

Ondas

Ricardo A. Depine

Departamento de Física
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Versión: 15 de abril de 2026

(comprobar actualizaciones en <https://bit.ly/F2Depinev03>)

© 2026 Ricardo A. Depine. All rights reserved.

Índice general

1. Introducción	1
1.1. Las ondas como fenómeno unificador	1
1.2. ¿Qué es una onda?	2
1.3. Grados de libertad	4
1.4. Repaso de un GL	4
2. Del equilibrio a la perturbación	9
2.1. Dos grados de libertad: ejemplos introductorios	9
2.2. Dos grados de libertad: caso de estudio	10
2.2.1. Elección de coordenadas	11
2.2.2. Ecuaciones de movimiento	11
Comentario sobre acoplamiento	11
Comentario sobre homogeneidad	12
Posiciones de equilibrio	12
Transformación a un sistema homogéneo	13
2.3. Más allá del caso de estudio	13
2.4. Notación matricial	15
2.5. El club de los N grados de libertad	16
2.5.1. Tres grados de libertad: caso de estudio	16
Torques restitutivos	16
Partes móviles virtuales	17
Ecuaciones de movimiento	18
Propiedades de las matrices $\mathbb{M}$, $\mathbb{K}$ y $\mathbb{M}^{-1}\mathbb{K}$	18

2.5.2. De 3 a N : la cadena lineal	20
2.6. Vamos al laboratorio	21
3. Estructura detrás de los ritmos I: simetrías	27
3.1. La simetría que ya estaba allí	28
3.1.1. Cambiar de coordenadas	28
3.1.2. Ritmos propios y coordenadas normales	29
3.1.3. Selección de frecuencias mediante las condiciones iniciales	29
3.2. Modos normales	32
3.3. Lo que realmente hicimos	34
3.4. El punto de vista matricial	34
3.5. El método sistemático cuando $\tilde{\mathbb{K}}$ es simétrica	36
3.5.1. Nota técnica: amplitudes, fases y normalización	37
3.5.2. Ejemplo: procedimiento para el caso desacoplado «a mano»	38
3.5.3. Volvamos a mirar los resultados obtenidos en el laboratorio	39
4. Simetrías y asimetrías	41
4.1. El método sistemático: caso general	41
4.2. El caso de estudio con masas y resortes distintos	43
4.3. Desacople modal completo	46
4.4. Condiciones iniciales y no ortogonalidad	47
4.5. Diagonalización y energía	49
4.6. Ejemplo integrador	51
4.6.1. Autovalores	52
4.6.2. Autovectores	52
4.6.3. Normalización y desarrollo modal	53
4.6.4. Condiciones iniciales	53
4.6.5. Regímenes de acoplamiento	54
Acoplamiento débil	55
Acoplamiento fuerte	56
4.6.6. Estructura modal e interpretación geométrica	56

4.7. Modulación y batidos	57
4.7.1. Batidos en un ejemplo: análisis e interpretación	58
4.7.2. Batidos en un ejemplo: simulaciones	60
4.7.3. Batidos y detectores cuadráticos	63
4.8. La perturbación que se propaga	64
5. Disipación y forzado	67
5.1. Disipación	68
5.1.1. Interacción con el medio y pérdidas internas	69
5.1.2. Estructura de $\mathbb{C}$	69
5.1.3. Condiciones de desacople	70
5.1.4. Regímenes de amortiguamiento	72
Interacción con el medio	73
Disipación interna y selección modal	75
5.2. Fuerzas externas y respuesta del sistema	77
5.2.1. Proyección modal de las fuerzas externas	77
5.2.2. Fuerzas externas armónicas	78
5.2.3. Respuesta modal de sistemas con varios grados de libertad	79
5.2.4. Ejemplos: respuesta en frecuencia y selección modal	80
6. Muchas partes móviles	85
6.1. Cadena con N cuentas idénticas equiespaciadas	86
6.1.1. Ecuaciones de movimiento para oscilaciones transversales	86
6.1.2. Modos normales	88
6.1.3. Relación de dispersión	89
6.1.4. Modos para dos extremos fijos	91
Condiciones iniciales	94
6.1.5. Condición para extremo libre	95
6.1.6. Oscilaciones longitudinales	95
6.2. Péndulos idénticos acoplados	96
6.3. Aproximación continua	98

6.3.1.	Cuerda $\Rightarrow$ ecuación de ondas clásica	99
6.3.2.	Slinky $\Rightarrow$ ecuación de ondas clásica	100
6.3.3.	Péndulos acoplados $\Rightarrow$ ecuación de Klein-Gordon	102
6.4.	La ecuación de ondas	102
6.5.	Modos normales en el medio continuo	103
6.6.	Condiciones en los extremos	104
6.7.	Cuerda con extremos fijos	105
6.7.1.	Discretización de los modos normales	105
6.7.2.	Condiciones iniciales	105
6.7.3.	Series de Fourier	107
	Coeficientes de la serie de Fourier impar	108
	Ejemplo: diente de sierra	110
	Ejemplo: pulso rectangular	111
	Condiciones iniciales	112
6.7.4.	Evolución temporal	113
6.8.	ondas estacionarias = suma de ondas viajeras	114
6.9.	Y en otros sistemas	116
7.	Ondas propagantes	121
7.1.	Ecuación clásica 1D	121
7.1.1.	Progresivas y regresivas 1D	121
7.1.2.	La solución más general 1D	122
7.1.3.	Fórmula de d'Alembert	123
7.2.	Ondas, medios y dimensiones	125
7.2.1.	La onda plana en un medio 3D	126
	La onda plana armónica en un medio 3D	127
7.2.2.	Fuentes con simetría esférica	127
7.2.3.	Fuentes con simetría cilíndrica	129
7.3.	Modulación	130
7.3.1.	Caso más sencillo	130

7.3.2. Modulaciones reales	132
7.3.3. Síntesis de señales	133
Mezclando N componentes armónicos	134
Límite $N \rightarrow \infty$, con $\Delta\omega$ constante	135
7.3.4. Transformada de Fourier	139
8. Respuesta forzada estacionaria	141
8.1. Forzado estacionario, cadena lineal	142
8.1.1. Rango dispersivo	143
8.1.2. Rango reactivo	143
8.2. Forzado estacionario, Klein-Gordon	144
8.2.1. Dispersivo vs reactivo	145
8.2.2. La parte con $e^{\kappa x}$ no está permitida en sistemas infinitos	146
8.2.3. Comentarios sobre el régimen reactivo	147
8.3. Descripción discreta	148
9. Descripción geométrica	151
9.1. Rayos	152
9.2. Leyes fenomenológicas	152
9.3. Algunas consecuencias de la ley de Ibn Sahl	154
9.3.1. Reflexión externa e interna	154
9.3.2. Reflexión total	154
9.4. La DG en otros marcos teóricos	155
9.4.1. Principio de Huygens	156
9.4.2. Principio de Fermat	160
9.5. Formación de imágenes	161
9.5.1. Dioptra plana	162
9.5.2. Dioptra esférica	163
9.5.3. Trazado de rayos	164
9.5.4. Lentes	165
9.6. La DG en el marco ondulatorio	167

9.6.1. Reflexión y transmisión de una onda plana en una superficie plana	168
9.6.2. Las condiciones de contorno imponen la cinemática de los rayos	169
9.6.3. Conservación de componentes tangenciales de $\vec{k}$	171
9.6.4. La reflexión total en el marco ondulatorio	172
10. Ondas transversales	175
10.1. Ondas planas armónicas transversales	175
10.2. Polarización lineal	176
10.3. Polarización elíptica	177
10.4. Polarización circular	178
10.5. Caso ϕ arbitrario	179
10.5.1. Luz natural	180
10.6. Polarización por absorción selectiva	180
10.7. Polarización por reflexión	181
10.8. Láminas retardadoras	186
10.9. Birrefringencia en medios quirales	187
10.10. Birrefringencia en cristales	188
11. Interferencia	193
11.1. Fuentes puntuales	194
11.2. Uso de fasores	196
11.3. Sistemas de franjas	197
11.4. Experiencia de Young	198
11.5. Coherencia	199
11.6. Interferómetros	200
11.6.1. División del frente de onda	201
11.6.2. División de amplitud	201
Interferómetro de Michelson	202
11.7. Interferencia en láminas	203
11.8. Localización de franjas	204

12.Difracción	205
12.1. Interferencia de N fuentes	205
12.1.1. Caso general	206
12.1.2. Fuentes alineadas y equiespaciadas	207
12.1.3. Intensidad en la zona lejana	210
12.1.4. Intensidad en la zona cercana	214
12.1.5. Experimentos numéricos	215
12.1.6. Límite continuo, $N \rightarrow \infty$, $d \rightarrow 0$	218
12.2. Difracción de Fresnel y de Fraunhofer	220
12.3. Difracción de Fraunhofer	221
12.4. Rendija de ancho b	222
12.5. Obstáculos planos con simetría de translación	223
12.5.1. Otra vez una rendija	224
12.5.2. N rendijas idénticas equiespaciadas	226
12.6. Obstáculos planos, caso general	229
12.6.1. Abertura rectangular	231
12.6.2. Abertura circular	231
12.7. Óptica de Fourier	231
Bibliografía	247

Capítulo 1

Introducción

1.1. Las ondas como fenómeno unificador

Las ondas están presentes en innumerables sistemas y procesos físicos. Desde las ondas mecánicas en una cuerda hasta las recientemente detectadas ondas gravitacionales (una nueva herramienta del siglo XXI para conocer el cosmos), pasando por el sonido y la luz (profundamente vinculados a nuestros sentidos), las ondas constituyen una forma fundamental de transmisión de información y energía en el universo.

La belleza de la teoría de las ondas radica en su carácter *unificador*: fenómenos muy distintos pueden describirse mediante estructuras matemáticas similares. Aunque cada tipo de onda se «viste» con un origen físico particular, muchas de sus propiedades —como la velocidad de propagación, la interferencia o la difracción— son universales. Esta universalidad es la razón principal para estudiar ondas en general: comprenderlas permite abordar fenómenos complejos y desarrollar modelos aplicables en diversos campos de la ciencia.

Como punto de partida nos enfocaremos en sistemas mecánicos, que permiten un tratamiento directo a partir de primeros principios: las leyes de Newton. Comenzaremos con sistemas de varias partes móviles que interactúan entre sí y estudiaremos cómo se comportan al ser apartados de su posición de equilibrio. Usaremos las mismas leyes físicas que aplican a sistemas más sencillos, como un péndulo o una masa atada a un resorte. Desarrollaremos herramientas para manejar las dificultades matemáticas que aparecen al aumentar el número de partes móviles y luego utilizaremos una aproximación muy útil para describir sistemas con muchísimas partes móviles muy próximas entre sí: la idealización del límite continuo.

Para estudiar otros tipos de ondas recurriremos a *analogías*. El comportamiento de las ondas mecánicas servirá como guía para comprender ondas electromagnéticas, acústicas, sísmicas o gravitacionales. Diremos que si con las ondas mecánicas aparecía una propiedad característica —como la difracción o la interferencia—, entonces con las electromagnéticas (o acústicas, o sísmicas, etc.) también debería observarse, por el simple hecho de ser ondas.

Seguramente este enfoque pueda generar cierta inseguridad, especialmente si no se conocen en detalle las ecuaciones fundamentales del electromagnetismo, la acústica o la sismología. Sin embargo, la analogía es una herramienta clave en el aprendizaje y ha sido fundamental en numerosos avances científicos. Permite comprender lo desconocido a partir de lo conocido. La analogía es útil, pero tiene limitaciones que debemos reconocer. Por ejemplo, decir que «un avión es un auto que vuela» puede ser útil para quien nunca ha visto uno, aunque para entenderlo plenamente sea necesario observarlo y estudiarlo en detalle. Lo mismo ocurre con un cohete espacial, que podría describirse como «un avión sin alas».

Aquí veremos cómo las analogías ayudan a entender conceptos complejos, aunque también aprenderemos a reconocer sus límites y cuándo dejan de ser válidas. La física, en última instancia, es una ciencia experimental, y toda conclusión obtenida mediante analogías debe ser contrastada con la experiencia. Un ejemplo histórico lo encontramos en las ondas gravitacionales: en 1893 Oliver Heaviside sugirió su existencia por analogía con las ecuaciones de Maxwell en el electromagnetismo. Décadas más tarde, en 1915, Einstein las incluyó como consecuencia natural de su Teoría General de la Relatividad. Sin embargo, hubo que esperar más de un siglo para su detección experimental, confirmada en el siglo XXI, un logro que culminó con el Premio Nobel de Física en 2017.

1.2. ¿Qué es una onda?

Como la palabra «onda» tiene varios significados en la vida cotidiana, empezaremos con una definición física minimalista. A continuación discutiremos los conceptos básicos de esta definición para que entren en escena otros conceptos que, si bien están implícitos en la definición minimalista, conviene tener presentes y suelen mencionarse en definiciones más verborágicas.

Según la definición minimalista, «una onda es una perturbación que se propaga».

Hablemos de *perturbación*. Para sistemas mecánicos compuestos por muchas partes móviles, la perturbación de la que estamos hablando es de alguna magnitud física (por ejemplo velocidad, posición, densidad, presión, etc.) asociada a cada parte móvil del sistema. Y cuando hablamos de perturbación, está implícito un estado de referencia -por lo general un estado de equilibrio- donde la perturbación de la magnitud física vale cero.

En general el *sistema* (o medio perturbado) puede ser de naturaleza muy diversa, como masas unidas por resortes, la membrana de un tambor, un líquido dentro de un recipiente, el aire en un tubo de órgano, la placa metálica de un vibráfono, la madera de un xilófono o los campos electromagnéticos en una fibra óptica. La definición dada incluye situaciones donde el medio puede no ser un medio material en el sentido convencional. Por ejemplo, la luz de una estrella se propaga en el espacio interestelar, que está mayormente vacío. O las distorsiones en la geometría del espacio-tiempo producidas por el movimiento orbital de las estrellas de un sistema binario se propagan como

ondas gravitacionales a través del universo.

Hablemos de *se propaga*. La perturbación se propaga a través del sistema, pero el sistema como un todo no se propaga. Por este motivo, es usual ampliar la definición minimalista anterior y explicitar que en una onda no hay transporte neto de materia. Por eso también se dice que una onda es «una perturbación que se propaga y que transporta energía pero no materia». Y como hay que saber parafrasear, también podríamos decir que «llamamos onda a la propagación de una perturbación, con transporte de energía pero sin transporte de materia».

Para que haya propagación, el medio debe tener cierta *extensión* o estar formado por varias partes móviles. Por ejemplo, el término «efecto dominó» alude a la caída sucesiva de fichas colocadas en fila: una perturbación inicial aplicada a la primera se transmite a las siguientes, generando una cascada de movimientos que recorre todo el sistema. Este ejemplo también pone en evidencia la necesidad de una *interacción* entre las partes móviles del sistema: la caída de una ficha debe provocar la caída de la ficha contigua, de lo contrario la propagación se interrumpe y el efecto desaparece.

En el ejemplo de las fichas de dominó, sin ningún mecanismo restitutivo, el apartamiento del equilibrio inicial (fichas que caen desde la posición vertical) —si bien se propaga— conduce inevitablemente a un nuevo estado de equilibrio final (fichas en la posición horizontal). Este comportamiento resulta ilustrativo para mostrar una propagación en cadena, pero no es adecuado para empezar a estudiar las ondas, ya que una vez que cae la última ficha el sistema queda detenido y sería necesario reponer manualmente las fichas en posición vertical para repetir el proceso.

En contraste, cuando el sistema parte de un equilibrio estable la situación es muy diferente. En este caso, el sistema se encuentra en el fondo de un pozo de potencial generado por fuerzas conservativas. Ante perturbaciones pequeñas, dichas fuerzas actúan de manera restitutiva, devolviendo al sistema a su posición inicial. Precisamente esta propiedad es la que permite la propagación de ondas en medios materiales: cada parte del sistema se aparta momentáneamente de su equilibrio, pero las fuerzas conservativas aseguran que el movimiento sea oscilatorio y que la perturbación pueda repetirse y transmitirse de manera continua.

Es conveniente recordar que bajo circunstancias adecuadas, el pozo de potencial asociado a un equilibrio estable puede aproximarse localmente a una parábola. En ese caso, el movimiento resultante es armónico: las fuerzas conservativas generan una restauración proporcional al desplazamiento respecto de la posición de equilibrio, y el sistema oscila de manera periódica alrededor de dicho mínimo. Este comportamiento es característico de numerosos sistemas físicos, como el resorte en la ley de Hooke o el péndulo simple para pequeñas oscilaciones.

En resumen, a partir de la definición minimalista (perturbación que se propaga), hemos podido introducir una serie de conceptos fundamentales. En primer lugar, la necesidad de un estado de referencia para la perturbación, que permite distinguir entre equilibrio inicial y desviaciones respecto de él. En segundo lugar, la importancia de que el sistema sea extenso o esté compuesto por partes móviles, de modo que la perturbación pueda recorrerlo. Además, distinguimos entre equilibrio estable e inestable, lo

que determina si existen fuerzas restitutivas capaces de devolver el sistema a su estado inicial. Finalmente, vimos que la presencia o ausencia de mecanismos restitutivos es decisiva para que la propagación pueda sostenerse en el tiempo y dar lugar a fenómenos ondulatorios.

1.3. Grados de libertad

Para describir la dinámica de un sistema físico necesitamos elegir un conjunto adecuado de coordenadas. ¿Cuántas son necesarias? Ni más ni menos que las indispensables para especificar cómo cambia la posición de cada parte con el tiempo. El número mínimo de coordenadas escalares que permiten determinar la posición de todas las partes móviles se denomina *grados de libertad* (GL).

Como analizaremos en el capítulo 2, el número de GL está íntimamente relacionado con el tipo de vibraciones que pueden manifestarse en sistemas mecánicos con varias partes móviles.

Una partícula en movimiento libre tiene tres GL, ya que para dar su posición en el espacio en cada instante se necesitan tres funciones del tiempo. En cambio, una partícula que cuelga del extremo de una cuerda tensa solo tiene dos GL, pues basta con dos funciones del tiempo independientes: las coordenadas que describen la proyección de su trayectoria sobre el plano horizontal.

Un conjunto de N partículas interactuantes que se mueven sin restricciones en el espacio tridimensional tiene $3N$ GL. Si el conjunto de partículas se mueve en una superficie, el número de GL es $2N$ y si se mueve a lo largo de una curva es N .

Un sólido rígido tiene seis GL. Por ejemplo, para un cubo se necesitan tres funciones del tiempo independientes para describir la posición de su centro de masa (las coordenadas cartesianas x, y, z) y otras tres para describir su orientación (en términos de ángulos de rotación alrededor de cada uno de los ejes x, y, z).

Cuando existen ligaduras entre las partículas, el número de GL de un sistema será el número de GL del sistema sin ligaduras, menos el número de ligaduras que relacionan las variables. Por eso, del primer ejemplo (partícula libre) al segundo ejemplo (partícula que cuelga del extremo de una cuerda), el número de GL pasa de tres a dos, porque la cuerda —considerada como una varilla rígida de masa despreciable— actúa como ligadura.

1.4. Repaso de un GL

Antes de adentrarnos en sistemas con muchos GL, conviene recordar algunos aspectos básicos de la mecánica elemental. En los cursos universitarios se discuten sistemas físicos sencillos, como masas puntuales o cuerpos rígidos, con vínculos que permiten analizar un solo GL. Ejemplos clásicos son el péndulo o el resorte [1].

Tal como se discutió en la sección 1.2, para poner en evidencia las ondas mecánicas estudiaremos la evolución temporal de perturbaciones alrededor de un estado de equilibrio estable en sistemas con muchas partes móviles, que interactúan entre sí mediante fuerzas restitutivas. En otras palabras, saldremos del espacio de confort de la mecánica de un GL. Y aunque nos adentremos en nuevos territorios, encontraremos que hay un lenguaje común y resultados que generalizan los ya vistos.

A continuación repasaremos brevemente algunos resultados de la mecánica elemental —como los presentados en [1]— que asumiremos conocidos y que servirán de base para el estudio de sistemas más complejos:

1. **Equilibrio estable e inestable.** En sistemas conservativos, la naturaleza del equilibrio depende de las condiciones sobre el potencial en el punto de equilibrio: un mínimo corresponde a equilibrio estable, un máximo a equilibrio inestable.
2. **Movimiento en equilibrio estable.** Una perturbación suficientemente pequeña de un estado de equilibrio estable produce un movimiento limitado en el espacio. O sea, si el sistema permanece dentro del pozo de potencial, se mueve alrededor de la posición de equilibrio.
3. **Oscilador armónico.** Si la fuerza restitutiva es lineal (potencial cuadrático), el movimiento es oscilatorio armónico y la frecuencia característica ω_0 satisface

$$\omega_0^2 = \frac{F_{\text{restit.}}}{m \Delta x},$$

es decir que el cuadrado de la frecuencia característica resulta igual a la fuerza restitutiva por unidad de desplazamiento y por unidad de masa. Este resultado, discutido en el libro de ondas de Crawford (Capítulo 1) [2, 3], ofrece un criterio físico muy claro para comparar frecuencias de distintos movimientos armónicos, incluso cuando su fenomenología es diferente. Permite desarrollar una intuición sobre por qué ciertos sistemas oscilan más rápido que otros: cuanto mayor es la rigidez relativa al valor de la masa, mayor será la frecuencia. Además, esta relación se generaliza de manera natural al analizar movimientos armónicos especiales en sistemas con múltiples grados de libertad, proporcionando una base conceptual para entender cómo surgen diferentes frecuencias características en configuraciones más complejas.

4. **Condición de pequeñas perturbaciones.** Cuando la perturbación del equilibrio estable es suficientemente pequeña, el potencial puede aproximarse localmente a una parábola. En ese caso, el movimiento es armónico simple con frecuencia ω_0 .
5. **Oscilador armónico amortiguado.** La ecuación diferencial de segundo orden, lineal y homogénea con coeficientes constantes describe la evolución temporal de un oscilador con disipación. La solución combina una dependencia periódica (cuando el amortiguamiento es débil) con un decaimiento exponencial de la amplitud. Los parámetros relevantes son la frecuencia natural ω_0 , el coeficiente de amortiguamiento γ y la frecuencia amortiguada $\omega' = \sqrt{\omega_0^2 - \gamma^2}$. Según la relación entre γ y ω_0 se distinguen tres regímenes:

- **Subamortiguado** ($\gamma < \omega_0$): oscilaciones con frecuencia reducida y amplitud decreciente.
 - **Críticamente amortiguado** ($\gamma = \omega_0$): retorno más rápido al equilibrio sin oscilaciones.
 - **Sobreamortiguado** ($\gamma > \omega_0$): retorno lento al equilibrio sin oscilaciones.
6. **Magnitudes características.** Amplitud, frecuencia, período, fase inicial, velocidad, aceleración, energía cinética, energía potencial, energía mecánica total y tiempo de decaimiento son parámetros fundamentales para describir el movimiento.
7. **Representaciones equivalentes.** La solución del oscilador armónico puede escribirse en distintas formas matemáticas:
- $x(t) = A \cos(\omega t + \varphi_A)$
 - $x(t) = B \sin(\omega t + \varphi_B)$
 - $x(t) = C \cos(\omega t) + D \sin(\omega t)$
 - $x(t) = \Re a \exp(\pm i\omega t)$ ($\Re$: parte real)

Todas estas expresiones son equivalentes y contienen la misma información física: amplitud, frecuencia y fase inicial. Es conveniente que el lector se plantee cómo pasar de una forma a otra y cómo obtener amplitudes y fases en cada caso. Por ejemplo, a partir de los coeficientes C y D se puede reconstruir la amplitud $A = \sqrt{C^2 + D^2}$ y la fase $\varphi = \arctan(D/C)$.

8. **Oscilador armónico forzado.** Es análogo al oscilador simple, donde las únicas fuerzas son las restitutivas, pero ahora se añade una fuerza exterior periódica cuya frecuencia no depende del sistema, sino de una fuente externa. La solución general se compone de dos partes:
- **Transitoria:** corresponde a la solución homogénea, depende de las condiciones iniciales y decae con el tiempo.
 - **Estacionaria:** es una solución particular de la ecuación no homogénea, no depende de las condiciones iniciales y oscila con la frecuencia de la fuerza externa.

En este contexto, se introduce el fenómeno de **resonancia**, cuando la frecuencia de la fuerza externa se aproxima a la frecuencia natural del sistema se produce una amplificación notable de la respuesta.

- En el caso ideal sin amortiguamiento, la amplitud crece indefinidamente (matemáticamente diverge).
- Con amortiguamiento, la amplitud máxima es finita y ocurre a una frecuencia ligeramente menor que ω_0 . Este resultado es fundamental para comprender la respuesta real de sistemas físicos y se generaliza en el estudio de modos normales.

9. **Principio de superposición.** En sistemas lineales, la suma de soluciones también es solución, lo que permite construir movimientos más complejos a partir de soluciones simples.

