pero no son independientes de \mathbf{X}_{τ_0} (el caminante no se han olvidado de esa posición) hasta pasado un cierto número de pasos τ , denominado **tiempo de autocorrelación**. Ese tiempo hay que dejarlo entre cada dos medidas para que los valores sean independientes.⁷⁹ De este modo

$$\langle A \rangle = \frac{1}{N} \sum_{j=1}^N A(\mathbf{X}_{\tau_0 + \tau j}) \pm \frac{\sigma_A}{\sqrt{N}}. \quad (6.162)$$

6.5.1. Condición de balance detallado

Para que a la larga el proceso markoviano muestree una ρ dada, hay dos condiciones claramente necesarias:

⁷⁸Cuando \mathbf{X} es una variable continua $W_k(\mathbf{y} | \mathbf{x})$ es una *densidad de probabilidad* respecto de la variable \mathbf{y} .

⁷⁹Usar N puntos seguidos no introduce un sesgo en el resultado (es decir, no modifica el valor esperado) pero como sólo N/τ puntos son realmente independientes el error va como $\sqrt{\tau/N}$ en vez de $1/\sqrt{N}$.

- a) La cadena markoviana ha de ser **irreducible**. Esto es, la función $W_k(\mathbf{y}|\mathbf{x})$ debe ser tal que en principio el caminante pueda ir (enlazando saltos) desde cualquier punto del soporte de $w(\mathbf{x})$ a cualquier otro.
- b) Si la distribución ya ha alcanzado el equilibrio, al aplicar un salto markoviano esta propiedad se debe mantener:

$$\int W_k(\mathbf{y}|\mathbf{x})\rho(\mathbf{x})d^n x = \rho(\mathbf{y}). \quad (6.163)$$

Notablemente si la cadena markoviana es *aperiódica* estas dos condiciones también son suficientes para que se satisfaga (6.161) [6].⁸⁰

Es fácil comprobar que cada paso markoviano acerca ρ_k a ρ :

$$\begin{aligned} \|\rho_{k+1} - \rho\| &:= \int d^n y |\rho_{k+1}(\mathbf{y}) - \rho(\mathbf{y})| = \int d^n y \left| \int d^n x W_k(\mathbf{y}|\mathbf{x})(\rho_k(\mathbf{x}) - \rho(\mathbf{x})) \right| \\ &\leq \int d^n y \int d^n x W_k(\mathbf{y}|\mathbf{x}) |\rho_k(\mathbf{x}) - \rho(\mathbf{x})| = \int d^n x |\rho_k(\mathbf{x}) - \rho(\mathbf{x})| \\ &= \|\rho_k - \rho\|. \end{aligned} \quad (6.164)$$

Pero por sí sólo esto no garantiza $\|\rho_k - \rho\| \rightarrow 0$.

Una forma práctica de garantizar que W_k deja invariante ρ es imponer la **condición de balance detallado**

$$W_k(\mathbf{y}|\mathbf{x})\rho(\mathbf{x}) = W_k(\mathbf{x}|\mathbf{y})\rho(\mathbf{y}), \quad (6.165)$$

que implica (6.163) usando (6.159). La misma ecuación se puede escribir usando $w(\mathbf{x})$ (el peso no normalizado) y equivale a decir que la función

$$S_k(\mathbf{y}, \mathbf{x}) \equiv W_k(\mathbf{y}|\mathbf{x})\rho(\mathbf{x}), \quad S_k(\mathbf{y}, \mathbf{x}) = S_k(\mathbf{x}, \mathbf{y}), \quad (6.166)$$

es una función simétrica.

Se puede dar un argumento de porqué la condición de balance detallado lleva a la distribución de equilibrio. Supongamos que tenemos una colectividad de caminantes distribuidos según ρ . Si el balance es detallado, los caminantes que saltan de \mathbf{x} a \mathbf{y} se compensan con los que saltan de \mathbf{y} a \mathbf{x} . (En un balance no necesariamente detallado la distribución se mantiene en equilibrio porque

⁸⁰Una condición suficiente para que una cadena markoviana sea aperiódica es que $W_k(\mathbf{x}|\mathbf{x}) > 0$ en alguna región de Ω .

los caminantes intercambian posiciones en cada paso, unos con otros pero no necesariamente por pares.) Si la distribución no está en equilibrio, sea $N(\mathbf{x})$ el número de caminantes en \mathbf{x} , ídem $N(\mathbf{y})$. Después de un salto

$$\begin{aligned}
 N(\mathbf{y} \leftarrow \mathbf{x}) - N(\mathbf{x} \leftarrow \mathbf{y}) &= W(\mathbf{y}|\mathbf{x})N(\mathbf{x}) - W(\mathbf{x}|\mathbf{y})N(\mathbf{y}) \\
 &= N(\mathbf{x})W(\mathbf{x}|\mathbf{y}) \left(\frac{W(\mathbf{y}|\mathbf{x})}{W(\mathbf{x}|\mathbf{y})} - \frac{N(\mathbf{y})}{N(\mathbf{x})} \right) \\
 &= N(\mathbf{x})W(\mathbf{x}|\mathbf{y}) \left(\frac{\rho(\mathbf{y})}{\rho(\mathbf{x})} - \frac{N(\mathbf{y})}{N(\mathbf{x})} \right).
 \end{aligned} \tag{6.167}$$

Por tanto, si $\frac{N(\mathbf{y})}{N(\mathbf{x})} \geq \frac{\rho(\mathbf{y})}{\rho(\mathbf{x})}$ se tendrá $N(\mathbf{x} \leftarrow \mathbf{y}) \geq N(\mathbf{y} \leftarrow \mathbf{x})$. El flujo va de donde hay más caminantes de los que debería a donde hay menos de modo que el equilibrio tiende a restablecerse. Cuando $\mathbf{X} \sim \rho$, $N(\mathbf{x}) \propto \rho(\mathbf{x})$ y $N(\mathbf{x} \leftarrow \mathbf{y}) = N(\mathbf{y} \leftarrow \mathbf{x})$ por lo que el equilibrio se mantiene.

6.5.2. Algoritmo de Metropolis

El **algoritmo de Metropolis** fue inventado por Rosenbluth para aplicarlo a la distribución de Boltzmann y extendido a una distribución general por Hastings. Tiene la virtud de que es muy flexible ya que no requiere propiedades específicas de $w(\mathbf{x})$, sólo saber calcular esta función en cada punto que se pida (de hecho basta conocer el cociente de esta función en dos puntos cualesquiera).

Para aplicar el método se necesita una densidad de probabilidad *auxiliar* $Q_k(z|\mathbf{x})$ (no negativa y normalizada respecto de z). Esta probabilidad dice cómo proponer un *candidato* z a la nueva posición del caminante cuando éste se encuentra en \mathbf{x} en tiempo k . No debe confundirse $Q_k(z|\mathbf{x})$ con la función $W_k(\mathbf{y}|\mathbf{x})$. Esta última se construye indirectamente mediante el algoritmo de Metropolis y no se necesita en forma explícita.

Dadas la funciones $Q_k(z|\mathbf{x})$ y $w(\mathbf{x})$, el algoritmo de Metropolis para actualizar el valor de \mathbf{X}_k al siguiente valor en la cadena markoviana, \mathbf{X}_{k+1} , es:

- 1) Dado \mathbf{X}_k se genera una propuesta \mathbf{Z} con $Q_k(\mathbf{Z}|\mathbf{X}_k)$, y se genera un número aleatorio $U \sim U(0, 1)$.
- 2) Se actualiza la posición de acuerdo con la regla

$$\mathbf{X}_{k+1} = \begin{cases} \mathbf{Z}, & \text{si } w(\mathbf{Z}) \geq U w(\mathbf{X}_k) \quad (\text{la propuesta es aceptada}) \\ \mathbf{X}_k, & \text{si } w(\mathbf{Z}) < U w(\mathbf{X}_k) \quad (\text{la propuesta es rechazada}) \end{cases} \tag{6.168}$$

Efectivamente la probabilidad de aceptación q es

$$q = \min \left(1, \frac{w(\mathbf{Z})}{w(\mathbf{X}_k)} \right), \quad (6.169)$$

Equivalentemente se puede decir que si $w(\mathbf{Z}) \geq w(\mathbf{X}_k)$ (si \mathbf{Z} es “más probable” que \mathbf{X}_k) el nuevo punto se acepta inmediatamente, en caso contrario se acepta con probabilidad $w(\mathbf{Z})/w(\mathbf{X}_k)$.⁸¹ Por supuesto, “más probable” es una forma de hablar, ya que la relación $w(\mathbf{Z}) \geq w(\mathbf{X})$, para los mismos puntos \mathbf{Z} y \mathbf{X} depende del sistema de coordenadas.

La función $Q_k(z|x)$ se puede elegir de muchas formas con tal de que satisfaga las dos condiciones siguientes:

- a) (*irreducibilidad*) que la función $Q_k(x|z)$ sea tal que mediante saltos sucesivos se pueda llegar a cualquier punto del soporte de la función $w(x)$, y
- b) (*simetría*) que la función $Q_k(x|z)$ sea simétrica

$$Q_k(z|x) = Q_k(x|z). \quad (6.170)$$

Es decir, la probabilidad de proponer \mathbf{Z} estando en \mathbf{X} debe ser igual a la de proponer \mathbf{X} estando en \mathbf{Z} .

Algunos casos patológicos quedan eliminados si se añade la condición de que $Q(x|x) > 0$ en alguna región, de modo que la cadena markoviana sea aperiódica.

El método descrito es el algoritmo de Metropolis usado más frecuentemente. Una generalización se obtiene no requiriendo que la función $Q_k(z|x)$ sea simétrica, lo que da lugar al algoritmo de **Metropolis-Hastings**. En este caso la probabilidad de aceptación pasa a ser

$$q = \min \left(1, \frac{Q_k(\mathbf{X}_k|\mathbf{Z}) w(\mathbf{Z})}{Q_k(\mathbf{Z}|\mathbf{X}_k) w(\mathbf{X}_k)} \right), \quad (6.171)$$

es decir, se acepta sii $U < q$ (con $U \sim U(0,1)$). Esta versión es covariante bajo cambios de coordenadas (siendo w una densidad y Q_k una densidad respecto de la primera variable y un escalar

⁸¹Es importante enfatizar que cuando el candidato es rechazado, el nuevo punto \mathbf{X}_{k+1} coincide con \mathbf{X}_k . Esto *no es lo mismo* que volver a generar nuevos candidatos hasta que uno sea aceptado y tomarlo como el nuevo \mathbf{X}_{k+1} (que es lo que se haría en aceptación-rechazo). Esa otra prescripción no produce una cadena markoviana con $\mathbf{X} \sim \rho$, sino que se termaliza a otra distribución de equilibrio ρ' que depende de la elección de Q_k , y por tanto produce valores esperados incorrectos.

respecto de la segunda variable, bajo cambios de coordenadas). Obviamente este método incluye el de Q_k simétrica como caso particular.

Vamos a hacer la demostración de que el método de Metropolis-Hastings satisface balance detallado usando una notación discreta, que es interesante por sí misma. Notación discreta quiere decir que denotamos las posibles opciones como i, j, \dots , el peso del caso i -ésimo como w_i y su probabilidad (en equilibrio) como ρ_i :

$$w_i = \mathcal{N} \rho_i, \quad \sum_i \rho_i = 1, \quad \rho_i \geq 0. \quad (6.172)$$

Denotamos la probabilidad de transición $j \rightarrow i$ en el proceso markoviano como $W_{ij}^{(k)}$, o simplemente W_{ij} dejando implícita la etiqueta del paso k -ésimo. La condición de equilibrio es

$$\sum_j W_{ij} \rho_j = \rho_i, \quad (6.173)$$

y la condición de balance detallado es

$$W_{ij} \rho_j = W_{ji} \rho_i \quad (\text{no hay suma de índices repetidos}), \quad (6.174)$$

o equivalentemente $W_{ij} w_j = W_{ji} w_i$. Igualmente Q_{ij} denota la probabilidad de proponer un candidato i si el estado actual es j . La probabilidad de aceptación será

$$q = \min \left(1, \frac{Q_{ji} w_i}{Q_{ij} w_j} \right). \quad (6.175)$$

Puesto que la condición de balance detallado es trivial si $i = j$, sólo consideramos el caso $i \neq j$ (el caminante cambia de sitio) y esto sólo ocurre si el candidato es aceptado. La probabilidad de $j \rightarrow i$ con aceptación del candidato es $q Q_{ij}$ (Q_{ij} es la probabilidad de proponer i y q la de ser aceptado si es propuesto),

$$W_{ij} = q Q_{ij} = \min \left(1, \frac{Q_{ji} w_i}{Q_{ij} w_j} \right) Q_{ij} \quad (i \neq j). \quad (6.176)$$

Se deduce que (dado que Q_{ij} , w_j son positivos)

$$W_{ij} w_j = \min(Q_{ij} w_j, Q_{ji} w_i) \quad (i \neq j), \quad (6.177)$$

que es manifiestamente simétrico respecto de ij , de acuerdo con la condición de balance detallado.