Capítulo 2

Del equilibrio a la perturbación

2.1. Dos grados de libertad: ejemplos introductorios

Para situarnos en el escenario más simple compatible con la definición de onda, prestaremos atención a la dinámica de sistemas mecánicos con dos partes móviles interactuantes que son apartados del equilibrio. Si cada parte tiene un grado de libertad, es claro que tales sistemas tienen dos grados de libertad (GL).

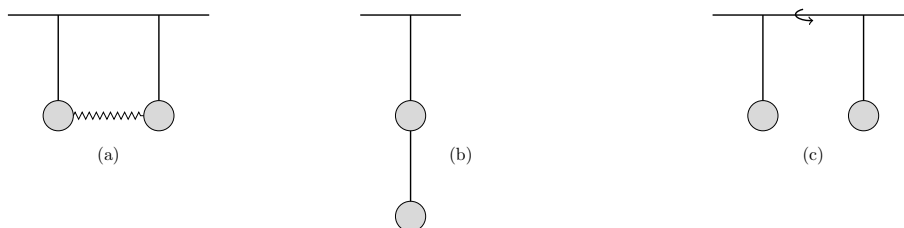


Figura 2.1: Sistemas de péndulos acoplados con dos grados de libertad.

En la figura 2.1 se muestran tres sistemas que cumplen con estas condiciones. En los tres sistemas interviene la gravedad como fuerza restitutiva. El sistema de la figura 2.1a consiste en dos péndulos simples acoplados mediante un resorte transversal. Cada péndulo oscila en el plano del dibujo y el resorte introduce el acoplamiento entre sus movimientos. La fuerza restitutiva tiene entonces dos aportes, la gravedad y la fuerza elástica del resorte. El sistema de la figura 2.1b es un péndulo doble idealizado. Se modela como dos masas puntuales unidas en serie mediante barras rígidas sin masa. La primera masa está suspendida de un soporte fijo y puede oscilar en el plano del esquema. De ella cuelga la segunda masa, que también oscila en el mismo plano. El sistema de la figura 2.1c corresponde a dos péndulos de torsión formados por varillas rígidas suspendidas de un mismo soporte, cada varilla con una masa en su extremo. Las oscilaciones se producen en dirección perpendicular al plano del dibujo, de manera que el acoplamiento entre péndulos surge de la torsión del soporte común: cuando una varilla rota, el soporte transmite un momento de torsión al otro péndulo. En este caso, las fuerzas restauradoras provienen de torques (momentos de torsión) que dependen

de los ángulos que forman las varillas con el plano vertical definido por la posición de equilibrio. Análogamente a lo que se observa en un resorte lineal para la dependencia entre fuerza aplicada y desplazamiento, para pequeñas perturbaciones angulares, la dependencia entre torque y ángulo de desviación puede considerarse lineal.

2.2. Dos grados de libertad: caso de estudio

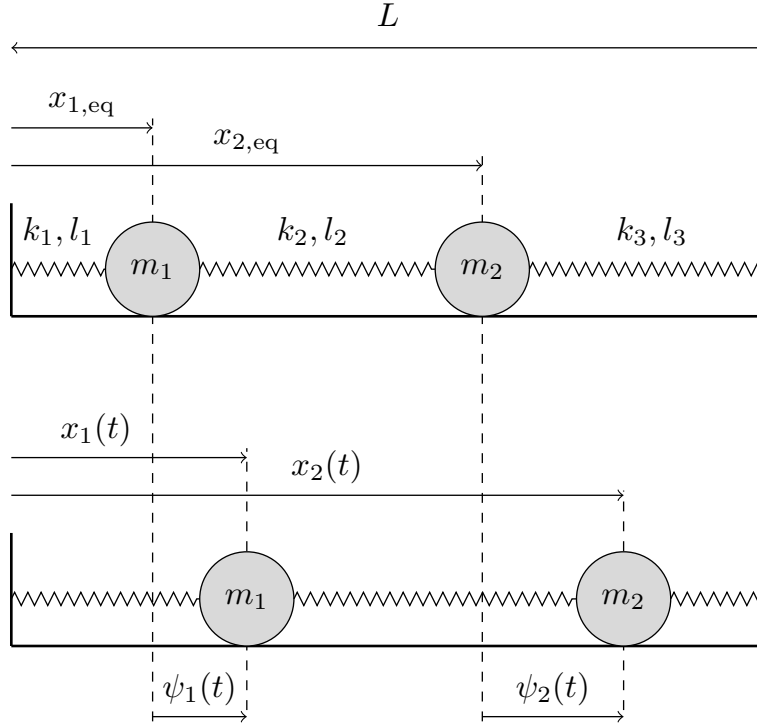


Figura 2.2: Las masas m_1 y m_2 interactúan entre sí y con las paredes a través de resortes. Panel superior: configuración de equilibrio. Panel inferior: fuera del equilibrio.

En lo que sigue vamos a limitarnos a considerar sistemas donde las fuerzas restitutivas son lineales. Si bien la física no lineal es un área sumamente rica y activa, con fenómenos propios como caos y bifurcaciones, abordar esos problemas requiere herramientas específicas que exceden el alcance de esta discusión inicial. Empezar con sistemas lineales permite obtener soluciones analíticas, construir intuiciones que luego sirven de base para enfrentar la no linealidad y desarrollar una notación clara.

En este contexto, elegir masas unidas por resortes ideales resulta especialmente conveniente porque la estructura lineal se hace evidente desde el inicio, en oposición a otros sistemas, como los que involucran péndulos, donde la linealidad surge de imponer la condición adicional de pequeñas perturbaciones alrededor del equilibrio.

Consideremos el sistema esquematizado en la figura 2.2: dos cuerpos de masas m_1 y m_2 que pueden moverse solamente en la dirección horizontal y que interactúan entre sí y con paredes rígidas a través de resortes ideales que siguen la ley de Hooke. La distancia total entre paredes es L y los resortes tienen constantes elásticas k_i y longitudes

naturales l_i ($i = 1, 2, 3$). Veremos que los resultados que obtengamos con este modelo son fácilmente generalizables a otros sistemas lineales con dos grados de libertad.

A continuación prestaremos atención a la elección de coordenadas adecuadas y a la identificación de las posiciones de equilibrio y perturbadas, un paso fundamental que permitirá simplificar las herramientas de análisis en todos estos escenarios.

2.2.1. Elección de coordenadas

La elección de coordenadas más natural e inmediata es medir las posiciones de las masas respecto de un punto fijo externo. Aquí elegimos la pared izquierda (ver figura 2.2, panel inferior), de modo que $x_1(t)$ y $x_2(t)$ representen directamente las posiciones absolutas de m_1 y m_2 . Esta elección es la que suele hacerse en primera instancia, pues refleja de manera intuitiva cómo se observa físicamente el sistema.

Sin embargo, resulta más conveniente trabajar con coordenadas $\psi_1(t)$ y $\psi_2(t)$ que miden las desviaciones de cada masa respecto su posición de equilibrio (ver figura 2.2, panel inferior)

$$\begin{aligned}\psi_1(t) &= x_1(t) - x_{1,\text{eq}}, \\ \psi_2(t) &= x_2(t) - x_{2,\text{eq}}.\end{aligned}\tag{2.1}$$

Veremos a continuación que las ecuaciones de movimiento escritas con las coordenadas $\psi_1(t)$ y $\psi_2(t)$ adoptan una forma más sencilla que con las coordenadas $x_1(t)$ y $x_2(t)$.

2.2.2. Ecuaciones de movimiento

Analicemos las fuerzas de los resortes sobre las partes móviles (ley de Hooke) en las coordenadas $x_1(t)$ y $x_2(t)$. Sobre la parte móvil 1 actúan dos fuerzas: la fuerza del resorte de la izquierda, $-k_1(x_1 - l_1)$, y la fuerza del resorte del centro, $k_2[(x_2 - x_1) - l_2]$. Una vez analizadas las fuerzas, la ecuación de Newton de la parte móvil 1 se obtiene pidiendo que la fuerza neta sea igual a la masa por la aceleración, es decir

$$m_1 \ddot{x}_1 = -k_1(x_1 - l_1) + k_2[(x_2 - x_1) - l_2]\tag{2.2}$$

Análogamente, hay dos fuerzas que actúan sobre la parte móvil 2: la fuerza del resorte de la derecha, $k_3[(L - x_2) - l_3]$ y la fuerza del resorte del centro, $-k_2[(x_2 - x_1) - l_2]$. La ecuación de Newton de la parte móvil 2 queda

$$m_2 \ddot{x}_2 = -k_2[(x_2 - x_1) - l_2] + k_3[(L - x_2) - l_3].\tag{2.3}$$

Comentario sobre acoplamiento

Las ecuaciones 2.2 y 2.3 forman un sistema de ecuaciones diferenciales ordinarias de segundo orden (porque involucran aceleraciones), lineal (porque las ecuaciones no contienen términos cuadráticos ni de orden superior en las incógnitas) y *acopladas* (porque el movimiento de cada parte móvil depende también de la otra). Respecto a

sistemas con un grado de libertad, acá la novedad es el acoplamiento: las dos coordenadas dinámicas están acopladas entre sí y ya no podemos estudiar cada parte móvil por separado.

Dicho de otra manera, esperamos soluciones que involucren *comportamientos colectivos* o *rasgos emergentes* del sistema en su conjunto, en los que se ponen de manifiesto fenómenos imposibles de entender desde el comportamiento aislado de cada parte móvil, y que emergen únicamente de la interacción entre ellas. Este tipo de fenómenos recuerda a las murmuraciones de aves —patrones colectivos que surgen en bandadas de estorninos y otras especies—, donde el movimiento coordinado no se explica por la acción individual de cada animal, sino por la interacción local que da lugar a un orden global sorprendente.

Para tratar el acoplamiento, vamos a necesitar herramientas matemáticas nuevas. Aunque todavía no explicaremos cómo se resuelve el sistema, conviene señalar que la situación recuerda a los sistemas algebraicos acoplados de N ecuaciones con N incógnitas, donde ninguna ecuación puede resolverse de manera aislada sin considerar las demás.

Comentario sobre homogeneidad

Un sistema de ecuaciones diferenciales lineales se denomina *homogéneo* cuando todas las ecuaciones están expresadas únicamente en función de las incógnitas y sus derivadas, sin términos adicionales. En cambio, se denomina *inhomogéneo* cuando aparecen términos constantes o funciones que actúan como fuerzas externas.

En nuestro caso, las ecuaciones de Newton obtenidas para $x_1(t)$ y $x_2(t)$ incluyen términos constantes asociados a las longitudes naturales de los resortes. Si pasamos estos términos al lado derecho, el sistema formado por las ecuaciones 2.2 y 2.3 se exhibe explícitamente como inhomogéneo

$$\begin{aligned} m_1\ddot{x}_1 + k_1x_1 + k_2(x_1 - x_2) &= k_1l_1 - k_2l_2, \\ m_2\ddot{x}_2 + k_2(x_2 - x_1) + k_3x_2 &= k_2l_2 + k_3(L - l_3). \end{aligned} \quad (2.4)$$

De este modo, se ve claramente que el sistema escrito en las coordenadas «absolutas» $x_1(t)$ y $x_2(t)$ es un sistema lineal *inhomogéneo*.

Posiciones de equilibrio

La presencia de inhomogeneidades está directamente relacionada con la existencia de posiciones de equilibrio no nulas: configuraciones estáticas en las que las fuerzas se compensan y las aceleraciones y velocidades se anulan.

En nuestro caso, las posiciones de equilibrio $x_{1,\text{eq}}$ y $x_{2,\text{eq}}$ se obtienen haciendo $\ddot{x}_1 = 0$ y $\ddot{x}_2 = 0$ en el sistema 2.4

$$\begin{aligned} k_1x_{1,\text{eq}} + k_2(x_{1,\text{eq}} - x_{2,\text{eq}}) &= k_1l_1 - k_2l_2, \\ k_2(x_{2,\text{eq}} - x_{1,\text{eq}}) + k_3x_{2,\text{eq}} &= k_2l_2 + k_3(L - l_3). \end{aligned} \quad (2.5)$$

Las posiciones de equilibrio dependen de las constantes elásticas k_1, k_2, k_3 , las longitudes naturales l_1, l_2, l_3 y la distancia fija entre paredes L . Si los tres resortes son idénticos, por razones de simetría debe ser $x_{1,\text{eq}} = L/3$ y $x_{2,\text{eq}} = 2L/3$, independientemente de la masa de las partes móviles.

Transformación a un sistema homogéneo

Una vez identificadas las posiciones de equilibrio, podemos usar la transformación 2.1 para expresar las ecuaciones de movimiento en las nuevas coordenadas $\psi_1(t)$ y $\psi_2(t)$. Al hacer la transformación, los términos constantes desaparecen y el sistema diferencial inhomogéneo 2.4 se transforma en el siguiente sistema homogéneo

$$\begin{aligned} m_1 \ddot{\psi}_1 &= -k_1 \psi_1 + k_2 (\psi_2 - \psi_1), \\ m_2 \ddot{\psi}_2 &= -k_2 (\psi_2 - \psi_1) - k_3 \psi_2. \end{aligned} \tag{2.6}$$

No hace falta hacer las cuentas detalladas para este ejemplo particular, porque todo sistema lineal con inhomogeneidad constante puede reducirse a un sistema homogéneo mediante un cambio de variables que traslada el origen al equilibrio. La demostración es sencilla: procedamos por el absurdo y supongamos que al medir las coordenadas como desviaciones del equilibrio el sistema resultara no homogéneo. Esto implicaría que en el equilibrio las nuevas coordenadas $\psi_1(t)$ y $\psi_2(t)$ no se anulan (porque estos valores serían solución de un sistema algebraico no homogéneo, análogo al sistema 2.5), lo cual contradice la propia definición de estas coordenadas como desviaciones del equilibrio. Por lo tanto, salvo situaciones patológicas —como sistemas sin equilibrio estático o con infinitas posiciones de equilibrio debido a direcciones no restringidas—, todo sistema lineal con inhomogeneidad constante puede reformularse como homogéneo mediante este cambio de variables.

Notar que en las ecuaciones originales 2.4 aparecen explícitamente las constantes geométricas del problema, como las longitudes naturales l_i de los resortes y la distancia fija L entre los puntos de anclaje. Sin embargo, al redefinir las coordenadas como apartamientos respecto del equilibrio, esas constantes desaparecen de la descripción dinámica: lo que queda es un sistema lineal acoplado cuyos coeficientes están determinados únicamente por las fuerzas restitutivas (las constantes elásticas k_i en este ejemplo). Este procedimiento pone de manifiesto la conveniencia de trabajar con apartamientos respecto del equilibrio, ya que simplifica la forma de las ecuaciones y permite concentrarse en la dinámica esencial: las oscilaciones y el acoplamiento entre las coordenadas.

2.3. Más allá del caso de estudio

Nuestro caso de estudio es representativo de una clase mucho más amplia de problemas lineales con dos grados de libertad. Esto es así porque el apartamiento del equilibrio de cualquier sistema de dos grados de libertad y con fuerzas restitutivas linealizadas siempre puede describirse mediante ecuaciones diferenciales lineales, formalmente idé-

licas al sistema 2.6

$$\begin{aligned} m_1 \ddot{\psi}_1 &= -K_{11} \psi_1 - K_{12} \psi_2, \\ m_2 \ddot{\psi}_2 &= -K_{21} \psi_1 - K_{22} \psi_2, \end{aligned} \quad (2.7)$$

donde los coeficientes K_{ij} (con $i, j = 1, 2$) son constantes que reflejan la linealización. Para el sistema esquematizado en la figura 2.2 estos coeficientes toman los valores

$$\begin{aligned} K_{11} &= k_1 + k_2, & K_{12} &= -k_2, \\ K_{21} &= -k_2, & K_{22} &= k_2 + k_3. \end{aligned} \quad (2.8)$$

Es fácil comprobar que los valores concretos de los coeficientes K_{ij} para el sistema de dos péndulos idénticos esquematizado en la figura 2.1a, son

$$\begin{aligned} K_{11} &= \frac{mg}{L} + k, & K_{12} &= -k, \\ K_{21} &= -k, & K_{22} &= \frac{mg}{L} + k. \end{aligned} \quad (2.9)$$

donde cada péndulo tiene masa m , longitud L , k es la constante elástica del resorte que provee la interacción y se ha linealizado la fuerza restitutiva gravitatoria. Para

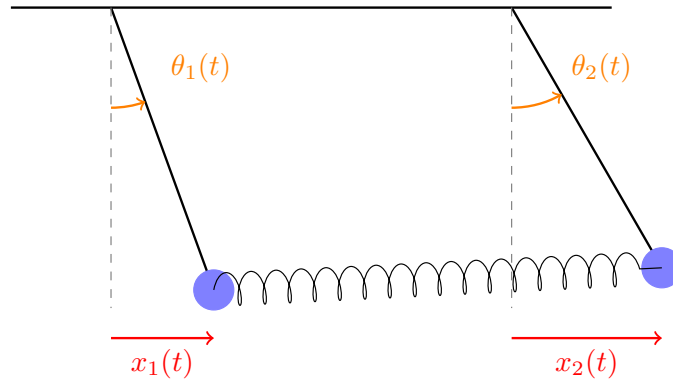


Figura 2.3: Dos péndulos acoplados por un resorte.

verlo basta con elegir las coordenadas angulares θ_1 y θ_2 que dan las posiciones de los péndulos, ya medidas desde el equilibrio, ver figura 2.3. Alternativamente, podríamos haber usado como coordenadas las proyecciones horizontales de las masas x_1 y x_2 , que en la hipótesis de pequeños desplazamientos del equilibrio resultan lineales con las coordenadas angulares $x_i \approx L\theta_i$, $i = 1, 2$. El sistema diferencial que surge del balance de fuerzas sobre cada parte móvil es

$$\begin{aligned} m\ddot{x}_1 &= -\frac{mg}{L}x_1 - k(x_1 - x_2), \\ m\ddot{x}_2 &= -\frac{mg}{L}x_2 - k(x_2 - x_1), \end{aligned} \quad (2.10)$$

que en forma ordenada se escribe

$$\begin{aligned} m\ddot{x}_1 &= -\left(\frac{mg}{L} + k\right)x_1 + kx_2, \\ m\ddot{x}_2 &= kx_1 - \left(\frac{mg}{L} + k\right)x_2. \end{aligned} \quad (2.11)$$

Este resultado no sólo muestra la equivalencia entre los dos casos particulares considerados, sino que establece una conclusión más general: todo sistema mecánico con 2GL y con fuerzas restitutivas lineales puede escribirse en la forma de la ecuación 2.7. En otras palabras, la estructura de las ecuaciones diferenciales es universal para este tipo de sistemas. Y la notación matricial resulta la herramienta natural para tratar con sistemas lineales y de paso compactar todo en una sola expresión.

2.4. Notación matricial

Para escribir las ecuaciones dinámicas en forma matricial definimos los vectores columna de coordenadas y de aceleraciones

$$\vec{\psi}(t) = \begin{bmatrix} \psi_1(t) \\ \psi_2(t) \end{bmatrix}, \quad \ddot{\vec{\psi}}(t) = \begin{bmatrix} \ddot{\psi}_1(t) \\ \ddot{\psi}_2(t) \end{bmatrix} \quad (2.12)$$

y las matrices $\mathbb{M}$ (matriz de masas) y $\mathbb{K}$ (matriz de coeficientes restitutivos)

$$\mathbb{M} = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}, \quad \mathbb{K} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}. \quad (2.13)$$

Con estas definiciones, el sistema se escribe como

$$\mathbb{M} \ddot{\vec{\psi}}(t) = -\mathbb{K} \vec{\psi}(t), \quad (2.14)$$

lo que equivale a

$$\ddot{\vec{\psi}}(t) = -\mathbb{M}^{-1}\mathbb{K} \vec{\psi}(t). \quad (2.15)$$

La matriz $\mathbb{M}^{-1}\mathbb{K}$ se conoce como *matriz dinámica* o *matriz del sistema*. Sus elementos tienen dimensiones de frecuencia al cuadrado, razón por la cual podríamos interpretarla como una extensión matricial del parámetro ω_0^2 que aparece en la ecuación diferencial del oscilador armónico unidimensional. En efecto, un oscilador armónico 1D cumple la ecuación

$$\ddot{x}(t) = -\omega_0^2 x(t)$$

y para el caso de un resorte tenemos $\omega_0^2 = k/m$. En esta interpretación formal, tanto la frecuencia ω_0^2 como la matriz $\mathbb{M}^{-1}\mathbb{K}$ juegan el mismo papel: ambos representan en un único objeto matemático la información sobre las relaciones entre masas y constantes restitutivas.

Si bien la notación es en parte una cuestión de gustos, en el contexto de sistemas acoplados la escritura matricial no es meramente estética: la notación matricial a) muestra claramente la estructura lineal del problema, b) permite capturar de manera compacta que el comportamiento es colectivo y que la dinámica ya no pertenece a cada parte móvil por separado, sino al conjunto completo y c) facilita aplicar técnicas de álgebra lineal que son naturales en este marco.

2.5. El club de los N grados de libertad

A esta altura nos damos cuenta que la formulación matricial desarrollada para el caso de 2GL se extiende sin cambios conceptuales a sistemas con un número mayor de coordenadas (N). La ecuación general

$$\mathbb{M} \ddot{\vec{\psi}}(t) + \mathbb{K} \vec{\psi}(t) = 0 \quad (2.16)$$

sigue siendo válida, con la diferencia de que ahora el vector columna $\vec{\psi}$ tiene N elementos y las matrices $\mathbb{M}$ y $\mathbb{K}$ tienen dimensión $N \times N$.

2.5.1. Tres grados de libertad: caso de estudio

A continuación consideramos un sistema de tres grados de libertad compuesto por tres varillas rígidas de longitud L , cada una de ellas con una masa puntual m_i en su extremo. Las varillas están suspendidas de un alambre cuya torsión genera el acoplamiento entre los péndulos. Las perturbaciones se producen en dirección perpendicular al plano de equilibrio, de modo que las oscilaciones corresponden a giros alrededor del eje del alambre (denominado «eje torsional»). Usaremos este caso de estudio no solo para adentrarnos en el manejo de la formulación matricial con coordenadas angulares, sino también para introducir una configuración particular que retomaremos más adelante. A diferencia del caso de estudio previo (2.2), aquí cada parte móvil está sometida a *dos fuerzas restitutivas* de distinta naturaleza, lo que da lugar a propiedades interesantes que enriquecerán el análisis. Además, aprovecharemos este escenario para destacar ciertos aspectos matemáticos que resultarán útiles al tratar el acoplamiento en el sistema de ecuaciones diferenciales y, de ese modo, resolver con rigor la evolución de las perturbaciones alrededor del equilibrio.

En la figura 2.4 se muestra la proyección de las posiciones perturbadas sobre un plano perpendicular al alambre. Las coordenadas angulares ψ_1 , ψ_2 y ψ_3 describen los apartamientos de cada masa m_i respecto de la vertical, y se toman positivas en sentido antihorario.

Fuera del equilibrio, el alambre entre péndulos se retuerce. Para el caso de la figura, el péndulo 2 se desplazó más que el 1, $\psi_2 > \psi_1$, el hilo entre ellos se retuerce y se genera un torque restaurador que intenta igualar los ángulos. El alambre retorcido actúa como un resorte torsional, como el de las pinzas con resorte, o “handgrips”, que se usan para fortalecer manos y antebrazos, o como una toalla mojada que se retuerce por los extremos para exprimirla.

Torques restitutivos

El torque restitutivo sobre cada péndulo no depende solo de su peso, sino también de la diferencia de posiciones angulares con los péndulos vecinos. En la aproximación de pequeños apartamientos angulares, también son pequeños los retorcimientos de los tramos de alambre. Y en este caso se pueden linealizar ambos torques restitutivos.

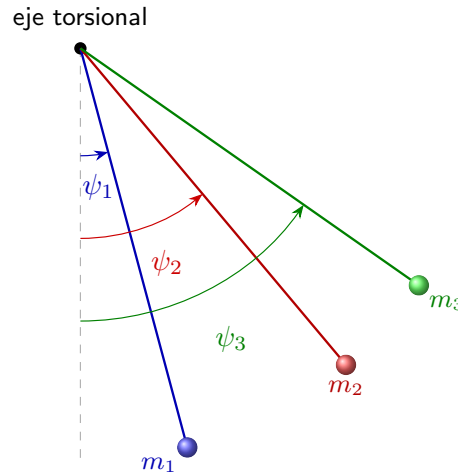


Figura 2.4: Configuración del tres péndulos con acoplamientos torsionales provistos por el soporte del que cuelgan. Las coordenadas angulares ψ_1 , ψ_2 y ψ_3 representan los apartamientos de las masas m_i respecto de la vertical, definidos positivos en sentido antihorario.

Analicemos los torques linealizados sobre el péndulo 2. El torque restitutivo gravitatorio es

$$-m_2 g L \sin \psi_2 \approx -m_2 g L \psi_2,$$

porque si $\psi_2 > 0$ (antihorario), la gravedad tira de la masa es sentido horario (negativo). El torque restitutivo torcional tiene dos aportes ¹: el ejercido por el tramo de alambre entre 2 y 3

$$+\kappa_{23}(\psi_3 - \psi_2),$$

(si $\psi_3 > \psi_2$, el alambre intenta hacer girar 2 hacia 3, sentido antihorario, es decir positivo). Y el ejercido por el tramo de alambre entre 1 y 2

$$\kappa_{12}(\psi_1 - \psi_2),$$

(si $\psi_1 < \psi_2$, el alambre intenta hacer girar 2 hacia 1, sentido antihorario, torque negativo). Las constantes de proporcionalidad κ_{12} y κ_{23} son las constantes de torsión de los tramos 1-2 y 2-3 de los alambres que soportan los péndulos (análogo rotacional de la constante de resorte lineal). El valor de las constantes de torsión depende del material y del grosor del alambre (un alambre más grueso tiene más resistencia al giro que uno más delgado), y de la distancia entre péndulos (cuanto más corto sea el tramo de alambre entre péndulos, más difícil será retorcerlo).

Partes móviles virtuales

Igualando el torque neto sobre 2 al producto de su aceleración angular ($\ddot{\psi}_2$) por el momento de inercia respecto al eje de rotación ($m_2 L^2$, consideramos que las varillas de

¹Para escribir bien los signos, la regla que no falla es «El alambre tira de mi masa hacia donde esté el vecino».

los péndulos no tienen masa), resulta

$$m_2 L^2 \ddot{\psi}_2 = -m_2 g L \psi_2 + \kappa_{12}(\psi_1 - \psi_2) + \kappa_{23}(\psi_3 - \psi_2). \quad (2.17)$$

Si a la izquierda del péndulo 1 hubiera otro péndulo (con etiqueta 0) también sostenido del mismo eje torsional, entonces la ecuación de movimiento para el péndulo 1 tendría la misma forma que la del péndulo 2

$$m_1 L^2 \ddot{\psi}_1 = -m_1 g L \psi_1 + \kappa_{01}(\psi_0 - \psi_1) + \kappa_{12}(\psi_2 - \psi_1),$$

con κ_{01} la constante de torsión del tramo de alambre 0-1. Análogamente, si a la derecha del péndulo 3 también hubiera otro péndulo (con etiqueta 4) sostenido del mismo eje torsional, la forma de la ecuación de movimiento para el péndulo 3 sería

$$m_3 L^2 \ddot{\psi}_3 = -m_3 g L \psi_3 + \kappa_{23}(\psi_2 - \psi_3) + \kappa_{34}(\psi_4 - \psi_3),$$

con κ_{34} la constante de torsión del tramo de alambre 3-4.

¡Pero a la izquierda del péndulo 1 y a la derecha del péndulo 3 no hay otros péndulos! Tanto el extremo izquierdo del tramo 0-1 como el extremo derecho del tramo 3-4 están atados a soportes rígidos. Sin embargo, y a riesgo de parecer demasiado fantasiosos, sigamos pensando que los péndulos 0 y 4 existen, pero que cada uno está incrustado en soportes rígidos a izquierda y derecha. ¿Hay diferencia entre la situación original que planteamos al comienzo y la situación que discutimos después, con péndulos *virtuales* 0 y 4 que no se pueden mover? Claramente no.

Vemos así que pensar en péndulos virtuales tiene sus ventajas, porque en lugar de tratar a los péndulos 1 y 3 como casos especiales «porque les falta un vecino», podemos aprovechar la ecuación general para un péndulo interno «con vecinos» (el péndulo 2) y adaptarla a la realidad imponiendo en ambos extremos virtuales *la condición de contorno de extremo fijo*

$$\psi_0 = 0 \quad \text{y} \quad \psi_4 = 0.$$

Ecuaciones de movimiento

Con las consideraciones anteriores, el sistema de ecuaciones dinámicas para el sistema de 3 péndulos torsionales acoplados resulta

$$\begin{aligned} m_1 L^2 \ddot{\psi}_1 &= -(m_1 g L + \kappa_{01} + \kappa_{12}) \psi_1 + \kappa_{12} \psi_2, \\ m_2 L^2 \ddot{\psi}_2 &= \kappa_{12} \psi_1 - (m_2 g L + \kappa_{12} + \kappa_{23}) \psi_2 + \kappa_{23} \psi_3, \\ m_3 L^2 \ddot{\psi}_3 &= -\kappa_{23} \psi_2 + (-m_3 g L + \kappa_{23}) \psi_3, \end{aligned} \quad (2.18)$$

que se puede escribir en la notación matricial compacta de la ecuación 2.15.

Propiedades de las matrices $\mathbb{M}$, $\mathbb{K}$ y $\mathbb{M}^{-1}\mathbb{K}$

La matriz de inercia $\mathbb{M}$ es diagonal, ya que cada masa aporta su momento de inercia $m_i L^2$ de forma independiente. Y como estos elementos diagonales son positivos, es una

matriz definida positiva. Parece un comentario físicamente descolgado, pero es una propiedad que garantiza que la energía cinética total del sistema sea siempre positiva y que habilita el uso de herramientas matemáticas asociadas con la matriz del sistema $\mathbf{M}^{-1}\mathbf{K}$.

En cuanto a las propiedades de $\mathbf{K}$

$$\mathbf{K} = \begin{bmatrix} m_1gl + \kappa_{01} + \kappa_{12} & -\kappa_{12} & 0 \\ -\kappa_{12} & m_2gl + \kappa_{12} + \kappa_{23} & -\kappa_{23} \\ 0 & -\kappa_{23} & m_3gl + \kappa_{23} + \kappa_{34} \end{bmatrix}, \quad (2.19)$$

observamos que se trata de una matriz *tridiagonal*: sólo tiene elementos en la diagonal principal y en las posiciones inmediatamente vecinas. Esto es así porque cada péndulo sólo interactúa directamente con sus vecinos inmediatos, no con todos los demás.

En los elementos de la diagonal aparece el efecto de la gravedad y la suma de las constantes de torsión de los alambres vecinos. Estos elementos son positivos y representan las fuerzas restitutivas y la tendencia de cada parte a volver al equilibrio. La suma de constantes de torsión se puede interpretar como una constante de torsión neta, como si cada péndulo estuviera sujeto por un único resorte torsional más duro.

La matriz $\mathbf{K}$ resulta simétrica porque el acoplamiento entre dos péndulos vecinos está gobernado por una misma constante de torsión κ_{ij} . Esa constante aparece en la ecuación de ambos péndulos con el mismo valor (aunque con signo contrario, según el sentido en que actúa la interacción). Los elementos fuera de la diagonal aparecen con signo negativo, lo cual refleja matemáticamente el efecto estabilizador del acoplamiento: estas contribuciones tienden a reducir las diferencias entre coordenadas y, en consecuencia, a mantener los ángulos de los péndulos vecinos próximos entre sí.

Para obtener la matriz del sistema hay que multiplicar $\mathbf{K}$ por la inversa de $\mathbf{M}$, lo que simplemente corresponde a dividir cada fila i de $\mathbf{K}$ por su respectivo m_iL^2 . La matriz resultante es:

$$\mathbf{M}^{-1}\mathbf{K} = \begin{bmatrix} \omega_0^2 + \frac{\kappa_{01} + \kappa_{12}}{m_1L^2} & -\frac{\kappa_{12}}{m_1L^2} & 0 \\ -\frac{\kappa_{12}}{m_2L^2} & \omega_0^2 + \frac{\kappa_{12} + \kappa_{23}}{m_2L^2} & -\frac{\kappa_{23}}{m_2L^2} \\ 0 & -\frac{\kappa_{23}}{m_3L^2} & \omega_0^2 + \frac{\kappa_{23} + \kappa_{34}}{m_3L^2} \end{bmatrix}. \quad (2.20)$$

Aquí ω_0 , definido por $\omega_0^2 = g/L$, representa la frecuencia natural de cada péndulo aislado, sin acoplamientos, es decir, en el caso límite en el que los alambres son completamente rígidos y no transmiten torsión entre vecinos. Esta frecuencia es un punto de referencia para analizar cómo la introducción de los acoplamientos modifica el comportamiento individual y da lugar a nuevos comportamientos colectivos.

Como $\mathbf{K}$ es simétrica, y si las masas son iguales, la matriz del sistema $\mathbf{M}^{-1}\mathbf{K}$ resulta simétrica. Y toda matriz real y simétrica es ortogonalmente diagonalizable: todos sus autovalores son reales y existe una base ortonormal de autovectores, incluso en casos de degeneración (es decir, cuando existen autovalores repetidos con multiplicidad mayor a uno).

Si las masas no son iguales, la matriz $\mathbf{M}^{-1}\mathbf{K}$ deja de ser simétrica, aun cuando $\mathbf{K}$ lo sea. Esto ocurre porque el producto entre la matriz de rigideces y la inversa de la matriz de masas no preserva en general la simetría.

A pesar de no ser simétrica, la matriz del sistema es *simetrizable*, porque puede escribirse como el producto de la inversa de una matriz simétrica definida positiva ($\mathbf{M}$) y otra simétrica ($\mathbf{K}$). Esta estructura garantiza que, aunque no simétrica, la matriz del sistema es *semejante* a una matriz simétrica. Las consecuencias de esto son relevantes: la matriz del sistema es diagonalizable y con autovalores reales.

Veremos en el próximo capítulo que estas propiedades serán esenciales para desarrollar la herramienta que permitirá desacoplar el movimiento y revelar los comportamientos colectivos del sistema.

2.5.2. De 3 a N : la cadena lineal

El análisis realizado para el sistema de tres péndulos torsionales sirve como modelo para comprender la estructura de sistemas similares con múltiples grados de libertad. A partir del ejemplo anterior podemos generalizar la formulación al caso de las oscilaciones transversales de una cadena lineal de N péndulos torsionales de longitud L , cada uno asociado a una masa m_i y acoplado a sus vecinos mediante constantes de torsión $\kappa_{i,i+1}$.

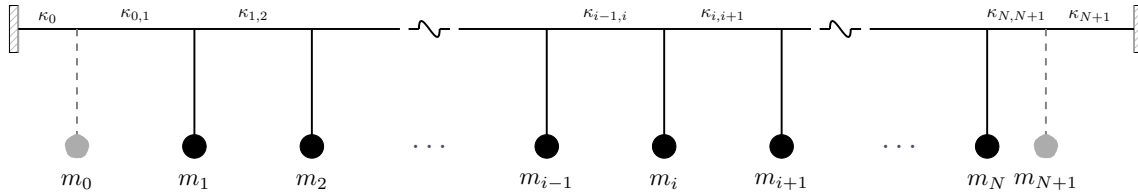


Figura 2.5: Configuración de equilibrio de una cadena de N péndulos torsionales acoplados con vecinos cercanos, con péndulos virtuales en los extremos. Las perturbaciones del equilibrio son transversales al plano del dibujo.

El término *cadena lineal* se utiliza porque los péndulos están dispuestos en una sucesión ordenada, uno tras otro, formando una *estructura unidimensional*. En este contexto, la posición de equilibrio y la identidad de cada parte móvil pueden describirse mediante una única coordenada cartesiana a lo largo de la dirección de la cadena, lo que simplifica la representación matemática del sistema. Por ejemplo, si tomamos el eje x paralelo al soporte, podemos identificar la parte móvil i -ésima como aquella que *en el equilibrio* se encuentra en la posición $x = x_i$. De este modo, cada elemento queda caracterizado por el valor de su coordenada en el equilibrio, en vez de por un número entero, lo que facilita la generalización a cadenas lineales con un número muy grande de partes móviles. Este modelo es análogo a otras cadenas lineales ampliamente estudiadas en la literatura, como las cadenas de masas unidas por resortes, las sogas con cuentas que oscilan transversal o longitudinalmente, y diversos sistemas discretos que sirven como aproximaciones de medios continuos. En todos estos casos, la linealidad de la

disposición y el acoplamiento local permiten aplicar métodos similares de análisis y obtener resultados comparables sobre los comportamientos colectivos.

Para evitar el tratamiento diferenciado de los extremos y unificar el tratamiento algebraico de todos los elementos de la cadena, introducimos dos partes móviles virtuales, etiquetadas con subíndices 0 y $N+1$. Esto permite modelar el sistema como una estructura de vecinos idénticos, desvincula la física de la interacción local de las restricciones globales de la cadena y otorga flexibilidad para imponer distintas condiciones de contorno de manera externa al desarrollo dinámico principal.

De manera análoga al análisis de la sección 2.5.1, el torque neto sobre la parte móvil i se compone del término gravitatorio y de la suma de los torques de los tramos izquierdo ($i-1$) y derecho ($i+1$):

$$\tau_{\text{neto},i} = -m_i gL \sin \psi_i + \underbrace{\kappa_{i-1,i}(\psi_{i-1} - \psi_i)}_{\text{Torque tramo izquierdo}} + \underbrace{\kappa_{i,i+1}(\psi_{i+1} - \psi_i)}_{\text{Torque tramo derecho}}, \quad i = 1, 2, \dots, N.$$

Las ecuaciones de movimiento se obtienen al igualar el torque neto con el producto del momento de inercia respecto al eje de rotación ($m_i L^2$) por la aceleración angular ($\ddot{\psi}_i$). En la aproximación de pequeños apartamientos del equilibrio queda

$$m_i L^2 \ddot{\psi}_i = m_i gL \psi_i + \kappa_{i-1,i}(\psi_{i-1} - \psi_i) + \kappa_{i,i+1}(\psi_{i+1} - \psi_i), \quad i = 1, 2, \dots, N. \quad (2.21)$$

y agrupando variables se obtiene

$$m_i L^2 \ddot{\psi}_i = -m_i gL \psi_i + \kappa_{i-1,i} \psi_{i-1} - (\kappa_{i-1,i} + \kappa_{i,i+1}) \psi_i + \kappa_{i,i+1} \psi_{i+1}, \quad i = 1, 2, \dots, N. \quad (2.22)$$

Este sistema diferencial lineal y acoplado se puede escribir en la forma matricial 2.16 y conserva un patrón común, independientemente de cuántas partes móviles se consideren: como en el ejemplo de 2.5.1, la matriz de masas (o de momentos de inercia) es diagonal y la matriz de rigidez es tridiagonal y simétrica

$$\mathbf{K} = \begin{bmatrix} d_1 & c_1 & 0 & \cdots & 0 \\ c_1 & d_2 & c_2 & \cdots & 0 \\ 0 & c_2 & d_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & c_{N-1} \\ 0 & \cdots & 0 & c_{N-1} & d_N \end{bmatrix}. \quad (2.23)$$

donde

$$d_j = m_j gL + \kappa_{j-1,j} + \kappa_{j,j+1}, \quad c_j = -\kappa_{j,j+1}. \quad (2.24)$$

Cuando las masas de los péndulos son distintas, la matriz del sistema $\mathbf{M}^{-1}\mathbf{K}$ no es simétrica. Sin embargo, y tal como se vió al final de la sección 2.5.1, es semejante a una matriz simétrica (y en consecuencia tiene autovalores reales y es diagonalizable).

2.6. Vamos al laboratorio

Luego de haber formalizado el lenguaje matricial y las propiedades matemáticas de sus matrices, estamos finalmente en condiciones de prestar atención a la naturaleza

dinámica de los problemas considerados. El tránsito del equilibrio a la dinámica de la perturbación $\vec{\psi}(t)$ marca un paso importante en nuestro recorrido. Aquí exploraremos si, al apartar el sistema de su estado de reposo, emergen esas perturbaciones que se propagan a las que llamamos ondas.

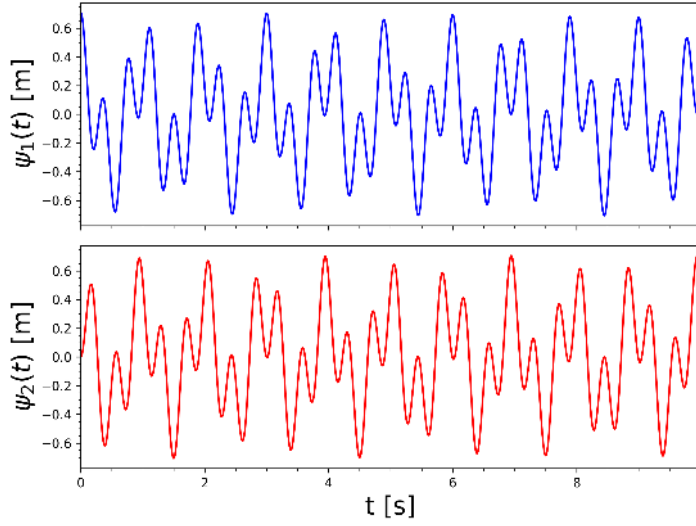
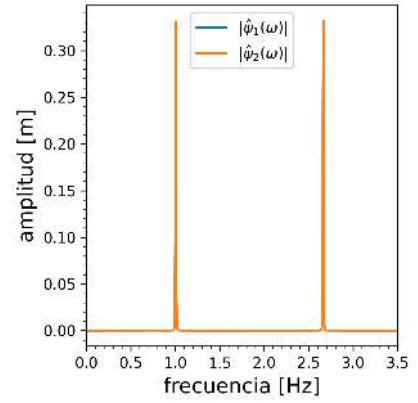
Si decidiéramos continuar únicamente por la vía matemática, el próximo desafío sería enfrentar de manera analítica el problema del acoplamiento. Lo recorreremos más adelante, cuando nos adentremos en un nuevo marco formal que exigirá un esfuerzo adicional. Por ahora, en cambio, vamos a detenernos en una etapa distinta: experimentar con simulaciones numéricas que nos permitan observar y comprender el fenómeno de manera directa, como si estuviéramos en el laboratorio. En estas simulaciones resolveremos las ecuaciones diferenciales numéricamente, con el objetivo de poner en evidencia los fenómenos antes de entrar en el formalismo. Al fin y al cabo, la física surge de la interacción constante entre formalismos y fenómenos. Al observar estas simulaciones, descubriremos que esta evolución temporal no solo confirma la propagación, sino que revela una complejidad inesperada: una danza de múltiples ritmos superpuestos que parecen esconder una estructura interna que los formalismos matemáticos deberán desentrañar en el capítulo siguiente.

Esta exploración no es un ejercicio puramente abstracto: los resultados que veremos son el reflejo fiel de experimentos que podríamos realizar hoy mismo en los laboratorios de la facultad. En este sentido, la *física computacional* actúa como un puente; nos permite asomarnos a la fenomenología de forma ágil y controlada, pero manteniendo siempre el rigor de saber que el experimento real es el juez final de cualquier formalismo que intentemos construir.

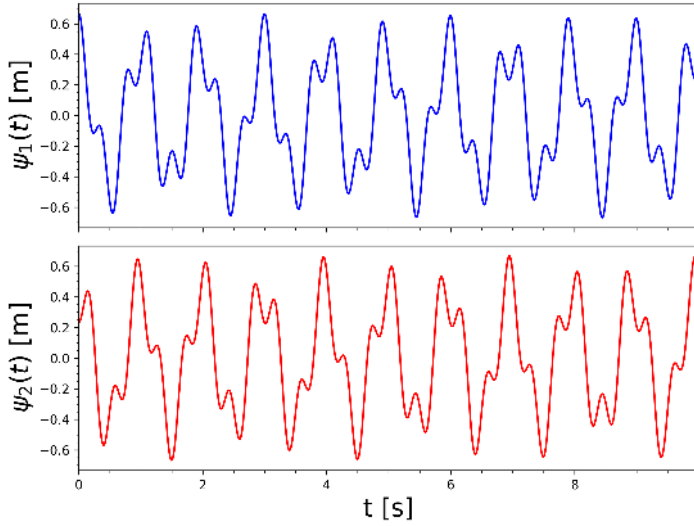
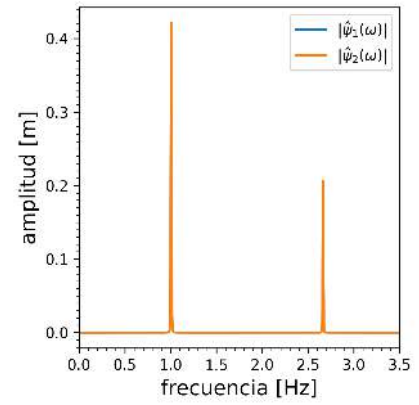
Hoy en día, la física computacional no es solo una extensión del cálculo manual, sino un pilar fundamental que nos permite asomarnos a sistemas de gran complejidad. En el ecosistema del lenguaje Python, contamos con bibliotecas consolidadas como `scipy.integrate`, que actúan como instrumentos de medición de alta precisión: tal como en el laboratorio usamos osciloscopios calibrados, aquí aprovechamos estas rutinas para resolver ecuaciones diferenciales con eficiencia, liberándonos de la carga mecánica del cálculo para centrarnos en la física del problema.

Del mismo modo, para analizar la «danza» de ritmos que observamos, contamos con un aliado matemático crucial: el *análisis espectral*. Pensemos, por ejemplo, en una aplicación de afinación en el celular: al tocar una nota, el dispositivo identifica la frecuencia dominante de la señal captada por el micrófono. La transformada de Fourier —implementada, por ejemplo, en herramientas como `numpy.fft`— realiza el núcleo matemático de ese procedimiento: toma una señal en el tiempo y la descompone en las frecuencias que la componen, indicándonos cuáles están presentes y con qué amplitud. Así, nuestras simulaciones dejan de ser simples gráficos para convertirse en experimentos virtuales donde podemos medir y comprender la estructura interna de cada componente de la perturbación $\vec{\psi}(t)$.

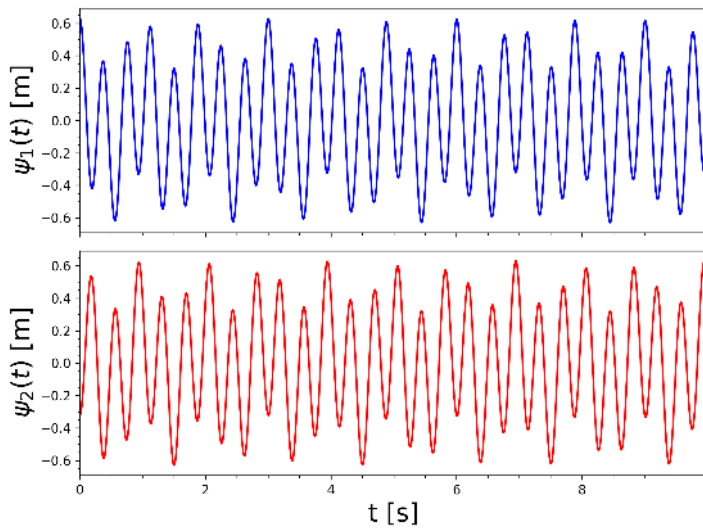
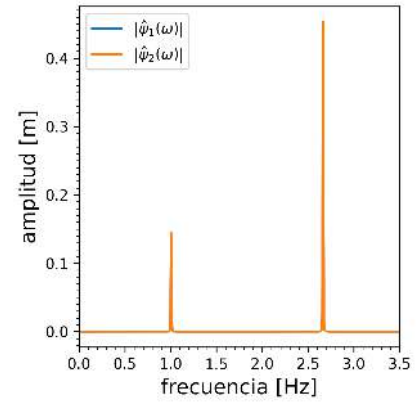
Con estas herramientas, veamos qué sucede cuando el equilibrio del sistema considerado en la sección 2.2 se perturba con distintas condiciones iniciales. Para el experimento numérico vamos a introducir cierta simetría. En particular, haremos $l_1 =$

(a) $\psi_1(0) = 0.7m$, $\psi_2(0) = 0$ 

(b)

(c) $\psi_1(0) = 0.67m$, $\psi_2(0) = 0.23m$ 

(d)

(e) $\psi_1(0) = 0.63m$, $\psi_2(0) = -0.32m$ 

(f)

Figura 2.6: Resultados del primer experimento numérico para tres condiciones iniciales distintas (ver detalles en el texto).

$l_2 = l_3 = L/3$ (en el equilibrio los tres resortes no están ni estirados ni comprimidos), $m_1 = m_2 = 1$ kg (las dos masas son iguales), $k_1 = k_3 = 40$ N/m (resortes de los extremos iguales, con elasticidad comparable a la de un resorte de acero disponible para demostraciones de laboratorio), $k_2 = 120$ N/m (un resorte más grueso o con las espiras más apretadas, tres veces más rígido que los exteriores). Esta simetrización del sistema no altera la esencia del problema, pero facilitará observar con claridad cómo se organiza su respuesta. En física, las simetrías cumplen un papel fundamental: simplifican el análisis y, al mismo tiempo, revelan propiedades profundas de los fenómenos. Aquí, la simetría servirá como recurso didáctico para destacar patrones en la evolución temporal y preparar el terreno para el marco teórico que desarrollaremos más adelante.