Aunque no se necesite explícitamente, nótese que la probabilidad de transición $j \rightarrow i$ se puede expresar como

$$W_{ij} = \sum_k Q_{kj} \int_0^1 du \left(\Theta(Q_{jk} w_k - u Q_{kj} w_j) \delta_{ik} + \Theta(u Q_{kj} w_j - Q_{jk} w_k) \delta_{ij} \right), \quad (6.178)$$

donde el primer sumando corresponde a aceptación (del candidato k) y el segundo a rechazo. Integrando u y sumando k en el término de aceptación se obtiene qQ_{ij} . El término de rechazo no contribuye si $i \neq j$.

Volviendo a Metropolis puro, un ejemplo de función $Q_k(z|x)$ válida es un salto aleatorio uniformemente distribuido dentro de una bola de radio R ,⁸²

$$Q_k(z|x) \propto \Theta(R - \|z - x\|). \quad (6.179)$$

La probabilidad de un salto $\xi = z - x$ es independiente de x y los saltos ξ y $-\xi$ son equiprobables. Otra elección válida es un salto gaussiano de tamaño R a elegir. Con cualquier elección, la función Q_k producirá un cierto salto típico $\|Z - X\| \sim R$. Aunque teóricamente el método converge en todo caso, si el valor de R se toma demasiado pequeño, y el caminante está lejos de la zona relevante, los candidatos serán aceptados pero el caminante tardará mucho en recorrer la distribución para poder muestrearla. Si por el contrario R es demasiado grande y el caminante ya está en la zona relevante, los candidatos serán casi siempre rechazados y el caminante se moverá poco, lo que también dificulta que explore la función $w(x)$. Más importante, en ambos casos el tiempo de autocorrelación será grande. Para la eficiencia del método es necesario que R se ajuste de modo que la proporción de candidatos aceptados esté lejos de 0% o 100%. Un valor típico es ajustar la proporción de aceptación/rechazo al 50% (o a un valor más adecuado haciendo las pruebas necesarias para reducir el tiempo de autocorrelación).

En la práctica, cuando hay muchas variables, lo que suele hacerse es proponer un salto Metrópolis para una de las coordenadas, manteniendo las demás fijas, a continuación se toca otra coordenada, y así sucesivamente hasta hacer un **barrido** completo sobre todas las coordenadas en \mathbf{X} . La elección de coordenada a actualizar puede hacerse de modo ordenado o bien de modo aleatorio, sorteando entre todas las coordenadas.

Como ejemplo de Metropolis, supongamos una red unidimensional con nodos $n = 1, 2, \dots, N$ en cada uno de los cuales hay un grado de libertad (una variable) real φ_n . La energía de la configuración $\varphi = (\varphi_1, \dots, \varphi_N)$ es

$$E(\varphi) = \sum_{n=1}^N \left(\frac{1}{2} (\varphi_{n+1} - \varphi_n)^2 + V(\varphi_n) \right). \quad (6.180)$$

Aquí suponemos condiciones de contorno periódicas: $\varphi_{N+1} \equiv \varphi_1$, $\varphi_0 \equiv \varphi_N$. El peso de la configuración φ a temperatura T es

$$w(\varphi) = e^{-\beta E(\varphi)}, \quad \rho(\varphi) = \frac{1}{Z} w(\varphi), \quad Z = \int d^n \varphi w(\varphi), \quad \beta \equiv \frac{1}{kT}. \quad (6.181)$$

⁸²Sería absurdo usar exactamente este método, más cómodo es generar z dentro de un hipercubo, con un efecto similar.

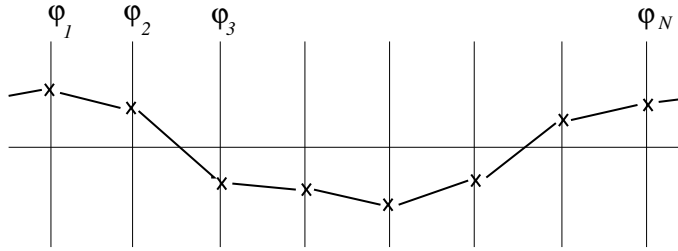


Figura 14: Una configuración de φ . Por las condiciones periódicas $\varphi_{N+1} \equiv \varphi_1$. Las configuraciones con grandes cambios en φ de un nodo al siguiente tienen más energía y por tanto son menos probables. Igualmente valores de φ_k alejados del mínimo del potencial también están suprimidos.

Z no es conocido pero no se necesita para hacer el paseo markoviano. Para aplicar Metropolis, no es adecuado mover φ mucho ya que (si N es grande) el movimiento será rechazado con mucha probabilidad. Lo usual es mover sólo uno de los φ_n para un n dado, luego actualizar otro, y así sucesivamente hasta completar un barrido de la red. Se aplica el número de barridos necesarios para obtener un número suficiente de configuraciones φ independientes termalizadas al peso w .

Cada vez que se toca un n se hace una propuesta de candidato φ'_n que se acepta o se rechaza. Por ejemplo $\varphi'_n = \varphi_n + \xi$ donde ξ es una variable aleatoria independiente de φ_n y con una distribución simétrica al cambiar ξ por $-\xi$ (esto garantiza que las probabilidades de $\varphi_n \rightarrow \varphi'_n$ y $\varphi'_n \rightarrow \varphi_n$ sean iguales),

La probabilidad de aceptación requiere el cálculo del cociente

$$q \equiv \frac{w(\varphi_1, \dots, \varphi'_n, \dots, \varphi_N)}{w(\varphi_1, \dots, \varphi_n, \dots, \varphi_N)}. \quad (6.182)$$

Se genera $U \sim U(0, 1)$ y el candidato se acepta si $U < q$ (obviamente si $q > 1$ no hace falta generar U).

Para un peso general el cálculo de q puede ser costoso (especialmente si N es grande), pero en nuestro caso la energía es local: cada φ_n sólo está conectada con $\varphi_{n\pm 1}$ (interacción de vecinos próximos), por tanto las variables no acopladas a φ_n se cancelan en numerador y denominador:

$$q = \frac{e^{-\beta(V(\varphi'_n) + \frac{1}{2}(\varphi_{n+1} - \varphi'_n)^2 + \frac{1}{2}(\varphi_{n-1} - \varphi'_n)^2)}}{e^{-\beta(V(\varphi_n) + \frac{1}{2}(\varphi_{n+1} - \varphi_n)^2 + \frac{1}{2}(\varphi_{n-1} - \varphi_n)^2)}} = e^{-\beta\xi(2\varphi_n - \varphi_{n+1} - \varphi_{n-1} + \xi)} e^{-\beta(V(\varphi_n + \xi) - V(\varphi_n))}. \quad (6.183)$$

Que la interacción sólo involucre nodos próximos entre sí (lo cual permite una gran simplificación)

es la regla más que la excepción, ya que no es natural que haya *acción a distancia* entre los grados de libertad de un sistema físico.

6.5.3. Baño térmico

También denominado **muestreo de Gibbs**. Es un proceso markoviano en el que en cada paso se elige un subconjunto de coordenadas de \mathbf{X} y se actualizan sólo esas coordenadas de acuerdo con su probabilidad condicionada, manteniendo las otras variables congeladas. Posteriormente se toma otro subconjunto y se procede igual. Los subconjuntos se pueden tomar de forma ordenada o bien aleatoria, con tal de que eventualmente se pase por todas las coordenadas. No se requiere conocer la normalización absoluta del peso.

Por ejemplo, si tenemos $\rho(x_1, \dots, x_d)$, para actualizar la coordenada n -ésima se genera y_n según la probabilidad condicionada $\rho(y_n | x_1, \dots, \hat{x}_n, \dots, x_d)$ y se actualiza $x_n \rightarrow y_n$. Es fácil ver que este procedimiento satisface balance detallado: para simplificar suponemos $d = 2$ y actualizamos la primera variable,

$$W(y_1, y_2 | x_1, x_2) \rho(x_1, x_2) = \rho(y_1 | x_2) \delta(y_2 - x_2) \rho(x_1, x_2) = \delta(y_2 - x_2) \frac{\rho(y_1, x_2) \rho(x_1, x_2)}{\rho(x_2)} \quad (6.184)$$

que es una función simétrica al cambiar x_i con y_i .

El baño térmico se aplica casi exclusivamente para variables sueltas o grupos de variables independientes entre sí, y es eficiente cuando es posible generar y_n de modo económico. Nótese que el nuevo valor de la coordenada x_n es completamente independiente del valor anterior, aunque sí depende del valor de las otras coordenadas: x_n se ha *termalizado* al valor de las otras coordenadas que forman el baño.

El baño térmico equivale a aplicar multi-hit en Metropolis, es decir, si se aplica Metropolis muchas veces a una misma coordenada antes de pasar a la siguiente. La variable queda distribuida según la probabilidad condicionada relativa a las demás variables.⁸³ Por ello el baño térmico termaliza la distribución en menos pasos que Metropolis, pero puede que cada paso sea demasiado costoso y no compense. En este sentido la aplicación de baño térmico es más limitada (hay menos problemas a los que puede aplicarse) que Metropolis.

⁸³El número de intentos no debe depender de los resultados individuales, ya que en otro caso se podría introducir un sesgo.

Como ilustración, en el ejemplo anterior, cuando $V(\varphi) = \frac{1}{2}k\varphi^2$, la probabilidad condicionada es una gaussiana y el muestreo de Gibbs es aplicable de modo eficiente.

Nota: La actualización (tanto en Metropolis como en el baño térmico) debe ser siempre secuencial: si las dos variables a actualizar x_1 y x_2 están acopladas (no son mutuamente independientes) debe actualizarse primero una de las dos y luego la otra (no importa el orden). Pero sería incorrecto generar simultáneamente x'_1 y x'_2 usando los valores actuales de x_1 y x_2 . Lo correcto es generar digamos $x_1 \rightarrow x'_1$ usando (x_1, x_2) y a continuación $x_2 \rightarrow x'_2$ usando (x'_1, x_2) . Si, como es frecuente, las variables se dividen en pares e impares, de modo que las pares sólo interaccionan con las impares y viceversa (por ejemplo ec. (6.180)), se pueden actualizar todas las variables pares a la vez y luego todas las impares, en cada barrido.

6.5.4. Método de búsqueda de mínimos por enfriamiento

Es una técnica para obtener máximos o mínimos de una función que es útil cuando el número de variables es grande. Por ejemplo, queremos minimizar una energía $E(\varphi)$, $\varphi \in \mathbb{R}^d$. Si se usa el método de máxima pendiente la búsqueda puede acabar en un mínimo local, en cuya cuenca esté el valor inicial.

El método MC se aplica generando un paseo markoviano (por ejemplo Metropolis) con peso $w(\varphi) = \exp(-\beta E(\varphi))$, para $\beta = 1/kT > 0$. Cuando $\beta \rightarrow \infty$ ($T \rightarrow 0$, enfriamiento), w selecciona los valores mínimos de $E(\varphi)$. Para que el proceso no se quede estancado en un mínimo local se usa T finita junto con el método markoviano y luego se baja lentamente la temperatura. Los teoremas de procesos markovianos aseguran que para tiempos de simulación suficientemente grandes el caminante explora adecuadamente todas las zonas relevantes.

6.5.5. Termalización y tiempo de autocorrelación

Aunque los métodos markovianos son muy flexibles, tienen el inconveniente de que la variable \mathbf{X}_k sólo está distribuida según la distribución de equilibrio $w(\mathbf{x})$ en el límite $k \rightarrow \infty$. Para k finito la distribución $\rho_k(\mathbf{x})$ tiene un cierto sesgo y depende del punto inicial \mathbf{X}_0 así como del algoritmo markoviano concreto. Un segundo problema es que no todos los \mathbf{X}_k son independientes.

El acercamiento al equilibrio (**termalización**) desde una configuración inicial cualquiera es típicamente de tipo exponencial. Se debe dejar un número τ_0 de pasos suficiente de modo que $\mathbf{X}_k \sim w$, para $k > \tau_0$. Aunque la termalización nunca es perfecta puede llegarse a un punto en el que el sesgo sea despreciable comparado con las fluctuaciones.

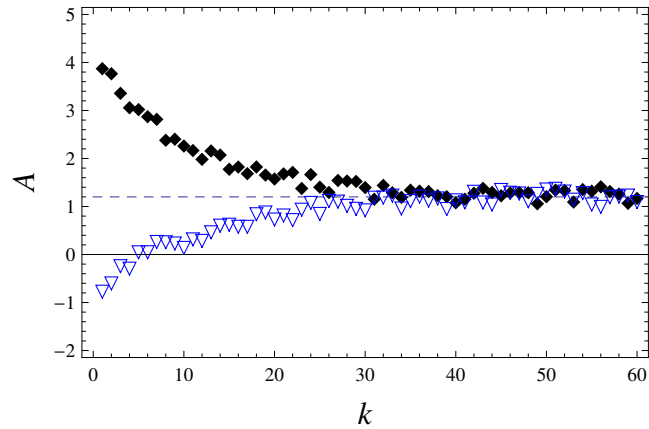


Figura 15: Criterio de termalización basado en un observable $A(\mathbf{X}_k)$ empezando con configuraciones iniciales bien separadas. Se puede considerar que hay termalización para $k \geq \tau_0 = 35$.