En la figura 2.6 presentamos tres experimentos numéricos sobre nuestro sistema de dos masas. Cada fila corresponde a distintas posiciones iniciales $\vec{\psi}(0)$, con velocidad inicial nula ($\dot{\vec{\psi}}(0)$) en todos los casos. En la columna izquierda se muestra la evolución de las posiciones $\psi_1(t)$ y $\psi_2(t)$. En la derecha graficamos el módulo de la *transformada de Fourier*, $|\hat{\psi}_1(\omega)|$ (curva azul) y $|\hat{\psi}_2(\omega)|$ (curva naranja), que indica qué frecuencias están presentes en cada señal y con qué amplitud. Cada pico del espectro corresponde a uno de los «ritmos» que componen el movimiento. No incluimos aquí la información de fase, ya que nuestro interés inmediato es identificar qué frecuencias intervienen y cómo cambia su peso relativo al modificar las condiciones iniciales.

Al observar las gráficas surge una primera evidencia: la evolución no corresponde a una oscilación armónica simple. A diferencia del oscilador armónico simple, la señal temporal es una superposición de varios ritmos, lo que impide identificar a simple vista una única frecuencia dominante.

Sin embargo, el análisis espectral revela un patrón notable. En todas las filas aparecen siempre las mismas dos frecuencias bien definidas, independientemente de la condición inicial. Lo que sí cambia entre los distintos experimentos es la amplitud relativa con que interviene cada una: en la primera fila ambas contribuyen de manera comparable, en la segunda predomina la primera frecuencia y en la tercera, la segunda.

Además, los espectros de ψ_1 y ψ_2 se superponen completamente en cada caso, lo que indica que ambas coordenadas comparten exactamente el mismo contenido frecuencial y con el mismo peso relativo. La coincidencia es tan precisa que en la columna derecha solo se distingue una curva visible, ya que la otra queda completamente superpuesta.

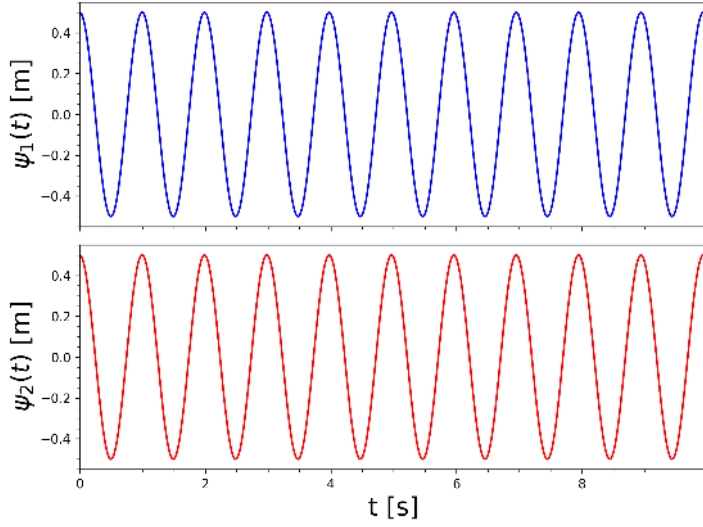
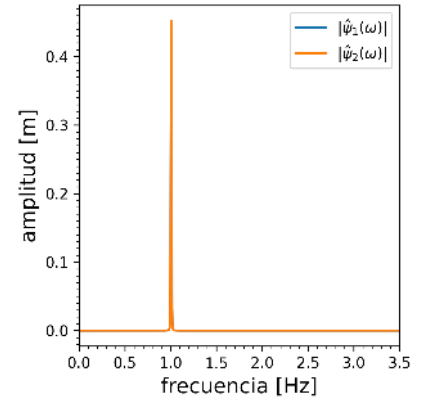
En la primera fila se eligió un desplazamiento inicial de la masa 1, lo que da lugar a una participación comparable de ambas frecuencias presentes en el espectro. En la segunda fila se preparó una condición inicial en la que predomina la primera frecuencia, mientras que en la tercera predomina la segunda. Los valores explícitos de las condiciones iniciales se indican en la leyenda de la figura.

Surge entonces una pregunta natural: ¿es posible elegir condiciones iniciales tales que el sistema oscile en una única frecuencia, aislando uno de estos ritmos fundamentales?

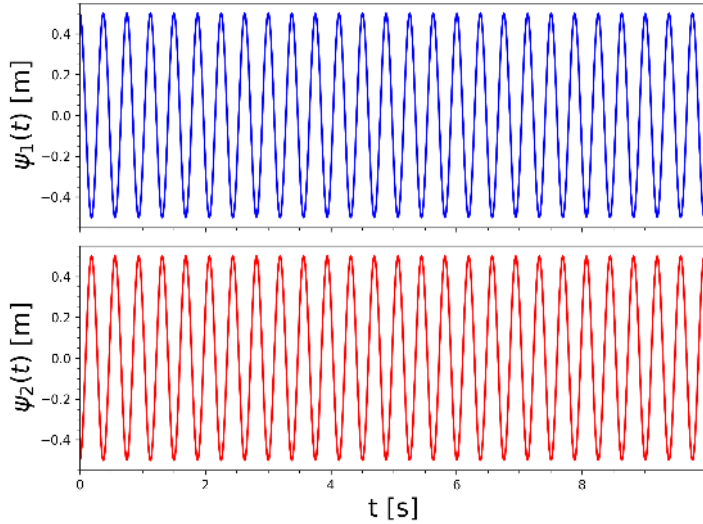
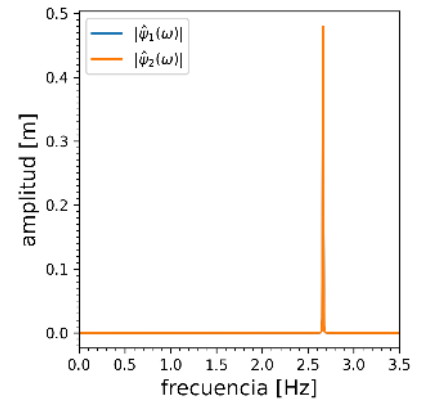
La clave para responderla parece residir en la simetría del propio sistema. Dado que ambas masas y los acoplamientos son idénticos, las ecuaciones de movimiento permanecen invariantes bajo el intercambio de los índices $1 \leftrightarrow 2$, es decir, al permutar

las dos partes móviles. En consecuencia, resulta razonable buscar movimientos que reflejen directamente esa simetría.

Existen, en particular, dos configuraciones iniciales que permanecen invariantes bajo este intercambio: una en la que ambas masas se desplazan de manera idéntica, y otra en la que lo hacen con igual amplitud y en sentidos opuestos. Si el sistema se prepara inicialmente respetando alguna de estas configuraciones —por ejemplo, imponiendo $\psi_1(0) = \psi_2(0)$ junto con $\dot{\psi}_1(0) = \dot{\psi}_2(0)$, o bien $\psi_1(0) = -\psi_2(0)$ junto con $\dot{\psi}_1(0) = -\dot{\psi}_2(0)$ — es natural esperar que la evolución preserve esa relación. Veamos qué revela el experimento numérico cuando se adoptan estas condiciones.

(a) $\psi_1(0) = 0.5m$, $\psi_2(0) = 0.5m$ 

(b)

(c) $\psi_1(0) = 0.5m$, $\psi_2(0) = -0.5m$ 

(d)

Figura 2.7: Resultados del segundo experimento numérico, con condiciones iniciales sugeridas por la simetría del sistema (ver detalles en el texto).

En la figura 2.7 se muestran los resultados correspondientes. A diferencia de los casos anteriores, ahora el espectro presenta un único pico bien definido en cada fila, lo que indica que el movimiento ocurre en una sola frecuencia. El comportamiento temporal, que antes parecía una superposición compleja, se transforma aquí en una oscilación armónica simple. El sistema no solo posee dos frecuencias bien definidas, sino que admite movimientos en los que cada una puede manifestarse de manera aislada. ¿De dónde surgen estas frecuencias y por qué el sistema parece “seleccionarlas” de manera tan precisa?

Capítulo 3

Estructura detrás de los ritmos I: simetrías

Los experimentos numéricos del capítulo anterior revelaron una regularidad notable: independientemente de las condiciones iniciales, en la dinámica siempre aparece un conjunto discreto de frecuencias bien definidas. En el sistema estudiado este conjunto estaba formado por dos valores y vimos además que cada uno de ellos podía excitarse de manera aislada eligiendo convenientemente las condiciones iniciales.

Estas observaciones sugieren que la complejidad aparente del movimiento encubre una estructura interna más simple. ¿Se trata de una peculiaridad del ejemplo analizado, o de una propiedad más general de los sistemas mecánicos lineales? ¿Qué ocurre cuando el número de grados de libertad aumenta? Encontrar respuestas para estas preguntas es uno de los objetivos de este capítulo.

Para responderlas conviene comenzar por reexaminar el sistema de dos grados de libertad ya estudiado, pero ahora desde una perspectiva diferente.

La simulación nos mostró ciertos comportamientos que emergen al perturbar un sistema, como si hubiéramos estado experimentando en el laboratorio para identificar fenómenos característicos. Ese paso resulta valioso para reconocer patrones observables, pero no alcanza para predecir qué ocurriría al sustituir una parte móvil por otra, o un resorte por otro. En otras palabras, no nos permite desentrañar la lógica interna de su dinámica, cuya organización solo puede comprenderse examinando la estructura de las ecuaciones que lo gobiernan.

A continuación veremos que es posible reescribir las ecuaciones de movimiento en términos de nuevas variables en las que la dinámica se simplifica notablemente. Este procedimiento permitirá identificar con claridad los patrones fundamentales que gobiernan la evolución y servirá como punto de partida para una formulación más general.

3.1. La simetría que ya estaba allí

El papel de la simetría en los ejemplos numéricos del capítulo anterior no puede pasar desapercibido: ciertas condiciones iniciales, simétricas o antisimétricas respecto del centro del sistema (ver figura 2.7), daban lugar a movimientos particularmente simples, asociados a una única frecuencia. Esta observación sugería que la clave residía en la propia estructura del sistema: ambas masas iguales y los resortes extremos con la misma constante elástica.

Para este sistema simétrico, las ecuaciones de movimiento (ver (2.6)) pueden escribirse como

$$\begin{aligned} m \ddot{\psi}_1 &= -(k + k') \psi_1 + k' \psi_2, \\ m \ddot{\psi}_2 &= -(k + k') \psi_2 + k' \psi_1. \end{aligned} \tag{3.1}$$

Estas ecuaciones son invariantes bajo el intercambio $\psi_1 \leftrightarrow \psi_2$ y reflejan la simetría izquierda–derecha del sistema físico. La estructura en espejo de ambas expresiones no es accidental: expresa matemáticamente la equivalencia dinámica de las dos masas.

Es natural preguntarse si esta simetría puede aprovecharse para encontrar una descripción del movimiento en la que el acoplamiento entre las coordenadas desaparezca o, al menos, adopte una forma más simple.

3.1.1. Cambiar de coordenadas

La forma en espejo de las ecuaciones (3.1) sugiere que las coordenadas originales ψ_1 y ψ_2 quizá no sean las más adecuadas para describir la dinámica. En problemas algebraicos elementales, cuando un sistema de ecuaciones lineales aparece acoplado, es habitual combinar las ecuaciones —sumarlas o restarlas— para obtener nuevas expresiones más simples. Algo similar puede intentarse aquí: buscar combinaciones de las variables que reflejen la simetría del sistema.

Con esta idea en mente, introduzcamos las nuevas variables

$$\begin{aligned} \Phi_1 &= \psi_1 + \psi_2, \\ \Phi_2 &= \psi_1 - \psi_2. \end{aligned} \tag{3.2}$$

La primera combina los desplazamientos de manera simétrica; la segunda, de manera antisimétrica.

Sumando miembro a miembro las ecuaciones (3.1) se obtiene

$$m (\ddot{\psi}_1 + \ddot{\psi}_2) = -(k + k') (\psi_1 + \psi_2) + k' (\psi_2 + \psi_1),$$

es decir,

$$m \ddot{\Phi}_1 = -k \Phi_1. \tag{3.3}$$

Restándolas, en cambio,

$$m(\ddot{\psi}_1 - \ddot{\psi}_2) = -(k + k')(\psi_1 - \psi_2) - k'(\psi_2 - \psi_1),$$

lo que conduce a

$$m\ddot{\Phi}_2 = -(k + 2k')\Phi_2. \quad (3.4)$$

El resultado es notable: en términos de las nuevas variables, las ecuaciones quedan desacopladas. Cada combinación evoluciona de manera independiente, como si se tratara de un sistema de un solo grado de libertad. La simetría del sistema se traduce así en una descomposición natural del movimiento en dos comportamientos elementales.

3.1.2. Ritmos propios y coordenadas normales

Las nuevas coordenadas $\Phi_1(t)$ y $\Phi_2(t)$ no sólo satisfacen ecuaciones diferenciales desacopladas, sino que cada una obedece una ecuación de la misma forma que la del oscilador armónico simple:

$$\ddot{\Phi}_1 = -\omega_1^2 \Phi_1, \quad \ddot{\Phi}_2 = -\omega_2^2 \Phi_2, \quad (3.5)$$

con

$$\omega_1^2 = \frac{k}{m}, \quad \omega_2^2 = \frac{k + 2k'}{m}. \quad (3.6)$$

Cada variable evoluciona entonces de manera independiente, con su propia frecuencia característica. Usando lo ya conocido para el oscilador armónico, podemos afirmar que

$$\Phi_1(t) = A_1 \cos(\omega_1 t + \varphi_1), \quad \Phi_2(t) = A_2 \cos(\omega_2 t + \varphi_2), \quad (3.7)$$

donde las constantes A_1 , A_2 , φ_1 y φ_2 se determinan a partir de las posiciones y velocidades iniciales.

Las nuevas variables Φ_1 y Φ_2 que desacoplan las ecuaciones de movimiento se llaman *coordenadas normales* y las constantes ω_1 y ω_2 que aparecen en (3.5) se llaman *frecuencias normales* del sistema.

Hasta aquí, las coordenadas normales han simplificado el problema, reduciendo el sistema a dos osciladores armónicos independientes. Pero surge una duda natural: ¿poseen estos osciladores alguna entidad física, o son un mero artificio matemático? ¿Estamos ante objetos observables o ante una abstracción que solo existe en nuestro papel? Para descifrar su significado y entender cómo estos osciladores dictan la dinámica real, debemos regresar a nuestras coordenadas originales.

3.1.3. Selección de frecuencias mediante las condiciones iniciales

Las coordenadas originales se obtienen al invertir el sistema (3.2)

$$\psi_1 = \frac{1}{2}(\Phi_1 + \Phi_2), \quad \psi_2 = \frac{1}{2}(\Phi_1 - \Phi_2).$$

Usando las expresiones (3.7) para las coordenadas desacopladas, el movimiento general de cada parte móvil resulta

$$\begin{aligned}\psi_1(t) &= \frac{1}{2} \left[A_1 \cos(\omega_1 t + \varphi_1) + A_2 \cos(\omega_2 t + \varphi_2) \right], \\ \psi_2(t) &= \frac{1}{2} \left[A_1 \cos(\omega_1 t + \varphi_1) - A_2 \cos(\omega_2 t + \varphi_2) \right].\end{aligned}\tag{3.8}$$

Las expresiones anteriores muestran que, en general, el movimiento de cada parte móvil es la superposición de dos oscilaciones armónicas con frecuencias distintas ω_1 y ω_2 .

Las amplitudes A_1 y A_2 —determinadas por las condiciones iniciales— fijan cuánto contribuye cada frecuencia al movimiento total. Esto explica el resultado observado en el primer experimento numérico (figura 2.6): cualquiera sea la condición inicial genérica, ambas frecuencias están presentes, aunque con pesos relativos diferentes.

Además, los valores de las frecuencias obtenidas mediante la transformada rápida de Fourier coinciden, dentro de la precisión del cálculo, con los valores ω_1 y ω_2 calculados usando las ecuaciones (3.6), lo que confirma la consistencia del modelo.

La conclusión anterior parece sugerir que ambas frecuencias están siempre presentes. Pero ¿es realmente inevitable esta mezcla? Si alguna de las amplitudes fuese nula, el movimiento contendría una sola frecuencia. ¿Existen condiciones iniciales que produzcan esta situación? ¿Puede ocurrir esto?

Para responder a estas preguntas, calculemos explícitamente las amplitudes A_1 y A_2 para distintas preparaciones del sistema. Si bien el camino natural para encontrar las constantes A_1 , A_2 , φ_1 y φ_2 parece partir de evaluar en $t = 0$ las ecuaciones (3.8), el camino mas conveniente es hacerlo en las ecuaciones (3.7). Como $\Phi_1 = \psi_1 + \psi_2$ y $\Phi_2 = \psi_1 - \psi_2$, en $t = 0$ resulta

$$\begin{aligned}\Phi_1(0) &= \psi_1(0) + \psi_2(0), & \dot{\Phi}_1(0) &= \dot{\psi}_1(0) + \dot{\psi}_2(0), \\ \Phi_2(0) &= \psi_1(0) - \psi_2(0), & \dot{\Phi}_2(0) &= \dot{\psi}_1(0) - \dot{\psi}_2(0).\end{aligned}$$

es decir

$$\begin{aligned}A_1 \cos \varphi_1 &= \psi_1(0) + \psi_2(0), & -\omega_1 A_1 \sin \varphi_1 &= \dot{\psi}_1(0) + \dot{\psi}_2(0), \\ A_2 \cos \varphi_2 &= \psi_1(0) - \psi_2(0), & -\omega_2 A_2 \sin \varphi_2 &= \dot{\psi}_1(0) - \dot{\psi}_2(0).\end{aligned}\tag{3.9}$$

Las dos filas de la ecuación (3.9) no mezclan todas las incógnitas, las dos ecuaciones de la primera fila tienen sólo A_1 y φ_1 y las dos ecuaciones de la segunda fila tienen sólo A_2 y φ_2 . Para eliminar la fase φ_1 de las ecuaciones de la primera fila, elevamos al cuadrado ambas ecuaciones (dividiendo previamente la segunda por ω_1) y las sumamos miembro a miembro. En el miembro izquierdo aparece entonces el factor $\cos^2 \varphi_1 + \sin^2 \varphi_1 = 1$, lo que permite aislar A_1^2 y expresarlo únicamente en función de las condiciones iniciales. Repitiendo el mismo procedimiento para el segundo par de ecuaciones, se obtiene una expresión análoga para la amplitud A_2^2 en términos de las condiciones iniciales. Así

resulta

$$\begin{aligned} A_1^2 &= \left[\psi_1(0) + \psi_2(0) \right]^2 + \frac{1}{\omega_1^2} \left[\dot{\psi}_1(0) + \dot{\psi}_2(0) \right]^2, \\ A_2^2 &= \left[\psi_1(0) - \psi_2(0) \right]^2 + \frac{1}{\omega_2^2} \left[\dot{\psi}_1(0) - \dot{\psi}_2(0) \right]^2. \end{aligned} \quad (3.10)$$

Las fases iniciales también quedan determinadas por las condiciones iniciales. Dividiendo miembro a miembro la segunda ecuación de cada línea por la primera, se elimina la amplitud y se obtienen expresiones para la tangente de la fase

$$\begin{aligned} \tan \varphi_1 &= -\frac{1}{\omega_1} \frac{\dot{\psi}_1(0) + \dot{\psi}_2(0)}{\psi_1(0) + \psi_2(0)}, \\ \tan \varphi_2 &= -\frac{1}{\omega_2} \frac{\dot{\psi}_1(0) - \dot{\psi}_2(0)}{\psi_1(0) - \psi_2(0)}. \end{aligned} \quad (3.11)$$

En particular, notar que si las velocidades iniciales son nulas, resulta $\varphi_1 = \varphi_2 = 0$.

Las ecuaciones (3.11) no definen completamente las fases, porque la función arcotangente convencional, al ser periódica en π , no distingue entre cuadrantes opuestos. Para determinar unívocamente φ_1 y φ_2 en el intervalo $[0, 2\pi)$ es fundamental considerar los signos del seno y del coseno —determinados por las ecuaciones (3.9)— de manera independiente.

En el ámbito computacional, la ambigüedad del arcotangente convencional —que solo considera el cociente y/x y pierde información sobre los signos individuales— se resuelve mediante la función `atan2(y, x)`. Esta función devuelve el ángulo cuya tangente es efectivamente y/x , pero al recibir los argumentos por separado, emplea los signos de x e y para situar el resultado en el cuadrante correcto del plano cartesiano. La función `atan2` es un estándar en la programación científica y está disponible en los principales lenguajes. En el ecosistema de Python, por ejemplo, contamos con `math.atan2` para cálculos escalares y con `numpy.arctan2` para trabajar de forma vectorizada sobre arreglos de datos.

Los resultados de las ecuaciones (3.10) y (3.11) responden, finalmente, al interrogante que motivó este análisis: ¿es inevitable la superposición de frecuencias? Como ya daban indicios las simulaciones numéricas presentadas en la figura 2.7, la respuesta es negativa.

El formalismo desarrollado nos permite comprender analíticamente este fenómeno: existen preparaciones específicas donde una de las amplitudes se anula, confinando el movimiento a la oscilación pura de un único modo normal. Puesto que, de acuerdo con (3.10), cada amplitud es una suma de cuadrados, esta solo puede anularse si ambos términos son cero de manera independiente. Debido a la alta simetría de nuestro sistema, las configuraciones iniciales que dan lugar a este estado son:

- CONFIGURACIÓN 1: $A_2 = 0$, $\Phi_2(t) = 0 \forall t$

Cuando: $\psi_1(0) = \psi_2(0)$ y $\dot{\psi}_1(0) = \dot{\psi}_2(0)$.

En este caso: $\psi_1(t) = \psi_2(t) = A_1 \cos(\omega_1 t + \varphi_1)$.

- CONFIGURACIÓN 2: $A_1 = 0$, $\Phi_1(t) = 0 \forall t$

Cuando: $\psi_1(0) = -\psi_2(0)$ y $\dot{\psi}_1(0) = -\dot{\psi}_2(0)$.

En este caso: $\psi_1(t) = -\psi_2(t) = A_2 \cos(\omega_2 t + \varphi_2)$.

En la primera configuración (simétrica), ambas masas oscilan en fase y con idéntica amplitud. Este movimiento describe la traslación del centro de masa del sistema sin deformación interna. Dado que el resorte central mantiene su longitud constante en todo instante, no ejerce fuerza alguna sobre las masas; es decir, en esta configuración, el resorte central no interviene en la dinámica y su presencia es irrelevante. En consecuencia, el movimiento está determinado exclusivamente por los resortes externos, dando lugar a la frecuencia ω_1 .

En la segunda configuración (antisimétrica), las masas se mueven en sentidos opuestos, lo que maximiza la compresión y estiramiento del resorte central. A diferencia del caso anterior, el resorte central interviene activamente añadiendo una fuerza restitutiva adicional a la de los resortes externos. La mayor rigidez resultante se traduce en la frecuencia de oscilación ω_2 superior a ω_1 . Este modo corresponde a la oscilación interna del sistema, asociada a la coordenada relativa $\Phi_2(t) = \psi_1(t) - \psi_2(t)$.

En ambos casos todas las partes móviles oscilan con una única frecuencia y mantienen relaciones de amplitud constantes. El sistema vibra según un patrón colectivo bien definido.

Estos patrones especiales de oscilación no son una curiosidad accidental del ejemplo, sino que revelan una estructura más profunda del sistema.

3.2. Modos normales

Llamaremos *modo normal* a todo patrón colectivo de oscilación en el cual el sistema vibra armónicamente con una única frecuencia, con una estructura espacial que permanece fija en el tiempo. Esta definición implica que las relaciones de amplitud entre las distintas partes no cambian y que no aparecen desfases relativos nuevos durante la evolución: si dos puntos del sistema comienzan moviéndose en fase (o en oposición de fase), lo harán así en todo instante. La forma del movimiento es invariable; sólo cambia el factor temporal común que la multiplica.

Más allá del formalismo matemático, un modo normal puede reconocerse experimentalmente sin necesidad de introducir cambios de variables ni procedimientos de desacople. Un sistema se encuentra en un modo normal cuando adopta una configuración geométrica que oscila «en bloque»: todas sus partes móviles pasan simultáneamente por su posición de equilibrio y conservan entre sí proporciones fijas de amplitud durante toda la evolución. Lo que cambia es únicamente un factor común multiplicativo –los cosenos en nuestro ejemplo– mientras que la forma espacial del movimiento permanece inalterada.

Desde este punto de vista resulta conveniente adoptar una convención ligeramente distinta de la habitual al definir la amplitud de un movimiento armónico simple. En lugar de describir, por ejemplo, el modo antisimétrico diciendo que las amplitudes son positivas pero las fases difieren en π , consideraremos que todas las partes oscilan con una misma fase temporal, permitiendo que las amplitudes puedan tomar signo, para representar así *la orientación relativa* del desplazamiento dentro del patrón geométrico colectivo. Esta forma de describir los modos enfatiza que lo esencial no es la fase individual de cada coordenada, sino la estructura fija de relaciones espaciales que caracteriza a cada movimiento normal. Más adelante veremos que esta convención resulta especialmente natural al formular el problema en términos más generales.

La cantidad de modos normales coincide con el número de grados de libertad del sistema. En nuestro ejemplo con dos grados de libertad, aparecen dos modos. Veremos más adelante que un sistema con N grados de libertad posee, en general, N modos normales.

Más aún, las cantidades ω_1^2 y ω_2^2 pueden interpretarse como la fuerza restitutiva efectiva por unidad de masa y por unidad de desplazamiento cuando el sistema oscila según cada uno de estos patrones colectivos. Esta interpretación no depende de los detalles particulares del ejemplo, sino que constituye una propiedad general de los modos normales en sistemas físicos lineales.

Vale la pena detenerse un momento y comparar estos movimientos con la idea de «perturbación que se propaga» que motivó nuestra discusión sobre sistemas apartados del equilibrio. En los modos normales que hemos identificado no aparece nada semejante: ninguna parte del sistema pone en movimiento a otra de manera sucesiva ni se observa un desplazamiento de la perturbación a lo largo del sistema. Desde el primer instante queda establecida una configuración espacial completa que simplemente oscila como un todo coherente. La estructura del movimiento está fija desde el inicio; sólo varía su amplitud común en el tiempo.

Si bien en esta etapa no aparece aún la perturbación que se propaga que motivó nuestra búsqueda, no debemos desalentarnos: veremos más adelante que los fenómenos de transferencia surgen precisamente cuando se superponen varios modos normales. La propagación no está contenida en cada modo por separado, sino que emerge de su interferencia.

Como lo muestran las ecuaciones (3.8), el movimiento más general del sistema puede interpretarse como una superposición de estos modos fundamentales. Las condiciones iniciales no generan nuevas frecuencias: sólo determinan cuánto participa cada modo en el movimiento total.

Esto permite reinterpretar lo observado en los experimentos numéricos del Capítulo 2. Cuando la preparación fue perfectamente simétrica (figura 2.7a) apareció una única frecuencia —la menor—; cuando fue antisimétrica (figura 2.7b), emergió exclusivamente la frecuencia mayor. En cualquier otra preparación ambas frecuencias estuvieron presentes (figura 2.6). El sistema no inventa ritmos nuevos según cómo se lo excite: siempre vibra con sus ritmos propios. Las condiciones iniciales sólo seleccionan la combinación en la que esos ritmos se manifiestan.

Esta comprensión física de fenómenos que podrían observarse tanto en una simulación como en un experimento sencillo fue posible gracias a un cambio de perspectiva cuya naturaleza merece ser examinada con más cuidado.

3.3. Lo que realmente hicimos

¿En qué consistió exactamente ese cambio de perspectiva?

No adivinamos las soluciones ni apelamos a artificios especiales. Tampoco resolvimos directamente el sistema acoplado en las coordenadas originales. Lo que hicimos fue algo más simple: buscamos combinaciones lineales de las coordenadas que reflejaran la simetría del sistema.

Al introducir las coordenadas $\Phi_1 = \psi_1 + \psi_2$ y $\Phi_2 = \psi_1 - \psi_2$, el acoplamiento desaparece y la dinámica se separa en dos ecuaciones independientes. Aunque este paso nos obliga a renunciar a la descripción de cada masa por separado, la renuncia es coherente con el punto de vista anticipado en la sección 2.4: el comportamiento del sistema es esencialmente colectivo y, por tanto, la dinámica no pertenece a las partes individuales, sino al conjunto. Desde esta perspectiva, el desacople no es un accidente algebraico, sino la consecuencia de elegir coordenadas más adecuadas para describir el problema.

Este punto de vista será crucial a continuación. Reformularemos el mismo problema en notación matricial y veremos que el procedimiento seguido no es exclusivo de este ejemplo, sino la manifestación de una estructura más general relacionada con encontrar los autovalores y autovectores de una matriz. Coloquialmente se podría decir que se trata de descubrir los movimientos que «le gustan» al sistema, es decir, aquellos que surgen naturalmente de su estructura. Más adelante extenderemos estas ideas a sistemas con un número arbitrario de grados de libertad, donde la misma estrategia permitirá identificar todos los modos normales de manera sistemática.

3.4. El punto de vista matricial

Para el ejemplo considerado —con ecuaciones de movimiento dadas por (3.1)— las matrices de masas y de coeficientes restitutivos son

$$\mathbb{M} = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}, \quad \mathbb{K} = \begin{bmatrix} k + k' & -k' \\ -k' & k + k' \end{bmatrix}. \quad (3.12)$$

El sistema diferencial acoplado se escribe

$$\mathbb{M} \ddot{\vec{\psi}}(t) = -\mathbb{K} \vec{\psi}(t). \quad (3.13)$$

Multiplicando por $\mathbb{M}^{-1}$ se obtiene

$$\ddot{\vec{\psi}}(t) = -\tilde{\mathbb{K}} \vec{\psi}(t), \quad (3.14)$$

donde $\tilde{\mathbb{K}} \equiv \mathbb{M}^{-1}\mathbb{K}$ es la matriz dinámica. Esta matriz condensa toda la información dinámica del sistema, ya que determina las aceleraciones a partir de las posiciones.

En la base de coordenadas ψ_1 y ψ_2 $\tilde{\mathbb{K}}$ no es diagonal. Pero la simetría nos llevó a introducir el cambio de coordenadas $\Phi_1 = \psi_1 + \psi_2$, $\Phi_2 = \psi_1 - \psi_2$, mediante el cual las ecuaciones diferenciales quedan desacopladas.

En el lenguaje del álgebra lineal, este cambio de variables puede interpretarse como un cambio de base en el espacio vectorial $\mathbb{R}^2$ («el espacio de configuraciones»). Las nuevas variables Φ_1 y Φ_2 corresponden a describir el mismo vector $\vec{\psi}$, pero en una base distinta, adaptada a los patrones colectivos de movimiento del sistema. Esta base se denomina *base modal*.

Es importante notar que este cambio de base no es único. En particular, podríamos haber definido las nuevas coordenadas como

$$\Phi_1 = \frac{1}{\sqrt{2}}(\psi_1 + \psi_2), \quad \Phi_2 = \frac{1}{\sqrt{2}}(\psi_2 - \psi_1),$$

sin modificar el desacople de las ecuaciones. En este caso, el cambio de base se escribe como

$$\begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}.$$

Las nuevas coordenadas están ahora normalizadas, y la matriz de cambio de base es ortogonal con determinante +1, por lo que corresponde a una rotación de $\pi/4$ en el espacio de configuraciones.

La elección original, en cambio, conduce a una matriz con determinante negativo, que puede interpretarse como una reflexión adicional. Además, si las coordenadas originales estaban asociadas a versores unitarios, este cambio de base da lugar a nuevos vectores que ya no están normalizados.

Sin embargo, estas diferencias no tienen consecuencias dinámicas: el desacople depende únicamente de las direcciones de los modos normales, no de su norma ni de su orientación.

Y acá surge la pregunta natural: ¿funciona siempre este cambio de base?

En nuestro ejemplo, la respuesta es afirmativa. La matriz dinámica es simétrica, y los teoremas del álgebra lineal aseguran que toda matriz real y simétrica puede diagonalizarse mediante un cambio de base adecuado. Las nuevas coordenadas que introdujimos coinciden precisamente con las coordenadas asociadas a esa base privilegiada.

Esta propiedad no es accidental. En muchos sistemas mecánicos $\tilde{\mathbb{K}}$ es simétrica, lo que garantiza la existencia de una base de modos normales.

Si las masas fueran distintas, $\tilde{\mathbb{K}}$ dejaría en general de ser simétrica. Sin embargo, el sistema seguiría teniendo modos normales. Como veremos más adelante al analizar el problema en su forma general, la estructura particular de las matrices mecánicas garantiza que siempre es posible encontrar coordenadas que desacoplen el movimiento.

Las observaciones hechas en este apartado del capítulo anterior sobre la estructura

de las matrices $\mathbf{M}$ y $\mathbf{K}$ toman ahora relevancia: son precisamente esas propiedades las que garantizan la existencia de modos normales y permiten desacoplar el movimiento mediante un cambio de base adecuado.

3.5. El método sistemático cuando $\tilde{\mathbb{K}}$ es simétrica

¿Y si no estamos lo suficientemente inspirados como para encontrar directamente las combinaciones de coordenadas que desacoplan las ecuaciones de movimiento? Según lo discutido en las secciones 3.3 y 3.4, el camino a seguir está bien definido: desacoplar las ecuaciones equivale a encontrar un cambio de base que diagonalice la matriz dinámica. El desafío, por lo tanto, se traslada al terreno del álgebra lineal.

A continuación, construiremos sistemáticamente dicho cambio de base. Consideremos un sistema de N grados de libertad y recordemos que las ecuaciones de movimiento se escriben

$$\ddot{\vec{\psi}}(t) = -\tilde{\mathbb{K}}\vec{\psi}(t). \quad (3.15)$$

Para avanzar de manera ordenada, comenzaremos por el caso más simple: supondremos que la matriz dinámica es simétrica. Como veremos, las ideas fundamentales se mantienen en el caso general, aunque la formulación requiere algunos ajustes que introduciremos más adelante.

Si $\tilde{\mathbb{K}}$ es simétrica, como en el sistema (3.12), la diagonalización está garantizada. Los teoremas del álgebra lineal aseguran que la matriz dinámica posee un conjunto completo de autovectores linealmente independientes. Si estos autovectores se disponen como columnas de una matriz $\mathbb{D}$, entonces el producto

$$\mathbb{D}^{-1}\tilde{\mathbb{K}}\mathbb{D},$$

(transformación de semejanza) diagonaliza a $\tilde{\mathbb{K}}$.

¿Cómo manipular la ecuación (3.15) para que en vez de $\tilde{\mathbb{K}}$ aparezca $\mathbb{D}^{-1}\tilde{\mathbb{K}}\mathbb{D}$? La idea es sencilla: multiplicamos primero a izquierda ambos lados de la ecuación por $\mathbb{D}^{-1}$. Luego reemplazamos la matriz identidad —implícita entre los factores $\tilde{\mathbb{K}}$ y $\vec{\psi}(t)$ — por $\mathbb{D}\mathbb{D}^{-1}$. Así obtenemos

$$\mathbb{D}^{-1}\ddot{\vec{\psi}}(t) = -\mathbb{D}^{-1}\tilde{\mathbb{K}}\mathbb{D}\mathbb{D}^{-1}\vec{\psi}(t) \quad (3.16)$$

Ahora definimos el vector

$$\vec{\Phi}(t) = \mathbb{D}^{-1}\vec{\psi}(t), \quad (3.17)$$

y como $\mathbb{D}$ no depende del tiempo, las derivadas temporales de $\vec{\psi}(t)$ se convierten en las derivadas temporales de $\vec{\Phi}(t)$

$$\ddot{\vec{\Phi}} = \mathbb{D}^{-1}\ddot{\vec{\psi}}. \quad (3.18)$$

Este es el momento de observar que la matriz $\mathbb{D}^{-1}\tilde{\mathbb{K}}\mathbb{D}$, por construcción, es diagonal. Si escribimos sus elementos diagonales como ω_i^2 , $i = 1, \dots, N$, el sistema finalmente queda

$$\ddot{\Phi}_i + \omega_i^2 \Phi_i = 0, \quad i = 1, \dots, N. \quad (3.19)$$

Y así, bajo la hipótesis de que $\tilde{\mathbb{K}}$ es simétrica —una condición particular que simplifica notablemente el problema, pero que aparece en muchos sistemas físicos de interés— las ecuaciones de movimiento han quedado desacopladas: cada coordenada modal satisface la ecuación de un oscilador armónico independiente, cuyas soluciones son

$$\Phi_i(t) = A_i \cos(\omega_i t + \varphi_i). \quad (3.20)$$

Los cuadrados de las frecuencias propias ω_i son los autovalores de la matriz dinámica, mientras que las constantes A_i y φ_i quedan determinadas por las condiciones iniciales.

Para interpretar estas soluciones en términos de las coordenadas originales invertimos la definición (3.17)

$$\vec{\psi} = \mathbb{D} \vec{\Phi},$$

de modo que

$$\vec{\psi}(t) = \sum_i \vec{v}_i \Phi_i(t), \quad (3.21)$$

donde $\vec{v}_i$ denota el i -ésimo autovector de la matriz dinámica.

Cada término de esta suma describe un movimiento en el cual todas las coordenadas del sistema oscilan con la misma frecuencia ω_i y con relaciones de amplitud fijas determinadas por el autovector correspondiente. Estos movimientos particulares son los *modos normales* del sistema.

Para muchos lectores, los autovalores y autovectores aparecen por primera vez en cursos de álgebra lineal como conceptos bastante abstractos. Momentos como este, en los que esas mismas ideas revelan la estructura del movimiento de un sistema físico, suelen producir una sorpresa agradable. Este tipo de conexiones entre matemática y física es uno de los motivos por los cuales muchos científicos se entusiasman con ambas disciplinas.

3.5.1. Nota técnica: amplitudes, fases y normalización

Al introducir las funciones armónicas (3.20) en la superposición de modos normales (3.21) obtenemos

$$\vec{\psi}(t) = \sum_i \vec{v}_i A_i \cos(\omega_i t + \varphi_i). \quad (3.22)$$

Cada término de la suma describe un patrón de movimiento fijo —determinado por el autovector $\vec{v}_i$ — que oscila en el tiempo con frecuencia ω_i .

Las constantes A_i y φ_i dependen de las $2N$ condiciones iniciales del sistema. En este contexto conviene notar que los autovectores no están definidos de manera única: si $\vec{v}_i$ es un autovector, también lo es cualquier múltiplo $\alpha \vec{v}_i$, con α una constante distinta de cero. ¿Significa esto que aparecen $3N$ constantes a determinar con solo $2N$ condiciones iniciales? La respuesta es que no hay tal problema: las constantes de escala de los autovectores se absorben en las amplitudes A_i . En consecuencia, el número de parámetros independientes sigue siendo $2N$, exactamente los que hacen falta para fijar la solución a partir de las condiciones iniciales.

Por esta razón, es habitual adoptar alguna convención para fijar la escala de los autovectores. Una elección frecuente consiste en normalizarlos, es decir, elegirlos de manera que tengan norma unitaria.

Finalmente, cuando se enumeran los modos normales suele adoptarse la convención de ordenarlos según sus frecuencias propias, de menor a mayor. De esta manera, el modo $i = 1$ corresponde a la frecuencia más baja del sistema, mientras que los índices mayores describen oscilaciones progresivamente más rápidas.

3.5.2. Ejemplo: procedimiento para el caso desacoplado «a mano»

Apliquemos el procedimiento sistemático basado en la diagonalización de $\tilde{\mathbb{K}}$ al caso de estudio con ecuaciones de movimiento dadas por (3.1) y matrices dadas por (3.12).

Para este sistema, la matriz dinámica tiene la forma

$$\tilde{\mathbb{K}} = \begin{bmatrix} a & b \\ b & a \end{bmatrix}, \quad (3.23)$$

donde $a = (k + k')/m$ y $b = -k'/m$ dependen de la masa y de las constantes elásticas de los resortes.

Los autovalores de esta matriz se obtienen resolviendo la ecuación característica $\det(\tilde{\mathbb{K}} - \lambda \mathbb{I}) = 0$. En este caso se obtiene

$$\lambda_1 = a + b, \quad \lambda_2 = a - b. \quad (3.24)$$

Como vimos anteriormente, estos autovalores corresponden a los cuadrados de las frecuencias propias del sistema,

$$\omega_1^2 = \lambda_1 = a + b, \quad \omega_2^2 = \lambda_2 = a - b.$$

Ahora podemos calcular explícitamente los autovectores, aunque antes es interesante observar que la propia estructura de la matriz (3.23) sugiere dos direcciones especiales. En efecto,

$$\tilde{\mathbb{K}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = (a + b) \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \tilde{\mathbb{K}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = (a - b) \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Esto muestra que los vectores

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

son autovectores de $\tilde{\mathbb{K}}$, tal como se obtiene al reemplazar cada autovalor en la ecuación $(\tilde{\mathbb{K}} - \lambda \mathbb{I})\vec{v} = 0$ y resolver para $\vec{v}$. Es claro que estos autovectores describen los modos normales del sistema: $\vec{v}_1$ representa un movimiento en el que las dos coordenadas oscilan en fase, mientras que $\vec{v}_2$ representa un movimiento en oposición de fase.

Si reunimos estos autovectores como columnas de la matriz $\mathbb{D}$,

$$\mathbb{D} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

el cambio de variables

$$\vec{\Phi} = \mathbb{D}^{-1} \vec{\psi}$$

produce precisamente las combinaciones

$$\Phi_1 = \psi_1 + \psi_2, \quad \Phi_2 = \psi_1 - \psi_2.$$

De esta manera recuperamos exactamente las mismas coordenadas modales que habíamos encontrado anteriormente por simple inspección del problema. La diferencia es que ahora aparecen como consecuencia directa del procedimiento general basado en autovalores y autovectores.

Según (3.22), en este sistema, la evolución más general de la perturbación del estado de equilibrio se puede escribir de la siguiente manera

$$\begin{bmatrix} \psi_1(t) \\ \psi_2(t) \end{bmatrix} = A_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cos(\omega_1 t + \varphi_1) + A_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \cos(\omega_2 t + \varphi_2), \quad (3.25)$$

donde las amplitudes y fases se obtienen siguiendo el procedimiento explicado en la sección 3.1.3.

3.5.3. Volvamos a mirar los resultados obtenidos en el laboratorio

Los resultados obtenidos en el apartado anterior ilustran cómo el álgebra lineal permite encontrar, *de forma sistemática*, el desacople de las ecuaciones dinámicas y la forma de los modos normales, que para este caso particular habíamos podido anticipar gracias a la simetría del sistema. Veamos ahora cómo esta descomposición modal explica también los resultados del análisis espectral mostrado en la figura 2.6 para tres desplazamientos iniciales distintos (todos con velocidades iniciales nulas).

La primera fila de la figura 2.6 corresponde a la condición inicial

$$\begin{bmatrix} \psi_1(0) \\ \psi_2(0) \end{bmatrix} = \begin{bmatrix} 0.7 \text{ m} \\ 0 \end{bmatrix} = 0.35 \text{ m} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.35 \text{ m} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Es decir, el desplazamiento inicial puede escribirse como una combinación de los dos modos normales con coeficientes iguales. Estos coeficientes —las coordenadas del vector inicial en la base modal— determinan las amplitudes de los modos excitados. Las amplitudes permanecen constantes durante la evolución, mientras que cada coordenada modal oscila con su frecuencia propia, por eso las curvas de $\vec{\psi}(t)$ mostradas en la figura 2.6a (obtenidas integrando numéricamente el sistema de ecuaciones diferenciales) son las mismas curvas que se obtendrían usando la ecuación (3.25) con $A_1 = A_2 = 0.35 \text{ m}$,

$\phi_1 = \phi_2 = 0$ y ω_1 y ω_2 dados por la ecuación (3.6). Es en estos valores de ω_1 y ω_2 donde la transformada de Fourier obtenida numéricamente tiene sus máximos, que se manifiestan como picos de igual altura e indica que ambos modos contribuyen con la misma amplitud.

La segunda fila de la figura 2.6 corresponde a la condición inicial

$$\begin{bmatrix} \psi_1(0) \\ \psi_2(0) \end{bmatrix} = \begin{bmatrix} 0.67 \text{ m} \\ 0.23 \text{ m} \end{bmatrix} = 0.45 \text{ m} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.22 \text{ m} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

En este caso el desplazamiento inicial tiene una componente mayor en la dirección del modo simétrico. Por lo tanto, en la evolución temporal del sistema este modo contribuye con una amplitud mayor que el modo antisimétrico. En consecuencia, el análisis espectral de la señal muestra nuevamente dos picos en las frecuencias propias ω_1 y ω_2 , pero ahora con alturas distintas, reflejando la diferencia entre las amplitudes modales presentes en la condición inicial.

Finalmente, la tercera fila de la figura 2.6 corresponde a la condición inicial

$$\begin{bmatrix} \psi_1(0) \\ \psi_2(0) \end{bmatrix} = \begin{bmatrix} 0.63 \text{ m} \\ -0.32 \text{ m} \end{bmatrix} = 0.155 \text{ m} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.475 \text{ m} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

En este caso la condición inicial está dominada por la componente asociada al modo antisimétrico. En consecuencia, este modo contribuye con mayor amplitud a la evolución temporal del sistema. El análisis espectral muestra nuevamente dos picos en las frecuencias propias del sistema, pero ahora el pico asociado al modo antisimétrico es el de mayor altura, indicando que este modo es el que domina la dinámica resultante.

Capítulo 4

Estructura detrás de los ritmos II: simetrías y asimetrías

Retomamos ahora el caso general, en el que la matriz dinámica no es necesariamente simétrica. Veremos que las ideas introducidas para el caso simétrico en el capítulo anterior siguen siendo válidas, aunque su implementación requiere algunos ajustes.

En lo que sigue, desarrollaremos en detalle el procedimiento de desacople modal en este contexto más general, prestando especial atención al papel de la ortogonalidad y su generalización. Analizaremos también cómo se incorporan las condiciones iniciales y cómo se construye el desarrollo modal en ausencia de ortogonalidad en el sentido usual.

Para ilustrar estos aspectos, estudiaremos un ejemplo integrador que permitirá discutir de manera concreta la estructura de los modos normales, la forma de los autovalores y autovectores, y la aparición de distintos regímenes de acoplamiento. Este análisis se complementará con una interpretación geométrica del problema, así como con la descripción de fenómenos dinámicos característicos, como la modulación de amplitudes y los batidos.

4.1. El método sistemático: caso general

La diagonalización de $\tilde{\mathbf{K}}$ permite identificar los modos del sistema de manera sistemática, aún en situaciones donde no es simétrica.

Recordemos que el producto de dos matrices simétricas no necesariamente conserva la simetría. Sin embargo, si una de ellas es *definida positiva* (como la matriz de masa $\mathbf{M}$, que es diagonal con entradas $m_i > 0$), el producto resulta semejante a una matriz simétrica, y por lo tanto es diagonalizable. Para verlo, reescribimos la matriz dinámica $\mathbf{M}^{-1}\mathbf{K}$ introduciendo la matriz $\mathbf{M}^{-1/2}$, cuya existencia está asegurada porque $\mathbf{M}$ es definida positiva (toda matriz definida positiva admite raíz cuadrada simétrica):

$$\tilde{\mathbf{K}} = \mathbf{M}^{-1}\mathbf{K} = \mathbf{M}^{-1/2} \underbrace{(\mathbf{M}^{-1/2}\mathbf{K}\mathbf{M}^{-1/2})}_{\hat{\mathbf{K}}} \mathbf{M}^{1/2} \quad (4.1)$$

La matriz entre paréntesis $\hat{\mathbf{K}}$ es simétrica por construcción. En consecuencia, la matriz dinámica $\tilde{\mathbf{K}}$ es semejante a la matriz simétrica $\hat{\mathbf{K}}$, por lo cual hereda su diagonalizabilidad y posee un conjunto completo de autovectores linealmente independientes. Y esto garantiza que el sistema pueda descomponerse en modos bien definidos.

Como ambas matrices son semejantes, $\tilde{\mathbf{K}}$ y $\hat{\mathbf{K}}$ tienen exactamente los mismos autovalores $\lambda_n = \omega_n^2$. Sin embargo, los autovectores son distintos.

Si $\vec{u}_n$ es autovector de $\hat{\mathbf{K}}$, entonces

$$\hat{\mathbf{K}} \vec{u}_n = \lambda_n \vec{u}_n$$

o sea

$$(\mathbf{M}^{-1/2}\mathbf{K}\mathbf{M}^{-1/2}) \vec{u}_n = \lambda_n \vec{u}_n.$$

Multiplicando por $\mathbf{M}^{-1/2}$

$$\mathbf{M}^{-1}\mathbf{K} (\mathbf{M}^{-1/2} \vec{u}_n) = \lambda_n (\mathbf{M}^{-1/2} \vec{u}_n)$$

es decir que

$$\vec{v}_n = \mathbf{M}^{-1/2} \vec{u}_n \quad (4.2)$$

es autovector de la matriz dinámica $\tilde{\mathbf{K}}$.