Un método usado frecuentemente es considerar varias simulaciones con muy distintos puntos de partida \mathbf{X}_0 y esperar hasta que los valores esperados en las distintas simulaciones converjan (dentro de las fluctuaciones). Las configuraciones iniciales pueden ser por, ejemplo, con variables fijas a un valor común (inicio frío) y con variables tomando valores aleatorios (inicio caliente). No hay un criterio universalmente aceptado para reconocer cuando ha transcurrido un tiempo de termalización τ_0 suficiente, en este sentido es preferible ser conservador para evitar confundir el equilibrio con un estado metaestable.

Una vez alcanzada la termalización, todos los valores sucesivos \mathbf{X}_k , con $k > \tau_0$ están distribuidos según la distribución de equilibrio w . El problema es que \mathbf{X}_k condiciona los valores posteriores en la cadena markoviana: \mathbf{X}_k y los valores posteriores \mathbf{X}_m , $m = k + 1, \dots, k + \tau$ no son variables aleatorias independientes hasta que τ tiene un valor suficientemente grande, denominado **tiempo de autocorrelación**.

El estimador⁸⁴

$$\bar{f} = \frac{1}{N} \sum_{k=1}^N f_k, \quad f_k \equiv f(\mathbf{X}_k) \tag{6.185}$$

no está sesgado (su valor esperado coincide con $\langle f \rangle$) ya que $\mathbf{X}_k \sim w$. Igualmente la varianza de la muestra, $S_{\bar{f}}^2$, de los f_k es un estimador no sesgado de *la varianza de f* . Pero la cuestión es cuál

⁸⁴Par simplificar la notación, después de alcanzar la termalización, desechamos las primeras τ_0 configuraciones y ponemos $k = 1$ otra vez.

es la *varianza de \bar{f}* . Para N variables independientes la dispersión en la media de la muestra sería σ_f/\sqrt{N} , que se podría estimar como S_f/\sqrt{N} cuando la muestra es suficientemente grande. En nuestro caso, sólo uno de cada τ puntos \mathbf{X}_k es realmente independiente (siendo τ el tiempo de autocorrelación). Por tanto el número de puntos independientes es $N_{\text{ef}} = N/\tau$, y la dispersión de \bar{f} será aproximadamente $S_f/\sqrt{N_{\text{ef}}}$. Es decir,

$$\langle f \rangle = \bar{f} \pm \sqrt{\tau} \frac{S_f}{\sqrt{N}} \quad (\text{para } N \text{ suficientemente grande}). \quad (6.186)$$

Hay que notar que $\text{Var}(f)$, al igual que $\langle f \rangle$, depende sólo de $f(\mathbf{X})$ y de la distribución w , pero no del método de muestreo utilizado. Por la ley de los grandes números, eventualmente S_f^2 será (probablemente) una estimación aceptable de $\text{Var}(f)$. Por tanto, para reducir la dispersión, los detalles del método markoviano se deben disponer de modo que τ sea *lo menor posible*. En principio el tiempo de autocorrelación no depende del observable $f(\mathbf{X})$ concreto (para observables genéricos), sino que es una propiedad de la cadena markoviana \mathbf{X}_k , $k = 1, 2, \dots$

Ejemplo. Si el proceso markoviano consiste en $W(\mathbf{y}|\mathbf{x}) = \rho(\mathbf{y})$, siendo ρ la distribución de equilibrio, cada nuevo punto generado está distribuido de acuerdo con ρ y es independiente del valor anterior en la cadena. Por tanto $\tau = 1$ en este caso. Ahora modificamos el algoritmo de modo que en uno de cada τ pasos $W(\mathbf{y}|\mathbf{x}) = \rho(\mathbf{y})$ pero para los otros $\tau - 1$ pasos $W(\mathbf{y}|\mathbf{x}) = \delta(\mathbf{y} - \mathbf{x})$ (el punto no se mueve). Como es fácil comprobar este algoritmo satisface balance detallado y es perfectamente válido. Simplemente hace que la cadena evolucione τ veces más despacio y el tiempo de autocorrelación es τ . El valor medio y la dispersión de $f(\mathbf{X})$ son los mismos que antes (caso $\tau = 1$) (ya que son los mismos puntos). Sin embargo, la dispersión de \bar{f} será $\sqrt{\tau}$ veces mayor para el mismo N :

$$\text{Var}(\bar{f}) = \text{Var}\left(\frac{\tau}{N}(f_\tau + f_{2\tau} + \dots + f_N)\right) = \left(\frac{\tau}{N}\right)^2 \frac{N}{\tau} \text{Var}(f) = \tau \frac{\text{Var}(f)}{N}. \quad (6.187)$$

Estimación del tiempo de autocorrelación mediante submuestras

Dada una cadena markoviana, un método usual para estimar la dispersión de \bar{f} (siendo $f(\mathbf{X})$ un observable cualquiera) es distribuir los N valores f_k en K bloques de valores sucesivos cada uno con un tamaño $L = N/K$. A partir de los bloques se construye una muestra de tamaño K haciendo corresponder a cada bloque su media

$$F_j = \frac{1}{L} \sum_{k=(j-1)L+1}^{jL} f_k, \quad j = 1, \dots, K. \quad (6.188)$$

Para la muestra $\{F_j\}_{j=1}^K$ podemos construir su media

$$\bar{F} = \frac{1}{K} \sum_{j=1}^K F_j = \bar{f}. \quad (6.189)$$

Por construcción la variable \bar{F} coincide con \bar{f} y por tanto tienen la misma media (a saber $\langle f \rangle$) y dispersión. La ventaja de usar $\{F_j\}_{j=1}^K$ es que si se toma el valor de L (tamaño de los bloques) suficientemente grande se tendrá que $L \gg \tau$ de modo que los distintos bloques serán independientes unos de otros, y por tanto los K valores F_j serán variables independientes, lo cual permite escribir

$$\langle f \rangle = \bar{F} \pm \frac{\sigma_F}{\sqrt{K}}. \quad (6.190)$$

Por otro lado, si el valor de K es suficientemente grande, la ley de los grandes números implica que podemos estimar la varianza de los F_j por la varianza de la muestra

$$\text{Var}(F) \approx S_F^2 = \frac{1}{K-1} \sum_{j=1}^K (F_j - \bar{F})^2 \quad (K \text{ suficientemente grande}) \quad (6.191)$$

Finalmente obtenemos, teniendo en cuenta que los F_j son independientes,

$$\langle f \rangle = \bar{F} \pm \frac{S_F}{\sqrt{K}}. \quad (6.192)$$

Comparando con la estimación naive que suponía que los N puntos eran independientes, $\langle f \rangle = \bar{f} \pm \frac{S_f}{\sqrt{N}}$, se deduce que el tiempo de autocorrelación es

$$\tau = \frac{S_F^2}{S_f^2} L. \quad (6.193)$$

Puede entonces comprobarse si L es adecuado verificando que $L \gg \tau$.

Puesto que la estimación del error requiere a la vez L y K suficientemente grandes, y $LK = N$, hace falta que N sea grande y tomar un compromiso en la distribución entre L y K . Lo que suele hacerse es usar valores crecientes de K (por ejemplo, $K = 2, 4, 8, \dots$) de modo que inicialmente L es grande ($L = N/2, N/4, N/8, \dots$) lo cual asegura $L \gg \tau$. Cuando K es suficiente para que se aplique la ley de los grandes números S_F^2/K se estabiliza a $\tau S_f^2/N$ (una magnitud independiente de L y K). De aquí se puede extraer el valor de τ (S_f^2 se obtiene de la muestra completa). El τ así determinado

se mantiene constante para K mayores (plateau) mientras $L \gg \tau$. Todos estos valores de K y L son aceptables y dan la misma estimación para el error en \bar{F} . Si se tomara K mayor, el τ estimado empezaría a disminuir respecto de su valor real, ya que $\frac{S_F^2}{S_f^2} L = 1$ cuando $K = N$, $L = 1$. Para esos K el error estaría subestimado ya que las fórmulas suponen que hay más F_j independientes de los que realmente hay.

Estimación del tiempo de autocorrelación por correladores

Un método más sofisticado y preciso (pero también más costoso) para la estimación de τ y el error se obtiene estudiando la correlación entre los datos. Definimos la función de **correlación** mediante

$$C_k = \text{Cov}(f_i, f_{i+k}) = C_{-k}, \quad C_0 = \text{Var}(f). \quad (6.194)$$

Nótese que una vez en el equilibrio el proceso markoviano es estacionario, por tanto $\text{Cov}(f_i, f_{i+k})$ no depende de i . Podemos ahora calcular la dispersión en \bar{f} a partir de la correlación

$$\text{Var}(\bar{f}) = \frac{1}{N^2} \text{Var}\left(\sum_{i=1}^N f_i\right) = \frac{1}{N^2} \sum_{i,j=1}^N \text{Cov}(f_i, f_j) = \frac{1}{N^2} \left(NC_0 + 2 \sum_{j=2}^N \sum_{i=1}^{j-1} C_{j-i} \right) \quad (6.195)$$

$$\sum_{j=2}^N \sum_{i=1}^{j-1} C_{j-i} = \sum_{j=2}^N \sum_{k=1}^{j-1} C_k = \sum_{k=1}^{N-1} (N-k) C_k \quad (6.196)$$

$$\text{Var}(\bar{f}) = \frac{1}{N} \left(C_0 + 2 \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) C_k \right). \quad (6.197)$$

Típicamente C_k cae exponencialmente, $C_k \approx C_0 e^{-2k/\tau}$, $k \gg 1$. En este caso, k/N puede despreciarse ya que los $k > \tau$ no contribuyen y $N \gg \tau$.

$$\text{Var}(\bar{f}) \approx \frac{1}{N} \left(C_0 + 2 \sum_{k=1}^{N-1} C_k \right) \quad (N \gg \tau). \quad (6.198)$$

Aquí se ve que la correlación entre puntos muestrales modifica la estimación usual $\text{Var}(\bar{f}) = \text{Var}(f)/N$, aumentando la dispersión ya que los C_k más importantes (k pequeño) son positivos, igual que C_0 .

Para ver la relación con el tiempo de autocorrelación, usamos $\tau \gg 1$ para aproximar la suma $\sum_{k=1}^{N-1} C_k$ por una integral, junto con la aproximación $C_k \approx C_0 e^{-2k/\tau}$:

$$\text{Var}(\bar{f}) \approx \frac{1}{N} C_0 (1 + \tau) \approx \tau \frac{\text{Var}(f)}{N}. \quad (6.199)$$

La relación en ec. (6.198) es esencialmente correcta ya que en la práctica $N \gg \tau$, sin embargo se refiere a los valores esperados exactos de la distribución. En realidad lo que se tiene son muestras obtenidas con el proceso markoviano, éstas se utilizan para estimar C_k ,

$$\bar{C}_k \equiv \frac{1}{N-k-1} \sum_{j=1}^{N-k} (f_j - \bar{f})(f_{j+k} - \bar{f}), \quad \bar{C}_0 = S_f^2. \quad (6.200)$$

Obtenemos una estimación de la varianza mediante

$$\text{Var}(\bar{f}) \approx \frac{1}{N} \left(\bar{C}_0 + 2 \sum_{k=1}^L \bar{C}_k \right). \quad (6.201)$$

Nótese que la suma sobre k no llega hasta $N-1$ sino hasta un cierto L a elegir. \bar{C}_k decrece exponencialmente cuando k aumenta pero no llega a hacerse arbitrariamente pequeño sino que a partir de un cierto $k=L$ toma valores pequeños pero fluctuantes. Se puede demostrar que la suma hasta $N-1$ de hecho no converge cuando $N \rightarrow \infty$. Debe cortarse la suma cuando C_k deja de decrecer para pasar a fluctuar alrededor de cero.

Este método es más preciso que el basado en dividir la muestra en bloques, pero el cálculo de \bar{C}_k puede ser muy costoso computacionalmente en situaciones realistas.

7. Problemas resueltos

7.1. Relación 1: Grupos de Lie

1. Enunciado:

Consideremos un espacio-tiempo relativista con una dimensión temporal (t) y otra espacial (x). Los eventos tienen coordenadas $(t, x) \in \mathbb{R}^2$. Las transformaciones de Lorentz en $1 + 1$ dimensiones son de la forma (usamos unidades $c = 1$)

$$\begin{pmatrix} t \\ x \end{pmatrix} \mapsto \begin{pmatrix} t' \\ x' \end{pmatrix} = B(v) \begin{pmatrix} t \\ x \end{pmatrix} \quad B(v) := \begin{pmatrix} \gamma & \gamma v \\ \gamma v & \gamma \end{pmatrix}, \quad \gamma(v) := \frac{1}{\sqrt{1-v^2}},$$

donde el parámetro v es la velocidad y $|v| < 1$.