Notemos que los autovectores $\vec{u}_n$ de la matriz simétrica $\hat{\mathbf{K}}$ son ortogonales y se pueden normalizar con el producto escalar canónico

$$\vec{u}_n^T \vec{u}_m = \delta_{nm}, \quad (4.3)$$

donde el supraíndice T indica «transposición» y δ_{nm} es la delta de Kronecker, definida por

$$\delta_{nm} = \begin{cases} 1 & \text{si } n = m, \\ 0 & \text{si } n \neq m. \end{cases}$$

Sin embargo, en el espacio físico de las coordenadas $\vec{\psi}$, los autovectores físicos $\vec{v}_n = \mathbf{M}^{-1/2} \vec{u}_n$ (que determinan cómo se combinan los desplazamientos de los distintos grados de libertad en cada modo) no son, en general, ortogonales entre sí con respecto al producto escalar usual cuando las masas son distintas.

En efecto, a partir de la relación $\vec{v}_n = \mathbf{M}^{-1/2} \vec{u}_n$, se obtiene

$$\vec{v}_n^T \mathbf{M} \vec{v}_m = \vec{u}_n^T \vec{u}_m.$$

Por lo tanto, si los autovectores $\vec{u}_n$ de la matriz simetrizada se eligen ortonormales con el producto escalar euclídeo, como en la ec. (4.3), los autovectores físicos satisfacen una relación de ortogonalidad natural con respecto al producto escalar inducido por la matriz de masa:

$$\vec{v}_n^T \mathbf{M} \vec{v}_m = \delta_{nm}. \quad (4.4)$$

En cambio, si los vectores $\vec{u}_n$ no se encuentran normalizados de acuerdo con (4.3), se obtiene

$$\vec{v}_n^T \mathbf{M} \vec{v}_m = \|\vec{v}_n\|_{\mathbf{M}}^2 \delta_{nm}, \quad (4.5)$$

donde $\|\vec{v}\|_{\mathbf{M}} = \sqrt{\vec{v}^T \mathbf{M} \vec{v}}$ es la norma inducida por la matriz de masa.

Además, se verifica que

$$\|\vec{v}_n\|_{\mathbf{M}}^2 = \vec{v}_n^T \mathbf{M} \vec{v}_n = \vec{u}_n^T \vec{u}_n = \|\vec{u}_n\|^2,$$

lo que muestra que la transformación entre $\vec{u}_n$ y $\vec{v}_n$ preserva la norma, siempre que se consideren los productos escalares naturales en cada espacio. Este resultado pone de manifiesto que la noción de ortogonalidad en el espacio físico no es la euclídea usual, sino que está determinada por la métrica inducida por la matriz de masa.

Es interesante recapitular el camino seguido hasta aquí. En primer lugar, el problema se formuló en términos de la matriz dinámica, lo que permitió tratar de manera directa el caso en que dicha matriz es simétrica, mediante la ortogonalidad euclídea de sus autovectores. En el caso general, esta estrategia condujo a la introducción de la matriz $\hat{\mathbb{K}}$, cuyos autovectores pueden elegirse ortogonales con respecto al producto escalar usual.

En este contexto, la matriz de masa queda implícitamente incorporada en la definición de los autovectores físicos, lo que da lugar de manera natural a una noción de ortogonalidad ponderada por $\mathbb{M}$. Desde este punto de vista, la norma inducida por la matriz de masa no constituye una hipótesis adicional, sino una consecuencia directa de la reformulación del problema en términos de $\hat{\mathbb{K}}$.

Alternativamente, es posible abordar el problema desde un punto de vista más general, considerando directamente el problema generalizado de autovalores

$$\mathbb{K} \vec{v} = \omega^2 \mathbb{M} \vec{v},$$

en el cual la matriz de masa interviene explícitamente desde el comienzo. En esta formulación, la normalización con respecto a $\mathbb{M}$ surge de manera natural como condición para diagonalizar simultáneamente las matrices $\mathbb{M}$ y $\mathbb{K}$, que más adelante se interpretarán en términos de la energía cinética y potencial del sistema.

Ambos enfoques son completamente equivalentes desde el punto de vista matemático. El primero resulta más directo y concreto, mientras que el segundo pone de manifiesto con mayor claridad la estructura general del problema.

En lo que sigue, adoptaremos alternativamente uno u otro punto de vista según resulte más conveniente para el análisis.

4.2. El caso de estudio con masas y resortes distintos

Como ejemplo de aplicación del formalismo modal para una matriz dinámica no simétrica, consideremos el sistema totalmente asimétrico de dos masas acopladas distintas y resortes distintos. Las ecuaciones de movimiento de este sistema ya fueron

analizadas: están dadas por la ecuación (2.6) y pueden reescribirse como

$$\ddot{\vec{\psi}} = -\tilde{\mathbf{K}} \vec{\psi}, \quad (4.6)$$

donde la matriz dinámica está dada por

$$\tilde{\mathbf{K}} = \mathbf{M}^{-1} \mathbf{K}.$$

En este ejemplo la matriz de masa es

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \quad (4.7)$$

mientras que la matriz de constantes elásticas tiene la forma

$$\mathbf{K} = \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 + k_3 \end{bmatrix}. \quad (4.8)$$

Por lo tanto, la matriz dinámica resulta

$$\tilde{\mathbf{K}} = \begin{bmatrix} \frac{k_1 + k_2}{m_1} & -\frac{k_2}{m_1} \\ -\frac{k_2}{m_2} & \frac{k_2 + k_3}{m_2} \end{bmatrix}. \quad (4.9)$$

Si las masas son distintas, esta matriz ya no es simétrica. Sin embargo, como vimos anteriormente, siempre puede escribirse en la forma

$$\tilde{\mathbf{K}} = \mathbf{M}^{-1/2} \hat{\mathbf{K}} \mathbf{M}^{1/2},$$

donde

$$\hat{\mathbf{K}} = \mathbf{M}^{-1/2} \mathbf{K} \mathbf{M}^{-1/2}.$$

Así obtenemos una matriz manifiestamente simétrica

$$\hat{\mathbf{K}} = \begin{bmatrix} \frac{k_1 + k_2}{m_1} & -\frac{k_2}{\sqrt{m_1 m_2}} \\ -\frac{k_2}{\sqrt{m_1 m_2}} & \frac{k_2 + k_3}{m_2} \end{bmatrix}. \quad (4.10)$$

que coincide con $\tilde{\mathbf{K}}$ cuando $m_1 = m_2$. Si bien simétrica, esta matriz tiene la forma *abbc*, en vez de la forma *abba* de la ecuación (3.23). Entonces, sus autovectores $\vec{u}_n$ ya no son

las direcciones $(1, 1)$ y $(1, -1)$. Es fácil demostrar que vienen dados por

$$\vec{u}_n \propto \begin{bmatrix} b \\ \lambda_n - a \end{bmatrix}, \quad (4.11)$$

con λ_n los autovalores

$$\lambda_n = \frac{a+c}{2} \pm \sqrt{\left(\frac{a-c}{2}\right)^2 + b^2}. \quad (4.12)$$

Para una matriz simétrica de 2×2 , los autovectores asociados a autovalores distintos son ortogonales entre sí. Esto significa que pueden elegirse como un par de vectores unitarios que forman una base ortonormal del plano. En consecuencia, siempre es posible escribirlos en la forma

$$\vec{u}_1 = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}, \quad (4.13)$$

para cierto ángulo θ . Usando la expresión (4.11), se obtiene que el ángulo θ satisface

$$\tan(2\theta) = \frac{2b}{a-c}. \quad (4.14)$$

Desde este punto de vista, diagonalizar la matriz $\hat{\mathbf{K}}$ equivale simplemente a realizar una rotación del sistema de coordenadas en el espacio de configuraciones. El ángulo θ determina las combinaciones lineales de las coordenadas originales que desacoplan las ecuaciones de movimiento. En el caso particular en que el sistema es completamente simétrico ($m_1 = m_2$ y $k_1 = k_3$), la matriz tiene la forma de la ecuación (3.23), $\tan(2\theta) \rightarrow \infty$, lo que implica $\theta = \pi/4$, y los autovectores corresponden exactamente a las direcciones $\vec{u}_{1s} = (1, 1)$ y $\vec{u}_{2s} = (1, -1)$ (ver figura 4.1), es decir, a una rotación de 45° en el espacio de configuraciones (el espacio de los vectores $\vec{\psi}$).

Cuando las masas o las constantes elásticas dejan de ser iguales, esa simetría se pierde. Los modos normales siguen siendo direcciones ortogonales del plano, pero ya no coinciden exactamente con $(1, 1)$ y $(1, -1)$. En cambio, corresponden a un par de direcciones —las direcciones $\vec{u}_1$ y $\vec{u}_2$ de la figura 4.1— rotadas por un ángulo θ que depende de los parámetros del sistema. En este sentido, la asimetría del sistema se manifiesta simplemente como un cambio en el ángulo que define las combinaciones de coordenadas que desacoplan el movimiento.

Como los autovectores físicos $\vec{v}_n$ se obtienen aplicando la transformación $M^{-1/2}$ a los autovectores $\vec{u}_n$, la geometría del espacio de configuraciones se deforma. En consecuencia, los autovectores físicos aparecen como una base oblicua cuando se utilizan las coordenadas físicas ψ_1, ψ_2 , como se ilustra esquemáticamente en la figura 4.1. Sin embargo, estos autovectores son ortogonales en el espacio «deformado» por la transformación $M^{-1/2}$, donde hay otra métrica y vale la relación de ortogonalidad generalizada (4.4).

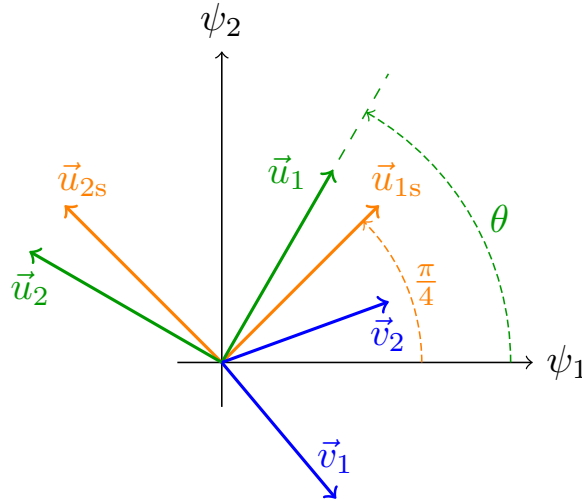


Figura 4.1: Representación esquemática de distintas bases en el espacio de configuraciones (ψ_1, ψ_2) . Los ejes cartesianos corresponden a las coordenadas físicas del sistema. La base ortonormal $(\vec{u}_1, \vec{u}_2)$, asociada a la matriz simetrizada $\tilde{\mathbf{K}}$, aparece rotada un ángulo θ respecto de estos ejes. En el caso simétrico, dicha base coincide con $(\vec{u}_{1s}, \vec{u}_{2s})$, rotada un ángulo $\pi/4$. Los autovectores físicos $(\vec{v}_1, \vec{v}_2)$, en cambio, forman en general una base oblicua cuando se representan en el sistema de coordenadas (ψ_1, ψ_2) . La diferencia entre los ángulos θ y $\pi/4$ puede interpretarse como una medida de la asimetría del sistema.

Conviene notar que esta interpretación geométrica no es particular del ejemplo considerado. En efecto, en el caso de dos grados de libertad, cualquier par de modos normales puede describirse mediante un único ángulo θ y la diferencia $\theta - \pi/4$ puede interpretarse como una medida de la ruptura de simetría del sistema: cuanto mayor es la desviación respecto de 45° , mayor es la desigualdad entre las propiedades dinámicas asociadas a las dos coordenadas originales.

4.3. Desacople modal completo

Ahora estamos en condiciones de ver que el procedimiento de desacople modal puede formularse de manera unificada tanto para el caso en que la matriz dinámica $\tilde{\mathbf{K}}$ es simétrica como para el caso general. La idea central consiste en encontrar un cambio de base que diagonalice la matriz dinámica y permita escribir las ecuaciones de movimiento como un conjunto de ecuaciones independientes.

En ambos casos seguimos los pasos esquematizados en la subsección 3.5. Partimos de la ecuación matricial (3.15) y observamos que la matriz dinámica *siempre posee un conjunto completo de autovectores linealmente independientes*, tanto para el caso simétrico como para el caso general. Con estos autovectores como columnas, se construye la matriz $\mathbb{D}$, que diagonaliza a $\tilde{\mathbf{K}}$ mediante la transformación de semejanza $\mathbb{D}^{-1} \tilde{\mathbf{K}} \mathbb{D}$.

En el caso simétrico, las columnas de $\mathbb{D}$ pueden elegirse ortonormales en el sentido usual. En cambio, en el caso general, las columnas de $\mathbb{D}$ están formadas por los autovectores físicos $\vec{v}_n$, que satisfacen una relación de ortogonalidad ponderada por la matriz

de masa. Sin embargo, esto no es obstáculo para diagonalizar a $\tilde{\mathbb{K}}$ y llegar al resultado obtenido en la ecuación (3.16),

$$\ddot{\vec{\Phi}} = -\mathbb{D}^{-1}\tilde{\mathbb{K}}\mathbb{D}\vec{\Phi}(t). \quad (4.15)$$

donde

$$\vec{\Phi}(t) = \mathbb{D}^{-1}\vec{\psi}(t),$$

es el vector de coordenadas modales.

$\mathbb{D}^{-1}\tilde{\mathbb{K}}\mathbb{D}$ es una matriz diagonal. Si a sus elementos diagonales los llamamos ω_i^2 , $i = 1, \dots, N$, el sistema queda finalmente

$$\ddot{\Phi}_i + \omega_i^2 \Phi_i = 0, \quad i = 1, \dots, N. \quad (4.16)$$

De este modo, las ecuaciones de movimiento han quedado desacopladas. Y cada coordenada modal satisface la ecuación de un oscilador armónico independiente, cuyas soluciones son

$$\Phi_i(t) = A_i \cos(\omega_i t + \varphi_i). \quad (4.17)$$

Vemos así que la diferencia entre el caso simétrico y el caso general no reside en la forma final de estas ecuaciones, sino en la estructura de la matriz $\mathbb{D}$. En ambos casos, el resultado es el mismo: el movimiento del sistema puede interpretarse como una superposición de modos normales independientes.

4.4. Condiciones iniciales y no ortogonalidad

Si bien la no ortogonalidad (en el sentido usual) de las columnas de $\mathbb{D}$ no impidió, en el caso asimétrico, obtener el desarrollo de la perturbación $\vec{\psi}(t)$ como combinación lineal de N modos normales independientes, veremos a continuación que dicha no ortogonalidad si afecta la manera en que se usan las condiciones iniciales para obtener las constantes A_i y φ_i , $i = 1, \dots, N$.

Consideremos la descomposición de la perturbación en la base de autovectores físicos

$$\vec{\psi}(t) = \sum_{n=1}^N \Phi_n(t) \vec{v}_n. \quad (4.18)$$

En el caso simétrico, donde los autovectores forman una base ortonormal en el sentido usual, los coeficientes modales se obtienen directamente mediante el producto escalar euclidiano: cuando se multiplica miembro a miembro la igualdad (4.18) por el vector $\vec{v}_m$ de la base modal, $m = 1, \dots, N$, se anulan todos los términos de la sumatoria, excepto el término m -ésimo

$$\vec{v}_m^T \vec{\psi}(t) = \sum_{n=1}^N \Phi_n(t) \underbrace{\vec{v}_m^T \vec{v}_n}_{\delta_{nm}} = \Phi_m(t).$$

Al evaluar la igualdad anterior en $t = 0$, vemos que el valor inicial de cada coordenada modal no es otra cosa que la proyección del vector perturbación inicial sobre su correspondiente autovector

$$\Phi_n(0) = \vec{v}_n^T \vec{\psi}(0) = A_n \cos \varphi_n.$$

Análogamente, el valor inicial de la velocidad de cada coordenada modal no es otra cosa que la proyección del vector velocidad inicial sobre su correspondiente autovector

$$\dot{\Phi}_n(0) = \vec{v}_n^T \dot{\vec{\psi}}(0) = -\omega_n A_n \sin \varphi_n.$$

En el caso general, en cambio, los autovectores físicos no son ortogonales con respecto a este producto escalar. Para lograr aislar un solo término del desarrollo modal se debe usar la relación de ortonormalidad generalizada (4.4)

$$\vec{v}_m^T \mathbf{M} \vec{v}_n = \delta_{nm}.$$

Con este fin, multiplicamos la igualdad (4.18), miembro a miembro por izquierda, por $\vec{v}_m^T \mathbf{M}$

$$\vec{v}_m^T M \vec{\psi}(t) = \sum_{n=1}^N \Phi_n(t) \underbrace{\vec{v}_m^T \mathbf{M} \vec{v}_n}_{\delta_{nm}} = \Phi_m(t), \quad (4.19)$$

de donde resulta que las funciones modales están dadas por

$$\Phi_m(t) = \vec{v}_m^T \mathbf{M} \vec{\psi}(t), \quad m = 1, \dots, N. \quad (4.20)$$

En particular, evaluando la expresión anterior en $t = 0$, se obtienen relaciones entre las constantes A_n y φ_n y las condiciones iniciales $\vec{\psi}(0)$ y $\dot{\vec{\psi}}(0)$,

$$\begin{aligned} \Phi_n(0): \quad \vec{v}_n^T \mathbf{M} \vec{\psi}(0) &= A_n \cos \varphi_n, \\ \dot{\Phi}_n(0): \quad \vec{v}_n^T \mathbf{M} \dot{\vec{\psi}}(0) &= -\omega_n A_n \sin \varphi_n. \end{aligned} \quad (4.21)$$

A_n y φ_n se obtienen exactamente del mismo modo que en el caso simétrico, pero utilizando estas proyecciones generalizadas en lugar del producto escalar usual. En ambos casos, el procedimiento sigue los mismos pasos que se usaron en la ecuación (3.9). Para eliminar la fase φ_n , se elevan al cuadrado ambas ecuaciones y se las suman convenientemente, como para luego usar la identidad $\cos^2 \varphi_n + \sin^2 \varphi_n = 1$ y así aislar A_n^2 en función de las condiciones iniciales. Y para eliminar A_n , se dividen ambas ecuaciones miembro a miembro y así se obtiene una expresión para la tangente de φ_n en función de las condiciones iniciales.

Queda demostrado entonces que la falta de ortogonalidad euclidiana de los autovectores físicos no afecta el desacople de las ecuaciones de movimiento, pero modifica la forma en que las condiciones iniciales se proyectan sobre los modos normales.

4.5. Diagonalización y energía

El procedimiento de desacople modal presentado en la sección anterior se formuló en términos de la matriz dinámica $\tilde{\mathbb{K}}$, cuya diagonalización permite reducir las ecuaciones de movimiento a un conjunto de ecuaciones independientes entre sí. Como se discutió previamente, esta formulación incorpora de manera implícita la matriz de masa.

Esta construcción puede reinterpretarse poniendo el énfasis en las *formas cuadráticas que describen la energía cinética y potencial* del sistema. En este enfoque, el problema se expresa directamente en términos de las matrices originales $\mathbb{M}$ (diagonal en los casos considerados, como consecuencia de haber elegido coordenadas cartesianas independientes, aunque no necesariamente en general) y $\mathbb{K}$.

A continuación veremos que las mismas técnicas utilizadas para diagonalizar la matriz dinámica pueden adaptarse para transformar simultáneamente las matrices $\mathbb{M}$ y $\mathbb{K}$, de modo que las formas cuadráticas de la energía queden desacopladas, sin productos cruzados entre coordenadas ni entre sus derivadas. En estas condiciones, cada coordenada normal contribuye de manera independiente a la energía del sistema.

Comentario. La descomposición en coordenadas normales, en la cual la energía del sistema se expresa como suma de contribuciones independientes, constituye una idea de alcance general en física. En particular, en el estudio de la radiación del cuerpo negro, el campo electromagnético puede descomponerse en modos independientes, cada uno de los cuales se comporta como un oscilador armónico clásico. Esta estructura permite interpretar la energía total como suma de contribuciones modales, un punto de partida clave en la formulación de la ley de Planck.

Las energías cinética y potencial pueden escribirse como

$$T = \frac{1}{2} \dot{\vec{\psi}}^T \mathbb{M} \dot{\vec{\psi}}, \quad V = \frac{1}{2} \vec{\psi}^T \mathbb{K} \vec{\psi}. \quad (4.22)$$

Para fijar ideas, resulta instructivo verificar estas expresiones en el sistema de dos masas y tres resortes considerado en la sección 4.2, con $\mathbb{M}$ y $\mathbb{K}$ dados por (4.7) y (4.8) respectivamente.

La energía cinética del sistema es

$$T(t) = \frac{1}{2} m_1 [\dot{\psi}_1]^2 + \frac{1}{2} m_2 [\dot{\psi}_2]^2 \quad (4.23)$$

lo que refleja que, en las coordenadas elegidas, cada grado de libertad contribuye de manera independiente a la energía cinética, sin términos cruzados entre velocidades. Esta forma particularmente simple de la energía cinética está ligada a la elección de coordenadas, y no es una propiedad general de sistemas mecánicos arbitrarios.

En cambio, la energía potencial elástica es

$$V(t) = \frac{k_1}{2} [\psi_1]^2 + \frac{k_2}{2} [\psi_2 - \psi_1]^2 + \frac{k_3}{2} [\psi_3]^2 \quad (4.24)$$

que al desarrollar puede escribirse como

$$V(t) = \frac{1}{2}(k_1 + k_2)[\psi_1]^2 - k_2\psi_1\psi_2 + \frac{1}{2}(k_2 + k_3)[\psi_2]^2. \quad (4.25)$$

Se trata, por lo tanto, de una forma cuadrática en los desplazamientos, que: i) está asociada con la matriz $\mathbb{K}$ y ii) contiene términos cruzados que reflejan el acoplamiento entre los grados de libertad. Puede observarse que los términos cruzados aparecen en combinaciones simétricas, lo que garantiza que la energía potencial pueda escribirse en la forma matricial (4.22). Es inmediato comprobar que, con las matrices $\mathbb{M}$ y $\mathbb{K}$ dadas, ambas expresiones se recuperan a partir de (4.22).

La presencia de los términos cruzados en la energía pone de manifiesto que las coordenadas ψ_i no son, en general, las variables más adecuadas para describir la dinámica del sistema. En efecto, aun cuando en el sistema considerado la energía cinética se encuentra desacoplada, como consecuencia de que cada grado de libertad describe el movimiento de una masa a lo largo de una dirección fija e independiente, esto no ocurre en general. Por ejemplo, en el caso de un doble péndulo en el régimen de pequeñas oscilaciones, puede verificarse que la energía cinética adopta la forma

$$T = \frac{1}{2}(m_1 + m_2)l_1^2\dot{\theta}_1^2 + \frac{1}{2}m_2l_2^2\dot{\theta}_2^2 + m_2l_1l_2\dot{\theta}_1\dot{\theta}_2,$$

lo que muestra que aparecen términos cruzados entre las velocidades. En consecuencia, en este tipo de sistemas el desacople requiere transformar simultáneamente tanto la energía cinética como la potencial.

Como ya se ha visto en el análisis del desacople modal, este acoplamiento puede eliminarse mediante un cambio de coordenadas apropiado, construido a partir de los autovectores de la matriz dinámica. La forma concreta que adopta esta transformación depende, sin embargo, del carácter de la matriz de masas, lo que lleva naturalmente a distinguir los casos ya analizados:

- **Masas iguales** ($\mathbb{M} = m\mathbb{I}$): En este caso, la energía cinética es proporcional a la identidad y, por lo tanto, los autovectores de la matriz dinámica pueden elegirse ortonormales en el sentido usual, es decir, $\mathbb{D}^T\mathbb{D} = \mathbb{I}$. En consecuencia,

$$\mathbb{D}^T\mathbb{M}\mathbb{D} = m\mathbb{I},$$

y la transformación definida por $\mathbb{D}$ elimina los términos cruzados en la energía potencial sin introducir acoplamientos en la energía cinética.

- **Masas distintas** ($\mathbb{M}$ diagonal): En este caso, la matriz de masas ya no es proporcional a la identidad. Sin embargo, la libertad en la elección de la normalización de los autovectores permite imponer la condición

$$\mathbb{D}^T\mathbb{M}\mathbb{D} = \mathbb{I},$$

lo que equivale a elegirlos ortonormales respecto de la forma cuadrática definida por $\mathbb{M}$. Con esta elección, la transformación vuelve a eliminar los términos cruzados, ahora de manera consistente en ambas formas cuadráticas.

En ambos casos, al introducir el cambio de variables $\vec{\psi} = \mathbb{D}\vec{\Phi}$, que en términos de las matrices originales equivale a resolver el problema $\mathbb{K}\vec{a} = \omega^2\mathbb{M}\vec{a}$, las energías cinética y potencial adoptan la forma

$$T = \frac{1}{2} \dot{\vec{\Phi}}^T (\mathbb{D}^T \mathbb{M} \mathbb{D}) \dot{\vec{\Phi}}, \quad V = \frac{1}{2} \vec{\Phi}^T (\mathbb{D}^T \mathbb{K} \mathbb{D}) \vec{\Phi}.$$

Utilizando la ecuación de autovalores que define los modos normales, se obtiene

$$\mathbb{D}^T \mathbb{K} \mathbb{D} = \text{diag}(\omega_1^2, \omega_2^2, \dots),$$

mientras que, según el caso considerado, la matriz $\mathbb{D}^T \mathbb{M} \mathbb{D}$ es diagonal y proporcional a la identidad, o bien la identidad misma. En consecuencia, las energías pueden escribirse como

$$T = \frac{1}{2} \sum_i \dot{\Phi}_i^2, \quad V = \frac{1}{2} \sum_i \omega_i^2 \Phi_i^2,$$

salvo, en el caso de masas iguales, por un factor constante global en la energía cinética.

De este modo, el cambio de coordenadas definido por $\mathbb{D}$ permite eliminar los términos cruzados en las formas cuadráticas asociadas a la energía, dando lugar a una descripción en la cual cada modo contribuye de manera independiente a la dinámica del sistema.

Esta propiedad puede interpretarse como una consecuencia de la estructura particular del problema conservativo: la dinámica está completamente determinada por dos formas cuadráticas, asociadas a la energía cinética y potencial, que pueden reducirse simultáneamente a una suma de contribuciones independientes mediante un cambio de coordenadas adecuado.

Esta situación se modifica de manera esencial al incorporar efectos disipativos o forzados. En presencia de disipación aparece, en general, una nueva contribución, típicamente dependiente de las velocidades, que no puede, en general, reducirse simultáneamente con las anteriores a una forma sin términos cruzados. Como consecuencia, el desacople completo de la dinámica deja de ser posible, y los modos normales del sistema conservativo pasan a desempeñar el papel de una base conveniente para describir una dinámica que es, en esencia, acoplada.

4.6. Ejemplo integrador: modos normales y regímenes de acoplamiento

Con el objetivo de integrar y profundizar los distintos aspectos del análisis modal desarrollados hasta aquí, retomemos el sistema de dos masas acopladas mediante tres resortes. Este ejemplo permitirá ilustrar el procedimiento de desacople en el caso asimétrico, la interpretación geométrica del problema, el papel del producto escalar ponderado y la aparición de distintos regímenes dinámicos según los parámetros del sistema.

4.6.1. Autovalores

Recordemos que este sistema tiene una matriz dinámica $\tilde{\mathbf{K}}$, en general no simétrica, dada por la ecuación (4.9) y una matriz simetrizada $\hat{\mathbf{K}}$, de la forma $abbc$, dada por la ecuación (4.10). Si bien $\hat{\mathbf{K}}$ es simétrica por construcción (ver ecuación (4.1)), es oportuno señalar que la asimetría del sistema físico se manifiesta en la diferencia entre elementos diagonales, mientras que el acoplamiento entre coordenadas está contenido en el término proporcional a k_2^2 .

Como $\tilde{\mathbf{K}}$ y $\hat{\mathbf{K}}$ son semejantes, tienen los mismos autovalores, que se obtienen a partir de la ecuación (4.12)

$$\lambda_{\pm} = \frac{1}{2} \left(\frac{k_1 + k_2}{m_1} + \frac{k_2 + k_3}{m_2} \right) \pm \frac{1}{2} \sqrt{\left(\frac{k_1 + k_2}{m_1} - \frac{k_2 + k_3}{m_2} \right)^2 + \frac{4k_2^2}{m_1 m_2}}, \quad (4.26)$$

y corresponden a los cuadrados de las frecuencias propias $\omega_{\pm}^2$.

4.6.2. Autovectores

Según la ecuación (4.11), los autovectores de $\hat{\mathbf{K}}$ son

$$\vec{u}_{\pm} \propto \begin{bmatrix} b \\ \lambda_{\pm} - a \end{bmatrix}, \quad (4.27)$$

y los autovectores físicos (ver (4.2)) resultan

$$\vec{v}_{\pm} = \frac{1}{g_{\pm}} \begin{bmatrix} 1/\sqrt{m_1} & 0 \\ 0 & 1/\sqrt{m_2} \end{bmatrix} \begin{bmatrix} b \\ \lambda_{\pm} - a \end{bmatrix} = \frac{1}{g_{\pm}} \begin{bmatrix} b/\sqrt{m_1} \\ (\lambda_{\pm} - a)/\sqrt{m_2} \end{bmatrix}, \quad (4.28)$$

donde $g_{\pm}$

$$g_{\pm}^2 = b^2 + (\lambda_{\pm} - a)^2. \quad (4.29)$$

son factores que normalizan euclídeamente a $\vec{u}_{\pm}$. Esta normalización garantiza automáticamente la normalización respecto a la matriz de masa de los autovectores físicos $\vec{v}_{\pm}$.

Comparando las ecuaciones (4.27) y (4.28), observamos los vectores $\vec{v}_{\pm}$ se obtienen mediante un reescalamiento componente a componente de los vectores $\vec{u}_{\pm}$. Cada coordenada se divide por la raíz cuadrada de la masa asociada, adaptando así los autovectores de la matriz simetrizada al peso dinámico de cada grado de libertad. Este resultado es siempre válido, porque proviene directamente de la semejanza entre $\tilde{\mathbf{K}}$ y $\hat{\mathbf{K}}$.

Las expresiones anteriores resultan particularmente adecuadas para su implementación directa en código, ya que permiten construir los autovectores a partir de los autovalores sin recurrir a funciones trigonométricas. En cambio, la parametrización angular —utilizada en la ecuación (4.13)— es útil para interpretar geoméricamente el

desacople, aunque a costa de introducir pasos intermedios que no son imprescindibles en un tratamiento explícito. En este punto es oportuno señalar que este problema, como tantos otros problemas físicos, admite más de una formulación igualmente válida. La elección entre ellas depende del objetivo: en este caso la formulación que estamos desarrollando resulta adecuada para obtener resultados numéricos explícitos, mientras que la parametrización de los autovectores en términos del ángulo θ hace más transparente la interpretación geométrica.

4.6.3. Normalización y desarrollo modal

Notemos que $\vec{u}_{\pm}$ y $\vec{v}_{\pm}$ están definidos a menos de un factor multiplicativo arbitrario. En aplicaciones prácticas —por ejemplo, en la implementación numérica del problema— resulta conveniente fijar este factor mediante una condición de normalización.

Como se discutió en 3.5.1, la normalización no es imprescindible, ya que si los autovectores no están normalizados, los factores de escala asociados pueden absorberse en las amplitudes de cada modo. No obstante, la normalización resulta conveniente porque permite identificar directamente las amplitudes modales a partir de proyecciones simples de las condiciones iniciales, evitando factores adicionales que pueden dificultar la interpretación de los resultados.

En particular, los autovectores $\vec{u}_{\pm}$ de la matriz simetrizada pueden elegirse ortonormales con respecto al producto escalar euclídeo, como en la ec. (4.3). En tal caso, los autovectores físicos $\vec{v}_{\pm}$ satisfacen automáticamente la relación de ortogonalidad generalizada (4.4).

De este modo, la normalización euclídea de los autovectores de $\hat{\mathbb{K}}$ induce de manera directa la normalización natural de los autovectores físicos con respecto a la matriz de masa. En estas condiciones, dichos autovectores constituyen bases particularmente convenientes para descomponer el movimiento en modos normales.

Una vez determinados $\lambda_{\pm}$ y $\vec{v}_{\pm}$, el problema dinámico queda completamente caracterizado y la evolución de la perturbación $\vec{\psi}(t)$ se puede calcular mediante el siguiente desarrollo modal

$$\vec{\psi}(t) = A_+ \cos(\omega_+ t + \varphi_+) \vec{v}_+ + A_- \cos(\omega_- t + \varphi_-) \vec{v}_-, \quad (4.30)$$

4.6.4. Condiciones iniciales

En este caso ($N = 2$), las ecuaciones (4.21) son cuatro igualdades

$$\begin{aligned} \vec{v}_+^T \mathbf{M} \vec{\psi}(0) &= A_+ \cos \varphi_+, & \vec{v}_+^T \mathbf{M} \dot{\vec{\psi}}(0) &= -\omega_+ A_+ \sin \varphi_+, \\ \vec{v}_-^T \mathbf{M} \vec{\psi}(0) &= A_- \cos \varphi_-, & \vec{v}_-^T \mathbf{M} \dot{\vec{\psi}}(0) &= -\omega_- A_- \sin \varphi_-. \end{aligned} \quad (4.31)$$

El lado izquierdo de estas igualdades contiene proyecciones de las condiciones iniciales $\vec{\psi}(0)$ y $\dot{\vec{\psi}}(0)$. A continuación se muestra el mecanismo que siguen las proyecciones de $\vec{\psi}(0)$ sobre $\vec{v}_\pm$

$$\frac{1}{g_\pm} \left[\frac{b}{\sqrt{m_1}} \quad \frac{\lambda_\pm - a}{\sqrt{m_2}} \right] \begin{bmatrix} m_1 \psi_1(0) \\ m_2 \psi_2(0) \end{bmatrix} = \frac{1}{g_\pm} [\sqrt{m_1} b \psi_1(0) + \sqrt{m_2} (\lambda_\pm - a) \psi_2(0)] .$$

donde se eligió hacer primero el producto de $\mathbf{M}$ por $\vec{\psi}(0)$. El mecanismo para las proyecciones de $\dot{\vec{\psi}}(0)$ es completamente análogo.

Usando estos resultados en el lado izquierdo de las ecuaciones (4.31), se obtiene el ya familiar sistema de ecuaciones

$$\begin{aligned} A_+ \cos \varphi_+ &= \frac{1}{g_+} [\sqrt{m_1} b \psi_1(0) + \sqrt{m_2} (\lambda_+ - a) \psi_2(0)] , \\ -\omega_+ A_+ \sin \varphi_+ &= \frac{1}{g_+} [\sqrt{m_1} b \dot{\psi}_1(0) + \sqrt{m_2} (\lambda_+ - a) \dot{\psi}_2(0)] , \\ A_- \cos \varphi_- &= \frac{1}{g_-} [\sqrt{m_1} b \psi_1(0) + \sqrt{m_2} (\lambda_- - a) \psi_2(0)] , \\ -\omega_- A_- \sin \varphi_- &= \frac{1}{g_-} [\sqrt{m_1} b \dot{\psi}_1(0) + \sqrt{m_2} (\lambda_- - a) \dot{\psi}_2(0)] . \end{aligned} \tag{4.32}$$

Las constantes A_+ , A_- , φ_+ y φ_- se obtienen siguiendo los procedimientos explicados para el caso general en la sección 4.4 y empleados en el caso simétrico en el contexto de la ecuación (3.9).

4.6.5. Regímenes de acoplamiento

Las expresiones (4.12) y (4.26) obtenidas para los autovalores, ponen de manifiesto que las frecuencias propias del sistema resultan de la competencia entre dos efectos: por un lado, la asimetría asociada a la diferencia entre los términos diagonales ($|a - c|$), y por otro, el acoplamiento entre las coordenadas, contenido en el término fuera de la diagonal (b).

En esta competencia se manifiestan explícitamente las frecuencias características ω_1 y ω_2 que tendría cada parte móvil en ausencia de acoplamiento. En efecto, los términos diagonales a y c corresponden a los cuadrados de esas frecuencias propias individuales, modificadas por la contribución del resorte intermedio.

El término proporcional a b^2 cuantifica el acoplamiento entre las coordenadas. La combinación $a + c$

$$a + c = \frac{k_1}{m_1} + \frac{k_3}{m_2} + k_2 \left(\frac{1}{m_1} + \frac{1}{m_2} \right) = \omega_1^2 + \omega_2^2 + k_2 \left(\frac{1}{m_1} + \frac{1}{m_2} \right),$$

involucra la suma entre las frecuencias propias individuales al cuadrado, modificada por un término que es cero en ausencia de acoplamiento. Por último, la combinación

$a - c$

$$a - c = \frac{k_1}{m_1} - \frac{k_3}{m_2} + k_2\left(\frac{1}{m_1} - \frac{1}{m_2}\right) = \omega_1^2 - \omega_2^2 + k_2\left(\frac{1}{m_1} - \frac{1}{m_2}\right),$$

involucra la diferencia entre las frecuencias propias al cuadrado, también modificada por otro término que es cero en ausencia de acoplamiento.

Como resultado de esta competencia entre asimetría y acoplamiento, el sistema presenta comportamientos cualitativamente distintos según la relación entre los términos que intervienen en la raíz cuadrada.

En efecto, pueden distinguirse dos casos límite. Si $b^2 \ll (a-c)^2$, la raíz está dominada por el término $(a-c)$, y los autovalores se aproximan a a y c . En cambio, si $(a-c)^2 \ll b^2$, el término dominante es el acoplamiento, y la estructura de los autovalores cambia cualitativamente, reflejando la aparición de modos colectivos.

Estos dos límites dan lugar a los regímenes de acoplamiento débil y acoplamiento fuerte que analizamos a continuación.

Acoplamiento débil

Consideremos en primer lugar el caso en que el acoplamiento es débil, en el sentido de que $b^2 \ll (a-c)^2$. En este régimen, el término dominante dentro de la raíz es $(a-c)^2$, por lo que resulta natural desarrollar la expresión de los autovalores en potencias del pequeño parámetro $b^2/(a-c)^2$.

En efecto, escribiendo

$$\sqrt{\left(\frac{a-c}{2}\right)^2 + b^2} = \frac{|a-c|}{2} \sqrt{1 + \frac{4b^2}{(a-c)^2}},$$

y desarrollando la raíz para valores pequeños de $b^2/(a-c)^2$, obtenemos

$$\sqrt{\left(\frac{a-c}{2}\right)^2 + b^2} \approx \frac{|a-c|}{2} \left(1 + \frac{2b^2}{(a-c)^2}\right).$$

Sustituyendo en la expresión de los autovalores, se encuentra que, a primer orden,

$$\lambda_+ \approx \max(a, c) + \frac{b^2}{|a-c|}, \quad \lambda_- \approx \min(a, c) - \frac{b^2}{|a-c|}. \quad (4.33)$$

Vemos así que, en este régimen, las frecuencias propias del sistema son próximas a las frecuencias características de cada parte móvil en ausencia de acoplamiento, con correcciones pequeñas debidas a la interacción entre ellas. En consecuencia, cada modo normal está dominado por el movimiento de una de las masas, con una participación secundaria de la otra. Desde el punto de vista geométrico, los autovectores están casi alineados con los ejes coordenados, lo que corresponde a un ángulo θ pequeño en la diagonalización de la matriz $\hat{\mathbf{K}}$ (ver ecuación (4.14)).

Acoplamiento fuerte

Consideremos ahora el caso opuesto, en el que $(a - c)^2 \ll b^2$. En este régimen, el término dominante dentro de la raíz es el acoplamiento, por lo que resulta conveniente desarrollar la expresión de los autovalores en potencias del pequeño parámetro $(a - c)^2/b^2$.

Escribimos entonces

$$\sqrt{\left(\frac{a - c}{2}\right)^2 + b^2} = |b| \sqrt{1 + \frac{(a - c)^2}{4b^2}},$$

y para valores pequeños de $(a - c)^2/b^2$, obtenemos

$$\sqrt{\left(\frac{a - c}{2}\right)^2 + b^2} \approx |b| \left(1 + \frac{(a - c)^2}{8b^2}\right).$$

Sustituyendo en la expresión de los autovalores, resulta

$$\lambda_{\pm} \approx \frac{a + c}{2} \pm |b|. \quad (4.34)$$

En este régimen, las frecuencias propias están dominadas por el acoplamiento, y los modos normales corresponden a movimientos colectivos de ambas masas. Dicho de otra manera: el carácter colectivo de los modos normales no es una propiedad adicional del sistema, sino una consecuencia directa del acoplamiento entre las coordenadas.

La diferencia entre los autovalores es del orden de $2|b|$, lo que indica que es el acoplamiento el que fija la escala dominante del problema. En términos de las frecuencias propias, esto se traduce en una separación cuya magnitud está controlada por b , aunque no coincide directamente con $2|b|$ debido a la relación no lineal $\omega = \sqrt{\lambda}$.

En el caso particular en que $a = c$, correspondiente a un sistema simétrico, se tiene $a - c = 0$, por lo que la condición de acoplamiento fuerte se cumple automáticamente siempre que $b \neq 0$. Esto explica por qué, en este caso, los modos normales son necesariamente combinaciones colectivas de las coordenadas. Desde este punto de vista, el caso simétrico puede interpretarse como un caso límite de acoplamiento fuerte.

4.6.6. Estructura modal e interpretación geométrica

Para comprender mejor la naturaleza de los modos normales, resulta útil analizar la forma de los autovectores físicos obtenidos anteriormente,

$$\vec{v}_{\pm} \propto \begin{bmatrix} \sqrt{m_1} b \\ \sqrt{m_2} (\lambda_{\pm} - a) \end{bmatrix}.$$

El cociente entre sus componentes,

$$\frac{v_2}{v_1} \propto \frac{\lambda_{\pm} - a}{b},$$

determina la relación entre las coordenadas ψ_1 y ψ_2 en cada modo. En general, este cociente es finito, lo que indica que ambos modos normales involucran el movimiento simultáneo de las dos masas. En este sentido, los modos son colectivos: el acoplamiento impide asociar cada frecuencia propia a una única coordenada.

Desde el punto de vista geométrico, esto se traduce en que los modos normales definen direcciones privilegiadas en el espacio de configuraciones, que pueden describirse como una base rotada respecto de las coordenadas originales. En el caso de la matriz simetrizada $\hat{\mathbf{K}}$, estas direcciones corresponden a una base ortonormal obtenida mediante una rotación de ángulo θ (ver ecuación (4.14)).

En el régimen de acoplamiento débil, uno de los modos está dominado por una de las masas, lo que se refleja en un cociente v_2/v_1 muy grande o muy pequeño. Geométricamente, esto corresponde a una base casi alineada con los ejes originales.

En cambio, en el régimen de acoplamiento fuerte, las componentes de los autovectores son comparables, y los modos corresponden a combinaciones equilibradas de ambas coordenadas. En este caso, la base modal está rotada de manera significativa respecto de los ejes originales.

Finalmente, en el caso simétrico ($a = c$), se obtiene $\theta = \pi/4$, y los modos se reducen a las direcciones $(1, 1)$ y $(1, -1)$, recuperando la estructura previamente discutida. La desviación $\theta - \pi/4$ puede interpretarse así como una medida de la asimetría del sistema.

4.7. Modulación y batidos

Cuando las frecuencias propias son cercanas entre sí, la superposición de los modos normales da lugar a fenómenos de modulación en la amplitud, conocidos como batidos. En este caso, la evolución temporal de las coordenadas puede interpretarse como una oscilación rápida modulada por una envolvente de frecuencia mucho menor, determinada por la diferencia entre las frecuencias propias.

Este comportamiento no se observa en las simulaciones mostradas en las figuras 2.6 y 2.7, ya que los parámetros elegidos en esos casos no conducen a frecuencias propias suficientemente cercanas. Esto muestra que la simulación numérica, si bien permite explorar la dinámica del sistema, no siempre revela de manera directa todos los fenómenos posibles y su interpretación depende de una elección adecuada de los parámetros. En cambio, la descomposición en modos normales permite anticipar la aparición de la modulación de amplitud a partir de la estructura espectral del sistema.

Veremos a continuación que el análisis modal no solo explica el origen de este comportamiento, sino que además permite relacionar cuantitativamente sus características con la cercanía entre las frecuencias propias. Este ejemplo pone de manifiesto que la simulación numérica y el análisis teórico cumplen roles complementarios: mientras la primera permite explorar el comportamiento del sistema para elecciones particulares de parámetros, el segundo permite identificar los mecanismos generales y anticipar la aparición de nuevos fenómenos.

4.7.1. Batidos en un ejemplo: análisis e interpretación

Para ilustrar el fenómeno de manera concreta, volvamos el sistema empleado en las figuras 2.6 y 2.7: dos masas acopladas en una configuración simétrica, con masas iguales y constantes elásticas de los extremos iguales.

Las frecuencias propias del sistema son $\omega_1^2 = k/m$ y $\omega_2^2 = (k + 2k')/m$ (ver ecuación (3.6)). La descomposición en modos normales conduce a una solución donde la evolución de la perturbación de las partes móviles $\vec{\psi}(t)$ viene dada por la ecuación (3.25), que por comodidad repetimos a continuación

$$\begin{aligned}\psi_1(t) &= \frac{1}{2} \left[A_1 \cos(\omega_1 t + \varphi_1) + A_2 \cos(\omega_2 t + \varphi_2) \right], \\ \psi_2(t) &= \frac{1}{2} \left[A_1 \cos(\omega_1 t + \varphi_1) - A_2 \cos(\omega_2 t + \varphi_2) \right].\end{aligned}$$

Si al igual que en las simulaciones de las figuras 2.6 y 2.7 elegimos condiciones iniciales con velocidad inicial nula y si además solo perturbamos la parte móvil 1, las ecuaciones (3.10) y (3.11) muestran que $\varphi_1 = \varphi_2 = 0$ y $A_1 = A_2 = 2A$. Es decir, la perturbación de las partes móviles se escribe como superposición de cosenos de frecuencias ω_1 y ω_2 , con amplitudes y fases iniciales iguales. En la parte móvil 1 los cosenos se suman y en la parte móvil 2 se restan

$$\begin{aligned}\psi_1(t) &= A \left[\cos(\omega_1 t) + \cos(\omega_2 t) \right], \\ \psi_2(t) &= A \left[\cos(\omega_1 t) - \cos(\omega_2 t) \right].\end{aligned}\tag{4.35}$$

¿Qué podemos decir sobre el comportamiento de funciones que son la suma o a la resta de dos cosenos en fase, igual amplitud A y frecuencias distintas?

Se podría pensar que los resultados mostrados en la figura 2.6a —donde aparecen precisamente combinaciones de este tipo para valores particulares de las masas y constantes elásticas— podrían ofrecer algún indicio para responder a esta pregunta. Sin embargo, como puede apreciarse, la forma de las curvas no permite extraer conclusiones claras a simple vista.

Para avanzar en la interpretación, resulta conveniente reescribir estas expresiones utilizando identidades trigonométricas. Partimos de una identidad fundamental —de las pocas que solemos recordar— las identidades trigonométricas de suma de ángulos. Con su ayuda es fácil demostrar que

$$\begin{aligned}\cos(x + y) + \cos(x - y) &= 2 \cos x \cos y, \\ \cos(x + y) - \cos(x - y) &= -2 \sin x \sin y.\end{aligned}\tag{4.36}$$

Para usar estos resultados, identificamos $\omega_1 t$ con $x + y$ y $\omega_2 t$ con $x - y$: de esta forma tenemos las combinaciones de cosenos que aparecen en (4.35). Con esta identificación, las variables auxiliares x e y resultan

$$x = \frac{\omega_1 + \omega_2}{2} t, \quad y = \frac{\omega_1 - \omega_2}{2} t.\tag{4.37}$$

Así se tiene

$$\cos(\omega_1 t) + \cos(\omega_2 t) = 2 \cos\left(\frac{\omega_1 - \omega_2}{2} t\right) \cos\left(\frac{\omega_1 + \omega_2}{2} t\right),$$

y

$$\cos(\omega_1 t) - \cos(\omega_2 t) = -2 \sin\left(\frac{\omega_1 - \omega_2}{2} t\right) \sin\left(\frac{\omega_1 + \omega_2}{2} t\right).$$

De esta manera, el movimiento de ambas partes móviles se puede escribir como producto de dos oscilaciones

$$\begin{aligned}\psi_1(t) &= 2A \cos\left(\frac{\omega_2 - \omega_1}{2} t\right) \cos\left(\frac{\omega_1 + \omega_2}{2} t\right), \\ \psi_2(t) &= 2A \sin\left(\frac{\omega_2 - \omega_1}{2} t\right) \sin\left(\frac{\omega_1 + \omega_2}{2} t\right),\end{aligned}\tag{4.38}$$

una a la frecuencia promedio

$$\omega_p = \frac{\omega_1 + \omega_2}{2},$$

y otra a la frecuencia semidiferencia

$$\omega_m = \frac{\omega_1 - \omega_2}{2}.$$

En el caso considerado en la figura 2.6, las frecuencias propias correspondientes a los parámetros, $m_1 = m_2 = 1$ kg, $k = 40$ N/m, $k' = 120$ N/m, difieren significativamente, por lo que ambas oscilaciones tienen períodos comparables. En consecuencia, la estructura de producto no se manifiesta de manera evidente en la forma de las curvas.