- Demuéstrese que el conjunto $G = \{B(v), |v| < 1\}$ forma un grupo (el grupo de Lorentz en $1 + 1$ dimensiones).⁸⁵ Obténgase la ley de composición $v_{12} = f(v_1, v_2)$. Esta es la ley relativista de suma de velocidades. Muéstrese que el grupo es abeliano y conexo y no compacto. ¿Es isomorfo a algún grupo conocido?
- Las matrices $B(v)$ definen una representación bidimensional de G . Se pide el generador infinitesimal, K , asociado a la coordenada v en esa representación.
- Claramente v no es una coordenada canónica (¿por qué?). Obténgase la coordenada canónica $\xi = \xi(v)$ que coincide con v a nivel infinitesimal, es decir, $\xi(v) = v + O(v^2)$.
- Puesto que el grupo es abeliano sus irreps (complejas) son unidimensionales. Hallar la reducción de la representación bidimensional $B(v)$.

Solución:

- Multiplicamos las matrices para ver la ley de composición

$$B(v_1)B(v_2) = B(v_{12})$$

$$\begin{pmatrix} \gamma_1 & \gamma_1 v_1 \\ \gamma_1 v_1 & \gamma_1 \end{pmatrix} \begin{pmatrix} \gamma_2 & \gamma_2 v_2 \\ \gamma_2 v_2 & \gamma_2 \end{pmatrix} = \gamma_1 \gamma_2 \begin{pmatrix} 1 + v_1 v_2 & v_1 + v_2 \\ v_1 + v_2 & 1 + v_1 v_2 \end{pmatrix} = \begin{pmatrix} \gamma_{12} & \gamma_{12} v_{12} \\ \gamma_{12} v_{12} & \gamma_{12} \end{pmatrix}$$

⁸⁵Notar: 1) La propiedad asociativa es automática para producto de matrices y no hay que comprobarla. 2) Para ver que es un grupo no basta obtener v_{12} , también hay que verificar que $\gamma_{12} = \gamma(v_{12})$ ya que sólo hay una coordenada independiente, v , y γ no es independiente.

se deduce

$$\gamma_{12} = \gamma_1 \gamma_2 (1 + v_1 v_2), \quad v_{12} = \frac{v_1 + v_2}{1 + v_1 v_2}.$$

Como puede comprobarse $\gamma_{12} = \gamma(v_{12})$, por tanto $GG \subseteq G$ (es cerrado bajo multiplicación). La propiedad asociativa es automática para el producto de matrices. El neutro es $B(0) = 1$ (matriz identidad) y el inverso de $B(v)$ es $B(-v)$ como se deduce de la ley de composición. Esto confirma que G es un grupo.

$$v_{12} = \frac{v_1 + v_2}{1 + v_1 v_2}$$

es la ley de composición. Puesto que $v_{12} = v_{21}$ el grupo es abeliano. Cuando v varía entre $] -1, 1[$ se obtienen todos los elementos del grupo, que entonces es conexo. No es compacto ya que $] -1, 1[$ es acotado pero no cerrado. O también, $B(v)$ como subconjunto de \mathbb{R}^4 (4 elementos de matriz reales) no es un conjunto compacto (no es acotado).⁸⁶

Por cada álgebra de Lie sólo hay un grupo conexo y simplemente conexo. G es conexo y simplemente conexo igual que $(\mathbb{R}, +)$ que tiene la misma álgebra (es abeliana), por tanto G es isomorfo a $(\mathbb{R}, +)$.

b) La expresión para el generador se obtiene tomando un elemento infinitesimal

$$B(\delta v) = 1 - i\delta v K.$$

Desarrollando $B(v)$ hasta primer orden se obtiene (notando que $\gamma(v) = 1 + O(v^2)$)

$$B(\delta v) = \begin{pmatrix} 1 & \delta v \\ \delta v & 1 \end{pmatrix}$$

y por tanto

$$K = i \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = i\sigma_1.$$

c) v no es una coordenada canónica ya que v_{12} no es $v_1 + v_2$. La coordenada canónica es aquella en la que $\xi_{12} = \xi_1 + \xi_2$ (y el isomorfismo entre G y $(\mathbb{R}, +)$ es manifiesto). Para la coordenada canónica

$$B(v) = e^{-i\xi K}$$

y ésta es una forma simple de obtener la expresión de $\xi(v)$ (teniendo en cuenta que $\sigma_1^2 = 1$):

$$e^{-i\xi K} = e^{\xi \sigma_1} = \cosh(\xi) + \sinh(\xi)\sigma_1 = \begin{pmatrix} \cosh(\xi) & \sinh(\xi) \\ \sinh(\xi) & \cosh(\xi) \end{pmatrix}.$$

⁸⁶Las representaciones de los grupos compactos son unitarias y $B(v)$ no lo es.

Comparando con la expresión de $B(v)$ se deduce que $\gamma = \cosh(\xi)$ y

$$v = \frac{\sinh(\xi)}{\cosh(\xi)} = \tanh(\xi).$$

Alternativamente se puede buscar la función $\xi(v)$ tal que

$$\xi(v_{12}) = \xi(v_1) + \xi(v_2), \quad \xi(v) = v + O(v^2).$$

Si tomamos $v_1 = v$ y $v_2 = \delta v$ (infinitesimal), desarrollando en δv hasta primer orden se tendrá

$$\begin{aligned} \xi(v_{12}) &= \xi((v + \delta v)/(1 + v\delta v)) = \xi((v + \delta v)(1 - v\delta v)) = \xi(v + \delta v - v^2\delta v) \\ &= \xi(v) + \xi'(v)(1 - v^2)\delta v, \\ \xi(v_1) + \xi(v_2) &= \xi(v) + \delta v. \end{aligned}$$

Comparando las dos expresiones se deduce

$$\xi'(v) = \frac{1}{1 - v^2}, \quad \xi(v) = \operatorname{atanh}(v).$$

d) Lo más simple es reducir K . Sus autovalores son $\pm i$ y por tanto, en su base diagonal K es $\operatorname{diag}(i, -i)$ y B es $\operatorname{diag}(e^\xi, e^{-\xi})$. $e^{\pm\xi}$ son las dos irreps unidimensionales que aparecen en la reducción. Los vectores propios son $(1, \pm 1)$. Estos dos subespacios unidimensionales son invariantes Lorentz (forman el cono de luz).

2. Enunciado:

Según la fórmula de Campbell-Hausdorff,

$$C = \log(e^A e^B) = A + B + \frac{1}{2}[A, B] + \text{términos cúbicos}.$$

Se pide verificar esta expresión hasta el orden indicado. Por ejemplo, desarrollando en serie las exponenciales, expresar $e^A e^B = 1 + X$ y usar el desarrollo en serie de $\log(1 + X)$.

Solución:

$$\begin{aligned}
 C &= \log(e^A e^B), \\
 e^A &= 1 + A + \frac{A^2}{2} + O(X^3), \\
 e^A e^B &= (1 + A + \frac{A^2}{2} + O(X^3))(1 + B + \frac{B^2}{2} + O(X^3)) \\
 &= 1 + A + B + \frac{A^2}{2} + \frac{B^2}{2} + AB + O(X^3) := 1 + Y, \\
 C &= \log(1 + Y) = Y - \frac{Y^2}{2} + O(X^3) \\
 &= A + B + \frac{A^2}{2} + \frac{B^2}{2} + AB - \frac{1}{2}(A + B)^2 + O(X^3) \\
 &= A + B + \frac{1}{2}(AB - BA) + O(X^3).
 \end{aligned}$$

3. Enunciado:

Sea $T(a, b)$ ($a, b \in \mathbb{R}$) el operador que actúa en $L^2(\mathbb{R})$ como $\psi(x) \rightarrow \psi'(x) = \psi(e^{-a}(x - b))$. Estos operadores forman un grupo (no se pide demostrarlo) que es isomorfo al de translaciones más dilataciones en \mathbb{R} , siendo a, b son las coordenadas de los elementos del grupo.

a) Obtén los generadores infinitesimales asociados a las coordenadas a y b (que denotamos D y P , respectivamente). ¿Son los operadores T unitarios? ¿Son D ó P operadores hermíticos?

b) Usando los generadores hallados en el apartado anterior, obtén las relaciones de conmutación del álgebra de Lie del grupo.

Solución:

$$(T(a, b)\psi)(x) = \psi(e^{-a}(x - b)).$$

a) Por definición de D y P

$$T(\delta a, \delta b) = 1 - i\delta a D - i\delta b P.$$

Tomando a y b infinitesimales

$$(T(\delta a, \delta b)\psi)(x) = \psi(e^{-\delta a}(x - \delta b)) = \psi((1 - \delta a)(x - \delta b)) = \psi(x - \delta b - \delta a x) = \psi - (\delta b + \delta a x)\partial_x \psi.$$

Comparando con

$$(T(\delta a, \delta b)\psi)(x) = (1 - i\delta a D - i\delta b P)\psi(x) = \psi - i\delta a D\psi - i\delta b P\psi$$

se deduce

$$-iD = -x\partial_x, \quad -iP = -\partial_x,$$

es decir,

$$P = -i\partial_x, \quad D = xP.$$

P es hermítico correspondiendo a que las traslaciones $\psi(x) \rightarrow \psi(x-b)$ definen un operador unitario (conserva la norma). D no es hermítico ($D^\dagger = Px = xP - i = D - i$) porque las dilataciones, $\psi(x) \rightarrow \psi(e^{-a}x)$ no conservan la norma.

b) La única relación de conmutación no trivial es $[D, P]$

$$[D, P] = [xP, P] = x[P, P] + [x, P]P = 0 + iP = iP,$$

(usando que la operación $[, A]$ es una derivación).

7.2. Relación 2: Grupo de rotaciones

4. Enunciado:

Sean \mathbf{A} y \mathbf{B} dos operadores vectoriales. Se pide comprobar que $\mathbf{A} \cdot \mathbf{B}$ conmuta con \mathbf{J} como debe hacerlo un escalar y $\mathbf{A} \times \mathbf{B}$ como debe hacerlo un vector.

Solución:

Que \mathbf{A} sea vectorial quiere decir que

$$[J_i, A_j] = i\epsilon_{ijk}A_k.$$

Definimos $S = A_i B_i$ (sumado sobre i). Entonces

$$\begin{aligned} [J_i, S] &= [J_i, A_j B_j] = [J_i, A_j] B_j + A_j [J_i, B_j] = i\epsilon_{ijk}A_k B_j + A_j i\epsilon_{ijk}B_k \\ &= i\epsilon_{ikj}A_j B_k + i\epsilon_{ijk}A_j B_k = 0, \end{aligned} \quad (7.1)$$

por $\epsilon_{ikj} = -\epsilon_{ijk}$.

Igualmente definimos $C_i = \epsilon_{ijk}A_j B_k$ ($\mathbf{C} = \mathbf{A} \times \mathbf{B}$):

$$\begin{aligned} [J_i, C_j] &= [J_i, \epsilon_{jab}A_a B_b] = \epsilon_{jab}([J_i, A_a]B_b + A_a[J_i, B_b]) \\ &= \epsilon_{jab}(i\epsilon_{iak}A_k B_b + i\epsilon_{ibk}A_a B_k). \end{aligned} \quad (7.2)$$

Usando la identidad

$$\epsilon_{ijk}\epsilon_{abk} = \delta_{ia}\delta_{jb} - \delta_{ib}\delta_{ja},$$

se tiene

$$\begin{aligned} [J_i, C_j] &= i(\delta_{ji}\delta_{bk} - \delta_{jk}\delta_{bi})A_k B_b - i(\delta_{ji}\delta_{ak} - \delta_{jk}\delta_{ai})A_a B_k \\ &= i(\delta_{ji}A_k B_k - A_j B_i) - i(\delta_{ji}A_k B_k - A_i B_j) = i(A_i B_j - A_j B_i). \end{aligned} \quad (7.3)$$

Por otro lado

$$i\epsilon_{ijk}C_k = i\epsilon_{ijk}\epsilon_{kab}A_a B_b = i(\delta_{ia}\delta_{jb} - \delta_{ib}\delta_{ja})A_a B_b = i(A_i B_j - A_j B_i).$$

Se deduce que \mathbf{C} es un vector ya que $[J_i, C_j] = i\epsilon_{ijk}C_k$.

5. Enunciado:

El espacio de Hilbert de una partícula con espín 1/2 es $\mathcal{H} = L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$. Es decir, sus funciones de onda son $\psi(\mathbf{x}, \alpha)$ donde $\mathbf{x} \in \mathbb{R}^3$ y $\alpha = 1, 2$ (ó \uparrow y \downarrow). Se suele escribir en la forma $\psi_\alpha(\mathbf{x})$ y se puede disponer en forma de biespinor, $\Psi(\mathbf{x}) = \begin{pmatrix} \psi_1(\mathbf{x}) \\ \psi_2(\mathbf{x}) \end{pmatrix}$. La base de estados con posición y espín bien definido son $|\mathbf{x}, \alpha\rangle$, de modo que $|\psi\rangle = \int d^3x \sum_{\alpha=1,2} \psi_\alpha(\mathbf{x}) |\mathbf{x}, \alpha\rangle$. El grupo de rotaciones $SO(3)$ actúa en este espacio como

$$U(R)|\mathbf{x}, \alpha\rangle = \sum_{\beta=1,2} D_{\beta\alpha}^{(1/2)}(R) |R\mathbf{x}, \beta\rangle$$

donde $D^{(1/2)}(R)$ es la matriz de $SU(2)$ que representa a R en la representación $j = 1/2$, a saber, $D^{(1/2)}(R) = \exp(-i\boldsymbol{\theta} \cdot \mathbf{S})$ con $\mathbf{S} = \frac{1}{2}\boldsymbol{\sigma}$. Se pide obtener la acción sobre los biespinores y calcular los generadores infinitesimales \mathbf{J} de esta representación.

Solución:

Para obtener la acción sobre la función de onda partimos de

$$|\psi\rangle = \int d^3y \sum_{\beta} \psi_{\beta}(\mathbf{y}) |\mathbf{y}, \beta\rangle,$$

y aplicamos $U(R)$:

$$|\psi^R\rangle = U(R)|\psi\rangle = \int d^3y \sum_{\beta} \psi_{\beta}(\mathbf{y}) U(R)|\mathbf{y}, \beta\rangle = \int d^3y \sum_{\beta, \alpha} \psi_{\beta}(\mathbf{y}) D_{\alpha\beta}^{(1/2)}(R) |R\mathbf{y}, \alpha\rangle,$$

Por otro lado

$$|\psi^R\rangle = \int d^3x \sum_{\alpha} \psi_{\alpha}^R(\mathbf{x}) |\mathbf{x}, \alpha\rangle$$

Usando $\mathbf{x} = R\mathbf{y}$ y $d^3x = d^3y$, se deduce

$$\psi_{\alpha}^R(\mathbf{x}) = D_{\alpha\beta}^{(1/2)}(R) \psi_{\beta}(R^{-1}\mathbf{x}),$$

o en notación matricial

$$\Psi^R(\mathbf{x}) = D^{(1/2)}(R) \Psi(R^{-1}\mathbf{x}).$$

Para los generadores infinitesimales, sabemos que

$$D^{(1/2)}(R) = e^{-i\boldsymbol{\theta} \cdot \mathbf{S}}, \quad \phi(R^{-1}\mathbf{x}) = e^{-i\boldsymbol{\theta} \cdot \mathbf{L}} \phi(\mathbf{x}),$$

(con $R = R(\boldsymbol{\theta})$), por tanto

$$\Psi^R(\mathbf{x}) = e^{-i\boldsymbol{\theta} \cdot \mathbf{S}} e^{-i\boldsymbol{\theta} \cdot \mathbf{L}} \Psi(\mathbf{x}) = e^{-i\boldsymbol{\theta} \cdot \mathbf{J}} \Psi(\mathbf{x}),$$

con

$$\mathbf{J} = \mathbf{S} + \mathbf{L},$$

o más explícitamente $\mathbf{J} = 1 \otimes \mathbf{S} + \mathbf{L} \otimes 1$ en $\mathcal{H} = L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$.

6. Enunciado:

Usando las propiedades de las matrices de Pauli, por ejemplo, $\sigma_i \sigma_j = \delta_{ij} + i\epsilon_{ijk} \sigma_k$, probar la relación

$$(\mathbf{a} \cdot \boldsymbol{\sigma})(\mathbf{b} \cdot \boldsymbol{\sigma}) = \mathbf{a} \cdot \mathbf{b} + i\mathbf{a} \times \mathbf{b} \cdot \boldsymbol{\sigma}, \quad \mathbf{a}, \mathbf{b} \in \mathbb{C}^3.$$

El grupo $SU(2)$ se puede parametrizar mediante $U(\mathbf{a}) = \sqrt{1 - \mathbf{a}^2} - i\mathbf{a} \cdot \boldsymbol{\sigma}$ para $\mathbf{a} \in \mathbb{R}^3$ y $\|\mathbf{a}\| \leq 1$. Obtener la ley de composición en las coordenadas \mathbf{a} . Verificar que estas coordenadas no son canónicas. Obtener la coordenada del elemento inverso.

Solución:

Identidad:

$$(\mathbf{a} \cdot \boldsymbol{\sigma})(\mathbf{b} \cdot \boldsymbol{\sigma}) = a_i b_j \sigma_i \sigma_j = a_i b_j (\delta_{ij} + i\epsilon_{ijk} \sigma_k) = a_i b_i + i\epsilon_{ijk} a_i b_j \sigma_k = \mathbf{a} \cdot \mathbf{b} + i\mathbf{a} \times \mathbf{b} \cdot \boldsymbol{\sigma}$$

Ley de composición:

$$U(\mathbf{c}) = U(\mathbf{a})U(\mathbf{b})$$

Usando la notación $a_0 = \sqrt{1 - \mathbf{a}^2}$, ídem b_0 ,

$$\begin{aligned} U(\mathbf{a})U(\mathbf{b}) &= (a_0 - i\mathbf{a} \cdot \boldsymbol{\sigma})(b_0 - i\mathbf{b} \cdot \boldsymbol{\sigma}) = a_0 b_0 - i(a_0 \mathbf{b} + b_0 \mathbf{a}) \cdot \boldsymbol{\sigma} - \mathbf{a} \cdot \boldsymbol{\sigma} \mathbf{b} \cdot \boldsymbol{\sigma} \\ &= a_0 b_0 - \mathbf{a} \cdot \mathbf{b} - i(a_0 \mathbf{b} + b_0 \mathbf{a} + \mathbf{a} \times \mathbf{b}) \cdot \boldsymbol{\sigma}. \end{aligned}$$

Comparando con

$$U(\mathbf{c}) = c_0 - i\mathbf{c} \cdot \boldsymbol{\sigma}$$

se deduce

$$\mathbf{c} = a_0 \mathbf{b} + b_0 \mathbf{a} + \mathbf{a} \times \mathbf{b} \quad (\text{ley de composición}).$$

También puede comprobarse que $\sqrt{1 - \mathbf{c}^2} = c_0 = a_0 b_0 - \mathbf{a} \cdot \mathbf{b}$.

Coordenadas canónicas: Si las coordenadas fueran canónicas $\mathbf{f}(\mathbf{a}, \mathbf{a}) = 2\mathbf{a}$ y aquí se encuentra que $\mathbf{c} = \mathbf{f}(\mathbf{a}, \mathbf{a}) = 2a_0 \mathbf{a}$.

Inverso: Puesto que $U(\mathbf{a})$ es unitaria

$$U(\mathbf{a})^{-1} = U(\mathbf{a})^\dagger = (a_0 - i\mathbf{a} \cdot \boldsymbol{\sigma})^\dagger = a_0 + i\mathbf{a} \cdot \boldsymbol{\sigma} = U(-\mathbf{a}).$$

Se deduce que las coordenadas del elemento inverso de \mathbf{a} son $-\mathbf{a}$. También se deduce de la ley de composición ya que $f(\mathbf{a}, -\mathbf{a}) = 0$.

7. Enunciado:

Probar que $U(n)/SU(n) \cong U(1)$.

Solución:

Para todo homomorfismo de grupos $f : G \rightarrow H$, $G/\text{Ker}f \cong \text{ran}f$.

Como f tomamos la aplicación $f(U) = \det(U)$, $U \in U(n)$. Como U es unitaria su determinante es una fase y pertenece a $U(1)$ (basta tomar el determinante en la base en la que U es diagonal). Dado que $\det(AB) = \det(A)\det(B)$, esta aplicación es un homomorfismo de $U(n)$ en $U(1)$.

La imagen de f son todos los determinantes de matrices unitarias, y estos son todos los elementos de $U(1)$ ya que basta tomar $U = \text{diag}(\omega, 1, \dots, 1)$ con $\omega \in U(1)$.

El núcleo de f son los U con determinante 1 (el neutro de $U(1)$). Los U con determinante 1 forman el grupo $SU(n)$.

Por tanto $U(n)/SU(n) \cong U(1)$.

8. Enunciado:

Derivar en detalle las relaciones de conmutación de J_+ , J_- y J_3 a partir de las de J_1 , J_2 y J_3 . Comprobar que dichas relaciones se verifican aplicándolas sobre la base estándar $|jm\rangle$ (usando las acciones de los operadores J_\pm, J_3 sobre $|jm\rangle$ obtenidas en clase).

Solución:

Relaciones de conmutación de J_\pm, J_3 :

$$\begin{aligned} [J_3, J_\pm] &= [J_3, J_1 \pm iJ_2] = iJ_2 \pm i(-iJ_1) = \pm J_1 + iJ_2 = \pm(J_1 \pm iJ_2) = \pm J_\pm, \\ [J_+, J_-] &= [J_1 + iJ_2, J_1 - iJ_2] = -i[J_1, J_2] + i[J_2, J_1] = -i(iJ_3) + i(-iJ_3) = 2J_3. \end{aligned}$$

Comprobación de las relaciones de conmutación en la base estándar:

$$\begin{aligned} J_3|j, m\rangle &= m|j, m\rangle, & J_\pm|j, m\rangle &= N_\pm(j, m)|j, m \pm 1\rangle, \\ N_\pm(j, m) &= \sqrt{j(j+1) - m(m \pm 1)} = \sqrt{(j \mp m)(j \pm m + 1)} \end{aligned}$$

Comprobamos las relaciones aplicándolas sobre la base:

$$\begin{aligned}
 [J_3, J_{\pm}]|j, m\rangle &= (J_3 J_{\pm} - J_{\pm} J_3)|j, m\rangle = N_{\pm}(j, m)J_3|j, m \pm 1\rangle - mJ_{\pm}|j, m\rangle \\
 &= N_{\pm}(j, m)(m \pm 1)|j, m \pm 1\rangle - mN_{\pm}(j, m)|j, m \pm 1\rangle = \pm N_{\pm}(j, m)|j, m \pm 1\rangle \\
 &= \pm J_{\pm}|j, m\rangle
 \end{aligned}$$

$$\begin{aligned}
 [J_+, J_-]|j, m\rangle &= (J_+ J_- - J_- J_+)|j, m\rangle = N_-(j, m)J_+|j, m - 1\rangle - N_+(j, m)J_-|j, m + 1\rangle \\
 &= (N_-(j, m)N_+(j, m - 1) - N_+(j, m)N_-(j, m + 1))|j, m\rangle \\
 &= (N_-(j, m)^2 - N_+(j, m)^2)|j, m\rangle = 2m|j, m\rangle = 2J_3|j, m\rangle.
 \end{aligned}$$

7.3. Relación 3: Grupo de rotaciones

9. Enunciado:

Calcular cuántos escalares y cuántos vectores independientes se pueden formar tensorialmente con tres vectores.

Solución:

Usamos la serie de Clebsch-Gordan de SU(2):

$$D^{j_1} \otimes D^{j_2} = \bigoplus_{j=j_m}^{j_M} D^j, \quad j_m = |j_1 - j_2|, \quad j_M = j_1 + j_2.$$

Así

$$D^1 \otimes D^1 = D^0 \oplus D^1 \oplus D^2,$$

con dos vectores (un vector cae en la representación $j = 1$) se puede construir un escalar ($j = 0$), un vector ($j = 1$) y un tensor cuadrupolar ($j = 2$). Si los vectores son \mathbf{A}_1 y \mathbf{A}_2 , el escalar es $\mathbf{A}_1 \cdot \mathbf{A}_2$ y el vector es $\mathbf{A}_1 \times \mathbf{A}_2$.

Con tres vectores

$$\begin{aligned} D^1 \otimes D^1 \otimes D^1 &= (D^0 \oplus D^1 \oplus D^2) \otimes D^1 = (D^0 \otimes D^1) \oplus (D^1 \otimes D^1) \oplus (D^2 \otimes D^1) \\ &= (D^1) \oplus (D^0 \oplus D^1 \oplus D^2) \oplus (D^1 \oplus D^2 \oplus D^3) \\ &= 1 \times D^0 \oplus 3 \times D^1 \oplus 2 \times D^2 \oplus 1 \times D^3. \end{aligned}$$

Se puede construir un escalar $\mathbf{A}_1 \times \mathbf{A}_2 \cdot \mathbf{A}_3$, y tres vectores⁸⁷

$$(\mathbf{A}_1 \cdot \mathbf{A}_2)\mathbf{A}_3, \quad (\mathbf{A}_1 \cdot \mathbf{A}_3)\mathbf{A}_2, \quad (\mathbf{A}_2 \cdot \mathbf{A}_3)\mathbf{A}_1.$$

Análogamente desarrollando el producto de cuatro vectores se obtiene

$$D^1 \otimes D^1 \otimes D^1 \otimes D^1 = 3 \times D^0 \oplus 6 \times D^1 \oplus 6 \times D^2 \oplus 3 \times D^3 \oplus 1 \times D^4.$$

Comprobación de dimensiones:

$$3^4 = 3 \times 1 + 6 \times 3 + 6 \times 5 + 3 \times 7 + 1 \times 9.$$

⁸⁷Más exactamente $A_1^i A_2^j A_3^k, A_1^j A_2^i A_3^k, A_1^i A_2^k A_3^j$, ya que no debe suponerse que los \mathbf{A}_i conmutan entre sí.

10. **Enunciado:**

Sea V_1 un espacio que lleva una irrep $j = 1$ de $SU(2)$ y denotemos los tres estados $|j, m\rangle = |1, 1\rangle, |1, 0\rangle, |1, -1\rangle$ de su base estándar como α, β y γ , respectivamente. El espacio $V = V_1 \otimes V_1$ admite una base desacoplada $\alpha\alpha, \alpha\beta, \alpha\gamma, \beta\alpha, \dots, \gamma\gamma$ (nueve estados en total). Si denotamos $|J, M\rangle$ ($J = 0, 1, 2, M = -J, \dots, J$) los estados de la base acoplada, se tiene (esto puede usarse sin demostración)

$$|2, 2\rangle = \alpha\alpha, \quad |2, 1\rangle = \sqrt{\frac{1}{2}}\alpha\beta + \sqrt{\frac{1}{2}}\beta\alpha, \quad |1, 1\rangle = \sqrt{\frac{1}{2}}\alpha\beta - \sqrt{\frac{1}{2}}\beta\alpha.$$

Se pide

- (i) Probar que $|2, 0\rangle = \sqrt{\frac{1}{6}}\alpha\gamma + \sqrt{\frac{2}{3}}\beta\beta + \sqrt{\frac{1}{6}}\gamma\alpha$,
- (ii) Expresar el estado $|1, 0\rangle$ en la base desacoplada,
- (iii) Lo mismo para $|0, 0\rangle$.

Solución:

Usando la expresión de los operadores escalera $J_{\pm}|j, m\rangle = N_{\pm}(j, m)|j, m \pm 1\rangle$, tenemos

$$J_- \alpha = \sqrt{2}\beta, \quad J_- \beta = \sqrt{2}\gamma, \quad J_- \gamma = 0$$

- (i) Para obtener $|2, 0\rangle$ usamos la relación

$$J_-|2, 1\rangle = \sqrt{6}|2, 0\rangle.$$

Por otro lado, teniendo en cuenta que $\mathbf{J} = \mathbf{J}^{(1)} + \mathbf{J}^{(2)}$,

$$\begin{aligned} J_-|2, 1\rangle &= (J_-^{(1)} + J_-^{(2)})\sqrt{\frac{1}{2}}(\alpha\beta + \beta\alpha) \\ &= \sqrt{\frac{1}{2}}((J_- \alpha)\beta + (J_- \beta)\alpha + \alpha(J_- \beta) + \beta(J_- \alpha)) \\ &= \beta\beta + \gamma\alpha + \alpha\gamma + \beta\beta \end{aligned}$$

se deduce

$$|2, 0\rangle = \sqrt{\frac{1}{6}}\alpha\gamma + \sqrt{\frac{4}{6}}\beta\beta + \sqrt{\frac{1}{6}}\gamma\alpha.$$

Nótese que este vector debe salir correctamente normalizado, y ésta es una comprobación del cálculo.

(ii) Por un lado

$$J_-|1,1\rangle = \sqrt{2}|1,0\rangle.$$

Por otro lado

$$\begin{aligned} J_-|1,1\rangle &= (J_-^{(1)} + J_-^{(2)})\sqrt{\frac{1}{2}}(\alpha\beta - \beta\alpha) \\ &= \sqrt{\frac{1}{2}}((J_- \alpha)\beta - (J_- \beta)\alpha + \alpha(J_- \beta) - \beta(J_- \alpha)) \\ &= \beta\beta - \gamma\alpha + \alpha\gamma - \beta\beta = \alpha\gamma - \gamma\alpha \end{aligned}$$

que implica

$$|1,0\rangle = \sqrt{\frac{1}{2}}\alpha\gamma - \sqrt{\frac{1}{2}}\gamma\alpha.$$

(iii) El estado $|00\rangle$ es de la forma

$$|00\rangle = a\alpha\gamma + b\beta\beta + c\gamma\alpha.$$

Por ortogonalidad con $|20\rangle$ y $|10\rangle$

$$|00\rangle \propto \alpha\gamma - \beta\beta + \gamma\alpha,$$

y por normalización

$$|00\rangle = e^{i\varphi} \sqrt{\frac{1}{3}}(\alpha\gamma - \beta\beta + \gamma\alpha).$$

La fase se fija por el convenio de que el primer vector de la base tenga un coeficiente positivo. En este caso la base es $\alpha\gamma, \beta\beta, \gamma\alpha$ y el primer vector es $\alpha\gamma$ porque tiene m_1 máximo. Finalmente

$$|00\rangle = \sqrt{\frac{1}{3}}(\alpha\gamma - \beta\beta + \gamma\alpha).$$

11. Enunciado:

Se sabe que cierto espacio de Hilbert \mathcal{H} de dimensión 4 soporta una representación del grupo $SU(2)$, y por tanto admite una reducción $\mathcal{H} \cong \bigoplus_j n_j V_j$, donde n_j indica la multiplicidad de la irrep $j = 0, 1/2, 1, 3/2, \dots$

1) Enumera las 5 reducciones compatibles con $\dim \mathcal{H} = 4$.

2) Al aplicar J_+ sobre cualquier vector de \mathcal{H} se obtiene siempre un múltiplo de un vector fijo, en otras palabras, el rango del operador J_+ es 1.⁸⁸ Selecciona las reducciones compatibles con esta nueva condición. ¿Cuál es ese vector fijo?

Solución:

1) Las particiones de 4 son: 4, 3 + 1, 2 + 2, 2 + 1 + 1 y 1 + 1 + 1 + 1, por tanto hay cinco reducciones posibles (teniendo en cuenta que para el grupo SU(2) cada dimensión corresponde exactamente a una irrep):

a) $\mathcal{H} = V_{j=3/2}$

b) $\mathcal{H} = V_{j=1} \oplus V_{j=0}$

c) $\mathcal{H} = V_{j=1/2} \oplus V_{j=1/2}$

d) $\mathcal{H} = V_{j=1/2} \oplus V_{j=0} \oplus V_{j=0}$

e) $\mathcal{H} = V_{j=0} \oplus V_{j=0} \oplus V_{j=0} \oplus V_{j=0}$

2) Al aplicar J_+ sobre $|j, m\rangle$ se obtiene $|j, m+1\rangle$ o cero si $m = j$, en consecuencia el rango de J_+ en un espacio V_j es $2j$. Por tanto los rangos para los casos a-e) son, respectivamente 3, $2+0=2$, $1+1=2$, $1+0+0=1$ y $0+0+0+0=0$. Sólo el caso d) tiene rango 1 y el vector que queda al aplicar J_+ sobre un vector cualquiera de ese espacio es proporcional a $|\frac{1}{2}, \frac{1}{2}\rangle$.

12. Enunciado:

El grupo SU(2) tiene una medida invariante normalizada que en coordenadas canónicas toma la forma

$$\langle f(U) \rangle_{\text{SU}(2)} = \int_0^{2\pi} d\psi \int_{S^2} d^2\Omega \frac{\text{sen}^2(\psi/2)}{4\pi^2} f(U(\psi)), \quad \psi = \psi \hat{n}.$$

Calcular los caracteres de SU(2), $\chi_j(U) = \text{tr}(D^j(U))$, correspondientes a $j = 0, \frac{1}{2}, 1$ y verificar que son ortogonales entre sí (respecto de la medida invariante del grupo).

Solución:

El producto escalar asociado a la medida invariante es

$$\langle \psi_1 | \psi_2 \rangle = \langle \psi_1^*(U) \psi_2(U) \rangle_{\text{SU}(2)}.$$

⁸⁸El rango de un operador es la dimensión de su espacio imagen.

Los caracteres sólo dependen de la clase de conjugación que a su vez sólo depende del ángulo de rotación. Se puede elegir una rotación sobre el eje z . En este caso la matriz de rotación D^j es diagonal ya que

$$|j, m\rangle \rightarrow e^{-i\psi J_3} |jm\rangle = e^{-i\psi m} |jm\rangle,$$

por tanto $D^j = \text{diag}(\dots, e^{-i\psi m}, \dots)$ con $m = j, j-1, \dots, -j$. La traza es la suma de la diagonal, es una serie geométrica y se obtiene

$$\chi_j(\psi) = \frac{\text{sen}(n\psi/2)}{\text{sen}(\psi/2)}, \quad n = 2j + 1.$$

En particular para $j = 1$ ($n = 3$)

$$\chi_1(\psi) = e^{-i\psi} + 1 + e^{i\psi} = 1 + 2\cos(\psi),$$

que es lo mismo que se obtiene tomando la traza de

$$R(e_3, \psi) = \begin{pmatrix} \cos(\psi) & -\text{sen}(\psi) & 0 \\ \text{sen}(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

(es la matriz de rotación en otra base y la traza no depende de la base).

Para ver que los caracteres son ortogonales, calculamos (usando $\int d^2\Omega = 4\pi$)

$$\langle \chi_{j_1} | \chi_{j_2} \rangle = \int_0^{2\pi} \frac{d\psi}{\pi} \text{sen}(n_1\psi/2) \text{sen}(n_2\psi/2).$$

Mediante la identidad $\cos(a \pm b) = \cos(a)\cos(b) \pm \text{sen}(a)\text{sen}(b)$ y el cambio de variable $\psi = 2x$

$$\langle \chi_{j_1} | \chi_{j_2} \rangle = \frac{1}{\pi} \int_0^\pi dx (\cos((n_1 - n_2)x) - \cos((n_1 + n_2)x)).$$

Como $n_1 \pm n_2$ es entero se puede usar

$$\int_0^\pi \cos(nx) dx = \begin{cases} \frac{1}{n} \text{sen}(n\pi) = 0 & n \neq 0 \\ \pi & n = 0 \end{cases} = \pi \delta_{n,0},$$

y se tiene

$$\langle \chi_{j_1} | \chi_{j_2} \rangle = \delta_{n_1 - n_2, 0} - \delta_{n_1 + n_2, 0}.$$

Finalmente $n_{1,2} = 1, 2, \dots$ (por lo que $n_1 + n_2 > 0$) y

$$\langle \chi_{j_1} | \chi_{j_2} \rangle = \delta_{j_1, j_2}.$$

7.4. Relación 1 de problemas laboriosos.

1. Enunciado:

El álgebra de Lie de dimensión 2 más general es $[X_1, X_2] = i\alpha X_1 + i\beta X_2$, $\alpha, \beta \in \mathbb{R}$.

i) Para toda álgebra de Lie, las propias constantes de estructura definen una representación (no necesariamente fiel) denominada representación adjunta. En esa representación los X_i son matrices $n \times n$ con elementos de matriz $(X_i)_{ab} = ic_{ib}^a$. Construir las dos matrices de la representación adjunta para el álgebra bidimensional y comprobar que en efecto satisfacen las relaciones de conmutación.⁸⁹

ii) Probar que haciendo el cambio de base apropiado, toda álgebra bidimensional adopta una de estas dos formas: a) $[Y_1, Y_2] = 0$ (abeliana) o b) $[Y_1, Y_2] = iY_1$ (no abeliana).

iii) Para $[Y_1, Y_2] = iY_1$, se pide calcular la ley de composición del grupo en las coordenadas canónicas asociadas a Y_1 e Y_2 .

Solución:

i) Para unas constantes de estructura cualesquiera se tiene $(X_k)_{ab} = ic_{kb}^a$, donde a (primer índice) es la fila y b (segundo índice) la columna:

$$X_k = i \begin{pmatrix} c_{k1}^1 & c_{k2}^1 & \cdots & c_{kn}^1 \\ c_{k1}^2 & c_{k2}^2 & \cdots & c_{kn}^2 \\ \vdots & \vdots & & \vdots \\ c_{k1}^n & c_{k2}^n & \cdots & c_{kn}^n \end{pmatrix}, \quad k = 1, 2, \dots, n.$$

De la relación de conmutación $[X_1, X_2] = i\alpha X_1 + i\beta X_2$ (teniendo en cuenta además que c_{ij}^k es antisimétrico en ij), se lee

$$c_{12}^1 = -c_{21}^1 = \alpha, \quad c_{12}^2 = -c_{21}^2 = \beta$$

y lo demás coeficientes se anulan. Por tanto

$$X_1 = i \begin{pmatrix} 0 & \alpha \\ 0 & \beta \end{pmatrix}, \quad X_2 = i \begin{pmatrix} -\alpha & 0 \\ -\beta & 0 \end{pmatrix}.$$

Comprobamos que cumplen las relaciones de conmutación

$$X_1 X_2 = \begin{pmatrix} \alpha\beta & 0 \\ \beta^2 & 0 \end{pmatrix}, \quad X_2 X_1 = \begin{pmatrix} 0 & \alpha^2 \\ 0 & \alpha\beta \end{pmatrix}$$

$$[X_1, X_2] = \begin{pmatrix} \alpha\beta & -\alpha^2 \\ \beta^2 & -\alpha\beta \end{pmatrix} = i\alpha X_1 + i\beta X_2.$$

⁸⁹Si se intenta hacer lo mismo en dimensiones mayores se encuentra que no cualquier elección de c_{ij}^k es aceptable. Además de ser antisimétricas, las constantes de estructura deben satisfacer la identidad de Jacobi.

ii) Si $\alpha = \beta = 0$ el álgebra es abeliana (en cualquier base). En otro caso, supongamos por ejemplo que $\alpha \neq 0$ (el caso $\beta \neq 0$ es análogo):

$$i\alpha X_1 + i\beta X_2 = [X_1, X_2]$$

$$i\left(X_1 + \frac{\beta}{\alpha} X_2\right) = \left[X_1, \frac{1}{\alpha} X_2\right] = \left[X_1 + \frac{\beta}{\alpha} X_2, \frac{1}{\alpha} X_2\right]$$

que es

$$iY_1 = [Y_1, Y_2], \quad \text{con} \quad Y_1 = X_1 + \frac{\beta}{\alpha} X_2, \quad Y_2 = \frac{1}{\alpha} X_2.$$

iii) La ley de composición $c = f(a, b)$ se puede obtener de

$$e^{-ic_1 Y_1 - ic_2 Y_2} = e^{-ia_1 Y_1 - ia_2 Y_2} e^{-ib_1 Y_1 - ib_2 Y_2},$$

usando cualquier representación fiel del álgebra. Podemos usar la representación adjunta y las matrices de $Y_{1,2}$ son las del apartado i) con $\alpha = 1$ y $\beta = 0$

$$-iY_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad -iY_2 = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Definimos $A = -ia_1 Y_1 - ia_2 Y_2$, ídem B y C de modo que $e^C = e^A e^B$.

$$A = \begin{pmatrix} -a_2 & a_1 \\ 0 & 0 \end{pmatrix}, \quad A^2 = \begin{pmatrix} a_2^2 & -a_1 a_2 \\ 0 & 0 \end{pmatrix} = -a_2 A.$$

Se deduce

$$A^n = \begin{cases} 1, & n = 0 \\ (-a_2)^{n-1} A, & n > 0 \end{cases},$$

lo cual permite calcular la exponencial en forma cerrada (separando el término $n = 0$)

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} = 1 + \sum_{n=1}^{\infty} \frac{A^n}{n!} = 1 + \sum_{n=1}^{\infty} \frac{(-a_2)^{n-1}}{n!} A = 1 - \frac{1}{a_2} \sum_{n=1}^{\infty} \frac{(-a_2)^n}{n!} A$$

$$= 1 - \frac{1}{a_2} (e^{-a_2} - 1) A$$

es decir

$$e^A = \begin{pmatrix} e^{-a_2} & \frac{1-e^{-a_2}}{a_2} a_1 \\ 0 & 1 \end{pmatrix}.$$

Multiplicando matrices

$$\begin{pmatrix} e^{-c_2} & \frac{1-e^{-c_2}}{c_2}c_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{-a_2} & \frac{1-e^{-a_2}}{a_2}a_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-b_2} & \frac{1-e^{-b_2}}{b_2}b_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{-a_2-b_2} & \frac{1-e^{-a_2}}{a_2}a_1 + \frac{1-e^{-b_2}}{b_2}e^{-a_2}b_1 \\ 0 & 1 \end{pmatrix},$$

de donde

$$c_2 = a_2 + b_2, \quad c_1 = \left(\frac{1-e^{-a_2}}{a_2}a_1 + \frac{1-e^{-b_2}}{b_2}e^{-a_2}b_1 \right) \frac{a_2 + b_2}{1-e^{-a_2-b_2}}.$$

La ley de composición es algo más simple en coordenadas auxiliares

$$e^A = \begin{pmatrix} 1-\alpha_2 & \alpha_1 \\ 0 & 1 \end{pmatrix}, \quad \text{ídem } e^B, e^C \text{ con } \beta_{1,2} \text{ y } \gamma_{1,2} \text{ respectivamente:}$$

$$\begin{pmatrix} 1-\gamma_2 & \gamma_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1-\alpha_2 & \alpha_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1-\beta_2 & \beta_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1-\alpha_2-\beta_2+\alpha_2\beta_2 & \alpha_1+\beta_1-\alpha_2\beta_1 \\ 0 & 1 \end{pmatrix}.$$

$$\gamma_1 = \alpha_1 + \beta_1 - \alpha_2\beta_1, \quad \gamma_2 = \alpha_2 + \beta_2 - \alpha_2\beta_2.$$

La ley de composición en coordenadas canónicas se obtiene invirtiendo la relación entre coordenadas a y α :

$$a_2 = -\log(1-\alpha_2), \quad a_1 = -\alpha_1 \frac{\log(1-\alpha_2)}{\alpha_2}.$$

2. Enunciado:

Usando la expresión explícita de las rotaciones,

$$R(\hat{n}, \theta) \mathbf{x} = \mathbf{x}_{\parallel} + \cos(\theta) \mathbf{x}_{\perp} + \sin(\theta) \hat{n} \times \mathbf{x}_{\perp}, \quad \mathbf{x} = \mathbf{x}_{\parallel} + \mathbf{x}_{\perp}, \quad \mathbf{x}_{\parallel} \parallel \hat{n} \perp \mathbf{x}_{\perp},$$

se pide verificar que los elementos de matriz de R dependen analíticamente de $\boldsymbol{\theta} = \theta \hat{n} = (\theta_1, \theta_2, \theta_3)$. Esto quiere decir que admiten desarrollo en serie de Taylor en potencias de θ_1 , θ_2 , y θ_3 . [Sabemos que tiene que ser así por $R(\hat{n}, \theta) = \exp(-i\boldsymbol{\theta} \cdot \mathbf{J})$, siendo \mathbf{J} tres matrices 3×3 , $(J_i)_{jk} = -i\epsilon_{ijk}$, pero se pide probarlo usando sólo la expresión explícita de $R(\hat{n}, \theta) \mathbf{x}$.]

Solución:

Función analítica (real) de $\boldsymbol{\theta}$ quiere decir

$$f(\boldsymbol{\theta}) = \sum_{n_1, n_2, n_3=0}^{\infty} f_{n_1 n_2 n_3} \theta_1^{n_1} \theta_2^{n_2} \theta_3^{n_3} = f_{000} + f_{100} \theta_1 + f_{010} \theta_2 + f_{001} \theta_3 + \dots,$$

y con radio de convergencia no nulo.

θ es una función analítica y lo mismo $\theta^2 = \theta_1^2 + \theta_2^2 + \theta_3^2$, pero $\theta = \sqrt{\theta_1^2 + \theta_2^2 + \theta_3^2}$ ó $\hat{n} = \theta/\theta$ no lo son. x es analítico (ya que es constante respecto de θ).

El problema consiste en probar que Rx puede expresarse usando funciones analíticas de θ (y en particular de θ^2).

El último término de Rx es analítico por sí mismo:

$$\sin(\theta)\hat{n} \times x_{\perp} = \frac{\sin(\theta)}{\theta}\theta \times x = (1 - \frac{\theta^2}{3!} + \dots)\theta \times x,,$$

es analítico porque $\sin(\theta)/\theta$ sólo tiene potencias pares de θ .

Para los otros dos términos

$$x_{\parallel} + \cos(\theta)x_{\perp} = (1 - \cos(\theta))x_{\parallel} + \cos(\theta)x.$$

El segundo sumando es analítico ($\cos(\theta)$ es par en θ). El término restante

$$(1 - \cos(\theta))x_{\parallel} = (1 - \cos(\theta))(\hat{n} \cdot x)\hat{n} = \frac{1 - \cos(\theta)}{\theta^2}(\theta \cdot x)\theta = (\frac{1}{2!} - \frac{1}{4!}\theta^2 + \dots)(\theta \cdot x)\theta$$

también es regular.

Todas las series involucradas tienen radio de convergencia infinito.

7.5. Relación 4: Monte Carlo

15. Enunciado:

Sean X e Y dos variables aleatorias independientes uniformemente distribuidas en $[0, 1]$. Se pide calcular la densidad de probabilidad de $Z = X/Y$, y calcular $\langle Z^\alpha \rangle$ para $|\alpha| < 1$.

Solución:

La distribución conjunta de X e Y es

$$p_{X,Y}(a,b) = \Theta(0 < a < 1)\Theta(0 < b < 1) = \begin{cases} 1 & a,b \in [0, 1] \\ 0 & \text{en otro caso} \end{cases}$$

Una opción es usar deltas de Dirac:

$$p_Z(c) = \langle \delta(c - Z) \rangle = \int da db p_{X,Y}(a,b) \delta(c - a/b) = \int_{[0,1] \times [0,1]} \delta(c - a/b) da db$$

Usamos la identidad

$$\delta(f(x)) = \sum_{x_i} \frac{\delta(x - x_i)}{|f'(x_i)|}, \quad \text{siendo } x_i \text{ los ceros de } f(x): f(x_i) = 0.$$

En nuestro caso $f = c - a/b$ como función de a :

$$\delta(c - a/b) = b \delta(a - cb) \quad (b > 0),$$

y obtenemos

$$p_Z(c) = \int_0^1 db b \int_0^1 da \delta(a - bc).$$

Aquí usamos la identidad

$$\int_a^b dx \delta(x - x_0) f(x) = f(x_0) \Theta(a \leq x_0 \leq b) \quad (a \leq b)$$

para obtener

$$p_Z(c) = \int_0^1 db b \Theta(1 - bc) = \int_0^{\min(1, 1/c)} db b = \frac{1}{2} \min(1, 1/c)^2.$$

Equivalentemente

$$p_Z(c) = \frac{1}{2} \Theta(1 - c) + \frac{1}{2} \frac{1}{c^2} \Theta(c - 1) = \begin{cases} \frac{1}{2} & c \leq 1 \\ \frac{1}{2c^2} & c > 1 \end{cases}.$$

En particular $\text{Prob}(Z > 1) = 1/2$.

Alternativamente se puede obtener $p_Z(c)$ por la propiedad (para $f(x)$ cualquiera)

$$\langle f(Z) \rangle = \int dc p_Z(c) f(c).$$

Expresando el valor esperado en términos de X e Y

$$\langle f(Z) \rangle = \int_{[0,1] \times [0,1]} f(a/b) da db.$$

La idea es hacer un cambio de variable eliminando a en favor de $c = a/b$ (para b fijo) y $da = bdc$. Sin embargo es conveniente pasar primero los límites de integración al integrando

$$\langle f(Z) \rangle = \int_0^\infty db \int_0^\infty da \Theta(1-b) \Theta(1-a) f(a/b) = \int_0^\infty db \int_0^\infty dc b \Theta(1-b) \Theta(1-bc) f(c).$$

Comparando con la expresión de más arriba y teniendo que $f(c)$ es arbitraria,

$$p_Z(c) = \int_0^\infty db b \Theta(1-b) \Theta(1-bc).$$

Esta expresión es equivalente a la obtenida anteriormente usando la delta de Dirac y da el mismo resultado.

El valor esperado de Z^α se puede obtener equivalentemente usando (X, Y) ó Z :

$$\langle Z^\alpha \rangle = \int_0^1 da \int_0^1 db (a/b)^\alpha = \int_0^1 da a^\alpha \int_0^1 db b^{-\alpha} = \frac{1}{1+\alpha} \frac{1}{1-\alpha}.$$

$$\langle Z^\alpha \rangle = \int_0^\infty dc p_Z(c) c^\alpha = \frac{1}{2} \int_0^1 dc c^\alpha + \frac{1}{2} \int_1^\infty dc c^{\alpha-2} = \frac{1}{2} \left(\frac{1}{\alpha+1} - \frac{1}{\alpha-1} \right) = \frac{1}{1-\alpha^2}.$$

16. Enunciado:

Sean X, Y dos variables aleatorias reales con densidad de probabilidad $p_{X,Y}(x, y)$ proporcional a $1/(1+x^2+y^2)^2$. Se pide especificar sendas funciones f_1 y f_2 tales que $X = f_1(u_1, \dots)$ e $Y = f_2(u_1, \dots)$ sigan la distribución pedida (siendo las u_i uniformemente distribuidas en $(0, 1)$ e independientes). De la covarianza de X e Y ¿se deduce que son variables independientes?

Solución:

Calculemos la normalización, usando coordenadas polares,

$$\begin{aligned}\int_{\mathbb{R}^2} dx dy \frac{1}{(1+x^2+y^2)^2} &= \int_0^{2\pi} d\phi \int_0^\infty dr r \frac{1}{(1+r^2)^2} = 2\pi \int_0^\infty d\xi \frac{1}{2(1+\xi)^2} \\ &= \pi \int_{\xi=\infty}^{\xi=0} d \frac{1}{1+\xi} = \pi.\end{aligned}$$

Así la probabilidad normalizada es

$$p_{X,Y}(x,y) = \frac{1}{\pi} \frac{1}{(1+x^2+y^2)^2}.$$

Para generar X e Y usamos coordenadas polares basándonos en la relación (para f arbitraria)

$$\begin{aligned}\langle f(X,Y) \rangle &= \int dx dy p_{X,Y}(x,y) f(x,y) \\ &= \frac{1}{\pi} \int_0^{2\pi} d\phi \int_0^\infty dr r \frac{1}{(1+r^2)^2} f = \int_0^{2\pi} d\left(\frac{\phi}{2\pi}\right) \int_{r=\infty}^{r=0} d\left(\frac{1}{1+r^2}\right) f \\ &= \int_0^1 du_\phi \int_0^1 du_r f.\end{aligned}$$

Por tanto el par (r, ϕ) se puede obtener con

$$r = \sqrt{\frac{1}{u_r} - 1}, \quad \phi = 2\pi u_\phi$$

siendo u_r, u_ϕ dos variables independientes uniformemente distribuidas en $[0, 1]$. X, Y se obtienen a partir de r y ϕ .

La covarianza de las variables X e Y se anula

$$\text{Cov}(X, Y) = \langle \Delta X \Delta Y \rangle = \langle XY \rangle = 0$$

usando que $\langle X \rangle = \langle Y \rangle = \langle XY \rangle = 0$ porque la distribución $P_{X,Y}(x,y)$ es par bajo $X \rightarrow -X$ y $Y \rightarrow -Y$ por separado. Que la covarianza se anule es condición necesaria para que X e Y sean independientes, pero no es suficiente. Las variables X, Y no son independientes porque $P_{X,Y}(x,y)$ no es separable.

18. Enunciado:

Queremos aplicar el método de aceptación-rechazo para hacer un muestreo del peso $w(x) = x^2$ con $0 \leq x \leq 1$. Disponemos de generadores para tres distribuciones de probabilidad auxiliares:

i) $q_1(x) \propto \Theta(0 \leq x \leq 1)$,

ii) $q_2(x) \propto x\Theta(0 \leq x \leq 1)$,

iii) $q_3(x) \propto x\Theta(0 \leq x \leq 2)$.

Explicar cómo se aplica el método en cada uno de los tres casos y determinar cuál de ellos es más eficiente (suponiendo que cuesta lo mismo generar cualquiera de esas variables aleatorias).

Solución:

i) La distribución normalizada es

$$q_1(x) = \Theta(0 \leq x \leq 1).$$

Queremos un K tal que $Kq_1(x) \geq w(x)$ para todo x , lo cual implica $K \geq 1$ en nuestro caso. El método es

1) generar un x de acuerdo con $q_1(x)$ (tenemos el generador),

2) generar u uniformemente en $[0, 1]$, y

3) aceptamos x si

$$u \leq \frac{w(x)}{Kq_1(x)} = \frac{x^2}{K},$$

en otro caso volvemos a 1).

La probabilidad de aceptación es \mathcal{N}/K donde, para nuestra w ,

$$\mathcal{N} = \int dx w(x) = \int_0^1 dx x^2 = \frac{1}{3}.$$

Para maximizar la probabilidad de aceptación tomamos el K mínimo: $K_1 = 1$ y la probabilidad de aceptación es $1/3$, es decir, de cada tres candidatos se acepta uno, en promedio.⁹⁰

ii) En este caso, después de normalizar,

$$q_2(x) = 2x\Theta(0 \leq x \leq 1).$$

⁹⁰Demuéstrase explícitamente que $\langle n \rangle = 1/p$ siendo n el orden del candidato aceptado en un experimento con probabilidad de aceptación p .

La condición $Kq_2 \geq w$ requiere $K \geq 1/2$. La elección óptima es $K_2 = 1/2$, y la probabilidad de aceptación es $2/3$. Se aceptan 2 candidatos de cada 3 propuestos. Si costara lo mismo generar con q_1 o con q_2 el segundo método sería más eficiente.

iii) En este caso

$$q_3(x) = \frac{1}{2}x\Theta(0 \leq x \leq 2)$$

x se va a generar en $[0, 2]$ en vez de $[0, 1]$. Obviamente todos los candidatos con $x > 1$ se van a rechazar (ya que w se anula en ese caso). El K óptimo es $K_3 = 2$ y la probabilidad de aceptación baja a $1/6$. De los $5/6$ rechazados, $3/4$ son los generados en el intervalo $[1, 2]$, del $1/4$ restante (generados en $[0, 1]$), se rechaza $1/3$, igual que en el caso ii).

19. Enunciado:

Cierto sistema puede hallarse en dos estados 1 y 2, con probabilidades p_1 y p_2 ($p_1 + p_2 = 1$). Queremos simularlo usando Metropolis y para hacer las propuestas se usa una probabilidad auxiliar $Q(1|2) = Q(2|1) = q$ y $Q(1|1) = Q(2|2) = 1 - q$ siendo q un parámetro entre 0 y 1. Se pide obtener la probabilidad $W(i|j)$ asociada y verificar que satisface balance detallado (para concretar supóngase que $p_1 \geq p_2$). Probar que la vida media para llegar al equilibrio es $\tau = -1/\log|1 - q/p_1|$.

Solución:

Es conveniente escribir las probabilidades de transición en forma de matriz:

$$W(i|j) = W_{ij}, \quad Q(i|j) = Q_{ij}$$

En nuestro caso son matrices 2×2 , y

$$Q = \begin{pmatrix} 1-q & q \\ q & 1-q \end{pmatrix}$$

Dado que $\sum_i Q(i|j) = 1$, la suma por columnas de las matrices de transición debe ser 1 en cada columna, como efectivamente ocurre. Lo mismo para W_{ij} .

De esta observación se deduce que basta calcular $W(i|j)$ para $i \neq j$. De acuerdo con el algoritmo de Metropolis $i \neq j$ sólo puede ocurrir si se propone un candidato i distinto del estado actual j y además se acepta. La probabilidad de que el candidato sea el otro estado es q y la probabilidad de aceptación es $\min(1, p_i/p_j)$:

$$(i \neq j) \quad W(i|j) = q \min\left(1, \frac{p_i}{p_j}\right) = \begin{cases} q & (i, j) = (1, 2) \\ q \frac{p_2}{p_1} & (i, j) = (2, 1) \end{cases}$$

teniendo en cuenta que $p_1 \geq p_2$. Completando $W(i|i)$ para que la suma sobre estados finales sea 1, tenemos

$$\mathbf{W} = \begin{pmatrix} 1 - q \frac{p_2}{p_1} & q \\ q \frac{p_2}{p_1} & 1 - q \end{pmatrix}.$$

Obsérvese que si $q = 1$ y además $p_1 = p_2 = \frac{1}{2}$, \mathbf{W} no es aperiódico y el método no va a converger.

Balance detallado requiere

$$W_{ij}p_j = W_{ji}p_i, \quad i, j = 1, 2$$

El único caso no trivial es $i \neq j$

$$W_{12}p_2 - W_{21}p_1 = qp_2 - q \frac{p_2}{p_1} p_1 = 0.$$

También se comprueba que (p_1, p_2) es una solución de equilibrio del proceso:

$$\sum_j W(i|j)p_j = p_i, \quad i = 1, 2$$

o en notación matricial

$$\mathbf{W}\mathbf{p} = \mathbf{p}, \quad \mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}.$$

Para ver el tiempo que tarda en llegar al equilibrio desde una configuración genérica (ρ_1, ρ_2) , notemos que cada paso markoviano equivale a aplicar la matriz \mathbf{W} sobre el vector $\boldsymbol{\rho}^{(k)}$ del paso k -ésimo:

$$\rho_i^{(k+1)} = \sum_j W(i|j)\rho_j^{(k)}, \quad \boldsymbol{\rho}^{(k+1)} = \mathbf{W}\boldsymbol{\rho}^{(k)}.$$

La matriz \mathbf{W} tiene dos vectores propios $|0\rangle$ y $|1\rangle$ con valores propios λ_0 y λ_1 , que forman una base,

$$\mathbf{W}|\alpha\rangle = \lambda_\alpha|\alpha\rangle, \quad \boldsymbol{\rho} = c_0|0\rangle + c_1|1\rangle$$

y en cada paso markoviano

$$\mathbf{W}\boldsymbol{\rho} = c_0\lambda_0|0\rangle + c_1\lambda_1|1\rangle.$$

Después de N pasos

$$\mathbf{W}^N\boldsymbol{\rho} = c_0\lambda_0^N|0\rangle + c_1\lambda_1^N|1\rangle.$$

Los $|\lambda_\alpha| < 1$ tienden a desaparecer para N grande. (Por el contrario un autovalor $|\lambda_\alpha| > 1$ explotaría y no se podría llegar al equilibrio.)

Por construcción \mathbf{p} es justamente un autovector $|0\rangle$ con autovalor $\lambda_0 = 1$. El otro autovalor se puede obtener de la ecuación característica, o más fácilmente de la traza de \mathbf{W} :

$$1 + \lambda_1 = \sum_{\alpha} \lambda_{\alpha} = \text{tr}(\mathbf{W}) = 2 - q - q \frac{p_2}{p_1} = 2 - \frac{q}{p_1},$$

es decir,

$$\lambda_1 = 1 - q/p_1.$$

Casos patológicos son: 1) $\lambda_1 = 1$, (si $q = 0$), la cadena no es irreducible, y 2) $\lambda_1 = -1$ que requiere $q = 1$, $p_1 = 1/2$, la cadena no es aperiódica. En estos casos no se llega al equilibrio a partir de una situación genérica. En los demás casos $|\lambda_1| < 1$. La evolución queda⁹¹

$$\mathbf{W}^N \boldsymbol{\rho} = c_0 \mathbf{p} + c_1 \lambda_1^N |1\rangle,$$

y la componente transitoria desaparece a un ritmo $|\lambda_1|^N$. Si se escribe en forma decaimiento exponencial, siendo N el tiempo,

$$|\lambda_1|^N = e^{-N/\tau}, \quad \frac{1}{\tau} = -\log |\lambda_1|, \quad \tau = -\frac{1}{\log(|1 - q/p_1|)}.$$

⁹¹De hecho $c_0 = 1$ necesariamente (eligiendo $|0\rangle = \mathbf{p}$) ya que $|1\rangle$ no contribuye a la normalización, su suma por columnas se anula.

Referencias

- [1] A. O. Barut y R. Rączka, *Theory of group representations and applications*, World Scientific Publishing, 1986.
- [2] S. Coleman, *Aspects of Symmetry*, Cambridge University Press, 1985.
- [3] K. S. Lam, *Topic in Contemporary Mathematical Physics*, World Scientific, 2003.
- [4] L. L. Salcedo, *Grupos continuos*, <http://www.ugr.es/local/salcedo/public/mt3/curso.pdf>
- [5] W-K. Tung, *Group Theory in Physics*, World Scientific, 1985.
- [6] N. Madras, *Lectures on Monte Carlo Methods*, The Fields Institute for Research in Mathematical Sciences, American Mathematical Society, 2002.
- [7] J. I. Illana, *Métodos Monte Carlo*, <http://www.ugr.es/local/jillana/Docencia/FM/mc.pdf>
- [8] K. Rummukainen, *Monte Carlo simulation methods*, http://www.helsinki.fi/~rummukai/lectures/montecarlo_oulu/lectures/mc_notes1.pdf
- [9] S. Weinzierl, *Introduction to Monte Carlo methods*, <http://arxiv.org/pdf/hep-ph/0006269.pdf>
- [10] L. L. Salcedo, *Integral de caminos*, <http://www.ugr.es/local/salcedo/public/mc2/curso.pdf>
- [11] <http://csrc.nist.gov/groups/ST/toolkit/rng/>
- [12] https://gcc.gnu.org/onlinedocs/gfortran/RANDOM_005fNUMBER.html
- [13] http://www.ugr.es/local/salcedo/public/fm/random_main.f
- [14] L. L. Salcedo, "Existence of positive representations for complex weights," *J. Phys. A* **40** (2007) 9399 [arXiv:0706.4359 [hep-lat]].