Sin embargo, la situación cambia cualitativamente si las frecuencias propias son cercanas entre sí. En ese caso, la cantidad ω_m es mucho menor que ω_p , y las dos oscilaciones ocurren en escalas temporales muy diferentes. La oscilación de mayor frecuencia (ω_p , con periodo $T_p = 2\pi/\omega_p$) actúa como *señal portadora*, cuya amplitud es modulada por la oscilación de baja frecuencia ω_m ; por esta razón, a esta última se la denomina *señal moduladora*. En estas condiciones, la expresión de $\psi_1(t)$ en la ecuación (4.38) puede interpretarse como un coseno de frecuencia ω_p , con amplitud instantánea dada por la función

$$A_{1m}(t) = 2A \cos \omega_m t,\tag{4.39}$$

que actúa como envolvente de la oscilación rápida. Análogamente, podemos interpretar la expresión de $\psi_2(t)$ en la misma ecuación como un seno de frecuencia ω_p , con amplitud instantánea

$$A_{1m}(t) = 2A \sin \omega_m t,\tag{4.40}$$

Este tipo de amplitudes moduladoras reciben el nombre de *amplitud de modulación* y varían en el tiempo mucho más lentamente que la función que modulan.

A pesar de esta interpretación, que sugiere oscilaciones armónicas con amplitud variable, las funciones $\psi_1(t)$ y $\psi_2(t)$ no son estrictamente armónicas, ya que su amplitud

no es constante en el tiempo. Sin embargo, cuanto más cercanas son las frecuencias ω_1 y ω_2 , más lenta es la modulación de estas funciones.

En el límite $\omega_m \rightarrow 0$, la función $\cos(\omega_m t)$ tiende a la unidad, por lo que $\psi_1(t)$ se reduce a una oscilación armónica pura de frecuencia ω_p . En cambio, la función $\sin(\omega_m t)$ tiende a cero, de modo que $\psi_2(t)$ se anula en este límite. Físicamente, esto refleja que cuando las frecuencias propias coinciden, desaparece el fenómeno de batidos y el sistema oscila en un único modo normal, sin intercambio de energía entre las partes móviles. Esto muestra que el fenómeno de batidos —y la consiguiente transferencia de energía entre las coordenadas— requiere la presencia de dos frecuencias distintas: cuando ambas coinciden, la dinámica se reduce a una oscilación modal pura.

En definitiva, este fenómeno —conocido como modulación de amplitud o batido— surge de manera natural como consecuencia de la cercanía entre las frecuencias propias del sistema y refleja la interferencia entre modos normales con escalas temporales bien diferenciadas.

4.7.2. Batidos en un ejemplo: simulaciones

A la luz de esta discusión, resulta claro por qué no se observa modulación de amplitud en el caso considerado en la figura 2.6a, con $k' = 120 \text{ N/m}$: en ese régimen, las frecuencias propias difieren significativamente, por lo que ambas oscilaciones tienen períodos comparables y la estructura de producto no se manifiesta de manera evidente en las curvas.

En cambio, cuando las frecuencias propias son cercanas, la modulación se vuelve claramente visible. Este comportamiento se ilustra en la figura 4.2. En cada columna se muestra la evolución de $\psi_1(t)$, la envolvente $\pm \cos(\omega_m t)$, la oscilación rápida $\cos(\omega_p t)$ y la evolución de $\psi_2(t)$.

El sistema difiere del considerado en la figura 2.6 solo en el valor de k' , la constante elástica del resorte central. Cambiando este valor es posible sintonizar la diferencia entre ω_1 y ω_2 . Es fácil ver que cuando $k'/k \ll 1$, $\omega_2 \approx \omega_1(+k/k')$. Para estas curvas se usó la condición inicial $\vec{\psi}(t) = (1, 0)$, que corresponde a $A = 0.5$.

En la columna izquierda ($k' = 3 \text{ N/m}$), las frecuencias propias son cercanas entre sí, lo que da lugar a una frecuencia de modulación ω_m pequeña. En consecuencia, el período de la envolvente es mucho mayor que el de la oscilación rápida, y los batidos se observan como una modulación lenta y bien definida.

En la columna derecha ($k' = 10 \text{ N/m}$), la separación entre las frecuencias propias es mayor, lo que incrementa la frecuencia de modulación. Como resultado, el período de la envolvente disminuye y la modulación resulta menos evidente.

Estas diferencias muestran cómo el valor del acoplamiento controla la escala temporal de los batidos a través de la separación entre las frecuencias propias.

Si bien la variación temporal de las perturbaciones depende de las condiciones iniciales, la modulación de amplitud persiste siempre que las frecuencias propias sean

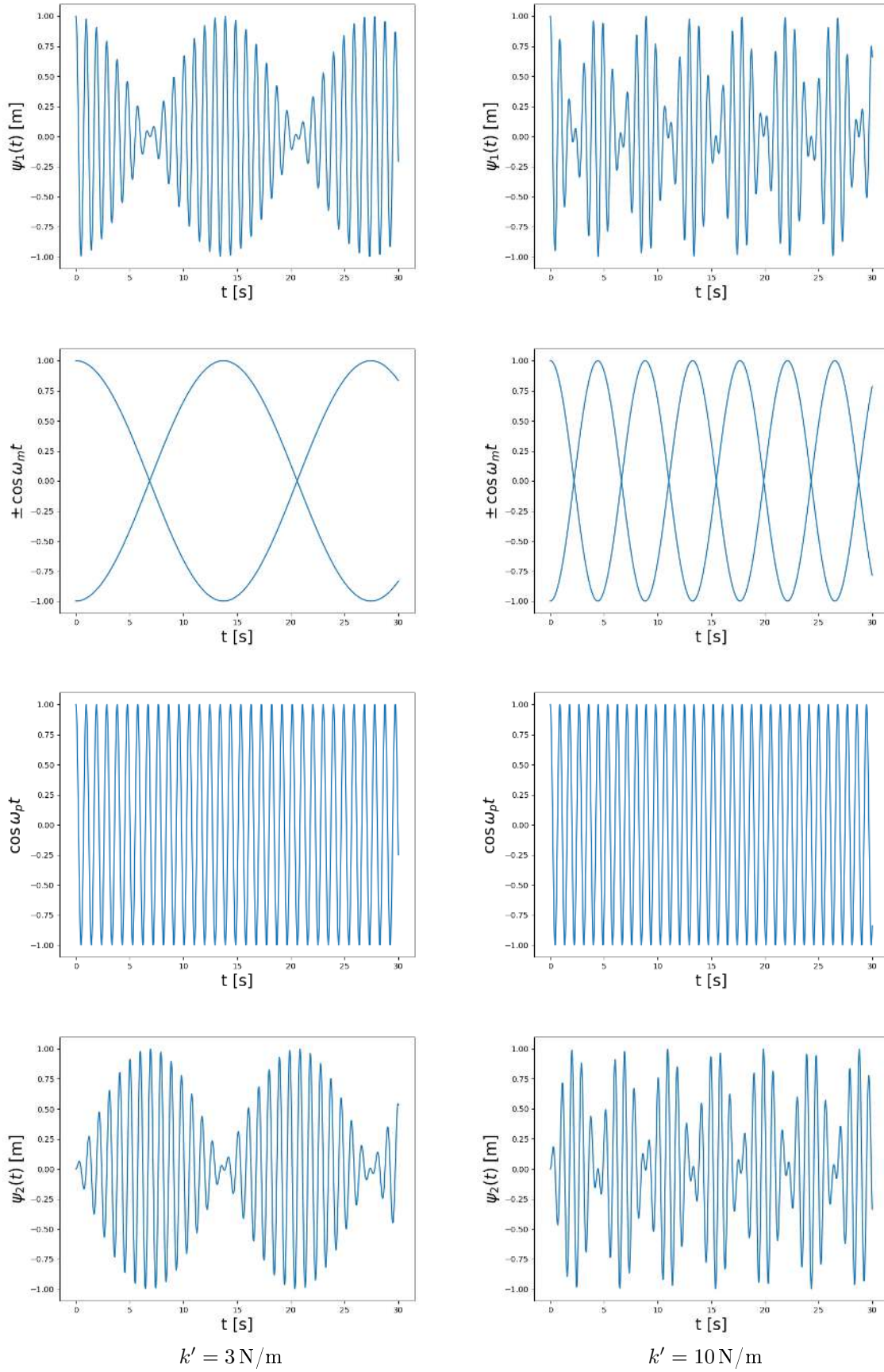


Figura 4.2: Batidos en un sistema simétrico de dos masas acopladas. De arriba abajo: $\psi_1(t)$, $\pm \cos \omega_m t$, $\cos \omega_p t$, $\psi_2(t)$. Izquierda: $k' = 3 \text{ N/m}$, derecha: $k' = 10 \text{ N/m}$.

cercanas. La igualdad de amplitudes y fases sólo da lugar a una expresión particularmente simple, pero no constituye una condición necesaria para la aparición de batidos. Esta persistencia puede observarse en la figura 4.3, aun cuando las condiciones iniciales no conducen a amplitudes y fases iguales en la descomposición modal.

En la columna izquierda, correspondiente a amplitudes modales distintas pero fases iniciales iguales, el patrón de batido se mantiene claramente reconocible, aunque la modulación ya no oscila alrededor de cero. En la columna derecha, donde además intervienen fases distintas, la señal presenta una estructura más compleja; sin embargo, sigue siendo posible identificar una modulación de la amplitud asociada a la cercanía entre las frecuencias propias.

Estos resultados muestran que el fenómeno de batido no depende de una elección particular de condiciones iniciales, sino que es una consecuencia robusta de la estructura espectral del sistema.

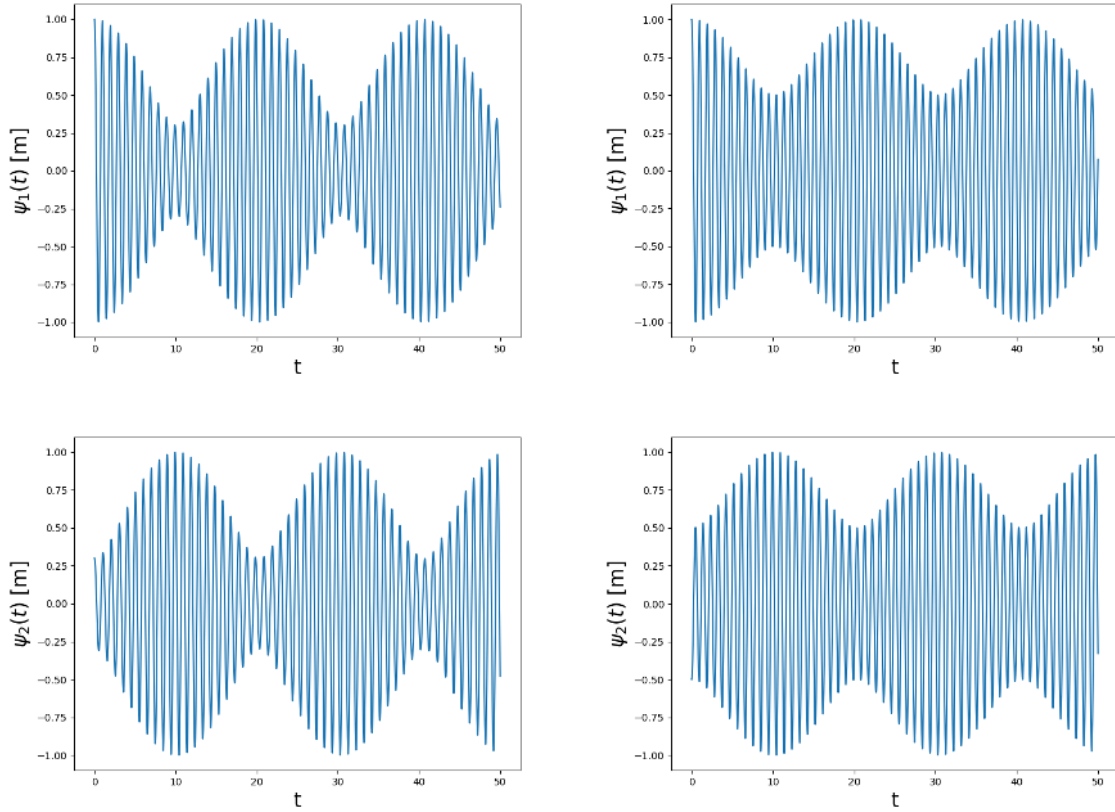


Figura 4.3: Efecto de las condiciones iniciales sobre la modulación de amplitud en un sistema simétrico como el de la figura anterior, pero con $k' = 2 \text{ N/m}$. En cada columna se muestran $\psi_1(t)$ (fila superior) y $\psi_2(t)$ (fila inferior), en la misma escala temporal. Columna izquierda: $(\psi_1(0), \psi_2(0)) = (1, 0.3)$, velocidades iniciales nulas. Columna derecha: $(\psi_1(0), \psi_2(0)) = (1, -0.5)$, $(\dot{\psi}_1(0), \dot{\psi}_2(0)) = (0.3, -0.2)$.

La inclusión de ambas coordenadas en la figura permite además visualizar otro aspecto fundamental de la dinámica: la transferencia de energía entre las partes móviles. En el caso simétrico con condiciones iniciales $(1, 0)$ y velocidades nulas, esta transferencia se manifiesta de manera particularmente clara al comparar las perturbaciones

$\psi_1(t)$ en la fila superior de la figura 4.2 con las perturbaciones $\psi_2(t)$ de la fila inferior: cuando una de las partes móviles alcanza un máximo de amplitud, la otra se encuentra próxima al reposo.

En los casos mostrados en la figura 4.3, la transferencia de energía sigue presente, aunque su manifestación es menos regular. La energía continúa intercambiándose entre ambas coordenadas, pero ya no de manera completamente simétrica, reflejando la diferencia en las amplitudes modales: la transferencia de energía no sigue un patrón simple, aunque sigue siendo posible reconocer intervalos en los que una coordenada domina sobre la otra.

4.7.3. Batidos y detectores cuadráticos

Un *detector cuadrático* es un dispositivo que mide una magnitud proporcional al cuadrado de la señal. En el contexto mecánico, esto corresponde naturalmente a la energía, que depende cuadráticamente de las coordenadas y sus velocidades. Este tipo de respuesta no es exclusiva de los sistemas mecánicos. En muchos contextos físicos y fisiológicos, las magnitudes observables están relacionadas con el cuadrado de una amplitud. En el sistema visual humano, la intensidad luminosa percibida está asociada al valor promedio en el tiempo del cuadrado del campo eléctrico de la onda electromagnética, mientras que en el sistema auditivo la percepción de la intensidad del sonido está vinculada a la energía transportada por la onda. En este último caso, la respuesta del sistema depende de la escala temporal de la señal: las variaciones rápidas no son resueltas, mientras que las variaciones más lentas sí pueden ser percibidas. En este sentido, ambos sistemas actúan, al menos de manera aproximada, como detectores cuadráticos, aunque con diferencias en su respuesta temporal.

Consideremos, por ejemplo, la expresión

$$\psi_1(t) = 2A \cos(\omega_m t) \cos(\omega_p t),$$

su cuadrado viene dado por

$$\psi_1^2(t) = 4A^2 \cos^2(\omega_m t) \cos^2(\omega_p t).$$

Dado que

$$\cos^2(\omega_m t) = \frac{1}{2}(1 + \cos(2\omega_m t)),$$

entonces

$$\psi_1^2(t) = 2A^2(1 + \cos(2\omega_m t)) \cos^2(\omega_p t).$$

La señal cuadrática contiene una modulación lenta con frecuencia $2\omega_m = |\omega_1 - \omega_2|$, llamada *frecuencia de pulsación*, que se manifiesta como una variación temporal de la intensidad de la oscilación rápida. Esta es la magnitud que resulta directamente accesible en sistemas cuya respuesta depende del cuadrado de la señal.

Desde el punto de vista físico, esto refleja que la energía del sistema no distingue el signo de la amplitud, por lo que los máximos de energía se producen tanto en los máximos (positivos) como en los mínimos (negativos) de la envolvente.

En sistemas como el auditivo, que responden a la intensidad de la señal, esta modulación se percibe como una variación periódica en el volumen del sonido, superpuesta a una señal de frecuencia aproximadamente igual a ω_p . Los músicos utilizan este efecto para afinar sus instrumentos, ajustando las frecuencias hasta que los batidos desaparecen, lo que indica que ambas coinciden.

Por ejemplo, si se superponen señales de 440 Hz y 439 Hz, se percibe una pulsación de 1 Hz, es decir, una variación en la intensidad una vez por segundo. Este principio se utiliza en la práctica musical para lograr afinaciones precisas; en particular, las orquestas suelen adoptar como referencia la nota La₄ (A4) afinada a 440 Hz, de acuerdo con el estándar internacional. Cuando la frecuencia de pulsación supera la capacidad de seguimiento temporal del sistema auditivo (del orden de 15–20 Hz), la modulación deja de percibirse como una variación de intensidad y se manifiesta como la coexistencia de dos frecuencias cercanas, dando lugar a una sensación de disonancia.

En contraste con el sistema auditivo, el sistema visual no puede seguir las oscilaciones rápidas de la señal en el rango óptico (con frecuencias del orden de 10^{14} – 10^{15} Hz) y responde al valor promedio de la intensidad en el tiempo. De este modo, actúa, en primera aproximación, como un detector cuadrático seguido de un promediado temporal de la señal.

4.8. La perturbación que se propaga

Un resultado físico muy interesante obtenido en el estudio numérico ilustrado en las figuras 4.2 y 4.3 es el siguiente. En ambos casos solo perturbamos inicialmente la parte móvil de la izquierda, pero no la de la derecha. Sin embargo, a medida que pasa el tiempo la perturbación se va transmitiendo a la parte móvil de la derecha, que comienza a moverse a partir del equilibrio, mientras que la parte móvil de la izquierda disminuye la amplitud de su movimiento. ¿No es esto justamente *una perturbación que se propaga*?

Vemos que la energía potencial que se entregó inicialmente a la parte móvil de la izquierda se va transfiriendo a la parte móvil de la derecha. Desde este punto de vista, el fenómeno de batidos puede interpretarse precisamente como una manifestación particular de este intercambio de energía en un sistema con un número finito de grados de libertad.

Podríamos especular que, si el último resorte estuviera unido a una tercera parte móvil (en lugar de estar fijado a una pared rígida), esta transferencia de energía continuaría alejándose de la parte móvil que le dio origen. Pero, como está la pared, es como si la perturbación se reflejara y regresara a la parte móvil de la izquierda.

Lo interesante de esta interpretación es que permite introducir de manera natural el concepto de onda que mencionamos en la sección 1.2, como *una perturbación que se propaga*. Es cierto que en un sistema tan pequeño y limitado como el que estamos considerando esta propagación es restringida, pero pronto veremos que en sistemas más extensos el movimiento producido por la perturbación en un extremo puede propagarse

hacia otras partes móviles más lejanas, hasta alcanzar el límite del sistema, donde necesariamente se refleja y se propaga en sentido contrario. Esta perspectiva resulta particularmente útil como punto de partida para el estudio de sistemas con muchos grados de libertad, donde la noción de propagación de perturbaciones adquiere un carácter continuo y conduce naturalmente al concepto de onda.

Capítulo 5

Disipación y forzado

Hasta aquí hemos considerado sistemas ideales, en los cuales la dinámica está gobernada exclusivamente por fuerzas conservativas lineales. En ese marco, la existencia de modos normales se vincula directamente con la posibilidad de diagonalizar el problema: la diagonalización reduce el sistema a un conjunto de grados de libertad independientes, cada uno de los cuales oscila con una frecuencia bien definida.

Esta estructura se modifica al incorporar efectos más realistas. Por un lado, la presencia de términos disipativos introduce una dinámica no conservativa: la energía del sistema deja de ser constante y el movimiento tiende a extinguirse con el tiempo. Por otro lado, la inclusión de forzados externos rompe la autonomía del sistema. Esto permite la inyección continua de energía y da lugar a fenómenos como la resonancia.

Desde el punto de vista formal, estos efectos se incorporan mediante la aparición de nuevos términos en las ecuaciones de movimiento. Por un lado, aparece un término dependiente de las velocidades de las partes móviles, asociado a los mecanismos de disipación, que se manifiestan únicamente en presencia de movimiento. Por otro lado, aparece un término externo a la dinámica del sistema, que introduce una inhomogeneidad en las ecuaciones.

En el caso en que las fuerzas disipativas sean lineales en las velocidades, las ecuaciones de movimiento pueden escribirse en la forma

$$\mathbb{M}\ddot{\vec{\psi}} + \mathbb{C}\dot{\vec{\psi}} + \mathbb{K}\vec{\psi} = \vec{f}(t), \quad (5.1)$$

donde $\mathbb{C}$ representa la contribución disipativa y $\vec{f}(t)$ las fuerzas externas aplicadas al sistema.

La presencia de estos términos adicionales altera de manera esencial la estructura del problema. En general, ya no es posible encontrar una transformación que desacople completamente la dinámica en modos independientes.

Dicho de otra manera, la transformación que permite eliminar los términos cruzados en las contribuciones asociadas a $\mathbb{M}$ y $\mathbb{K}$ no diagonaliza, en general, la matriz $\mathbb{C}$. Como consecuencia, las ecuaciones de movimiento pueden seguir acopladas aun en coordenadas modales.

El tratamiento del caso general requiere herramientas adicionales y no será considerado aquí. Nos restringiremos, por lo tanto, a situaciones en las que la matriz de disipación puede diagonalizarse simultáneamente con $\mathbb{M}$ y $\mathbb{K}$, lo que permite extender de manera directa el análisis en términos de modos normales.

En este marco, el estudio de sistemas con disipación y forzado puede entenderse como una extensión del problema de modos normales. La dinámica resulta del equilibrio entre la pérdida y la inyección de energía. Comenzaremos analizando los efectos de la disipación, luego continuaremos con la acción de fuerzas externas.

5.1. Disipación

Para investigar la incorporación de efectos disipativos en sistemas lineales, consideraremos que las fuerzas de disipación son proporcionales a las velocidades, de modo que pueden representarse mediante una matriz $\mathbb{C}$ constante.

Tales fuerzas disipativas aparecen en numerosas situaciones físicas. Un cuerpo que se desplaza en un fluido, como el aire o el agua, experimenta una fuerza de resistencia que se opone al movimiento. Para velocidades moderadas, esta fuerza puede modelarse como proporcional a la velocidad, $\vec{F}_d = -\gamma \vec{v}$, donde $\gamma > 0$ es un coeficiente que depende de las propiedades del medio y de la geometría del cuerpo. Esta fuerza da lugar a una aceleración de frenado proporcional a la velocidad, cuyo efecto es disipar la energía mecánica del sistema. Este modelo corresponde al llamado *amortiguamiento viscoso*, ampliamente utilizado en la descripción de sistemas lineales.

En sistemas con varios grados de libertad, este tipo de interacción se generaliza mediante términos lineales en las velocidades, que pueden escribirse en forma matricial a través de la matriz $\mathbb{C}$ introducida anteriormente. Como se discutió en la introducción, la presencia de esta matriz puede impedir el desacople completo de las ecuaciones de movimiento. En lo que sigue nos limitaremos a situaciones en las que $\mathbb{C}$ puede diagonalizarse simultáneamente con $\mathbb{M}$ y $\mathbb{K}$, lo que permite extender el análisis modal desarrollado para sistemas conservativos.

Bajo estas hipótesis, al introducir el cambio de coordenadas

$$\vec{\psi} = \mathbb{D} \vec{\Phi},$$

las ecuaciones de movimiento se desacoplan en la forma

$$\ddot{\Phi}_i + \gamma_i \dot{\Phi}_i + \omega_i^2 \Phi_i = 0,$$

donde ω_i son las frecuencias propias del sistema conservativo y γ_i caracteriza el amortiguamiento de cada modo y generaliza, en este contexto, el coeficiente de frenado introducido en el caso de un grado de libertad. Observar que cada coordenada modal evoluciona de manera independiente como un oscilador armónico amortiguado. En particular, distintos modos pueden presentar diferentes escalas de decaimiento temporal.

5.1.1. Mecanismos de disipación: interacción con el medio y pérdidas internas

Para ilustrar la estructura de la matriz de disipación, consideremos nuevamente el sistema de dos masas y tres resortes. Supongamos que existen mecanismos disipativos de dos tipos: i) fuerzas proporcionales a la velocidad de cada masa, que modelan la interacción con el medio, y ii) fuerzas proporcionales a la velocidad relativa entre los extremos de los resortes, que modelan la disipación interna del sistema.

Al incorporar las fuerzas disipativas, la única diferencia con respecto al sistema (2.6) radica en la aparición de términos proporcionales a las velocidades. En consecuencia, las ecuaciones de movimiento del sistema toman la forma

$$\begin{aligned} m_1 \ddot{\psi}_1 &= -k_1 \psi_1 - k_2(\psi_1 - \psi_2) - c_1 \dot{\psi}_1 - c_2(\dot{\psi}_1 - \dot{\psi}_2), \\ m_2 \ddot{\psi}_2 &= -k_3 \psi_2 - k_2(\psi_2 - \psi_1) - c_3 \dot{\psi}_2 - c_2(\dot{\psi}_2 - \dot{\psi}_1) \end{aligned} \quad (5.2)$$

Reordenando los términos, se obtiene la ecuación en forma matricial

$$\mathbb{M} \ddot{\vec{\psi}} + \mathbb{C} \dot{\vec{\psi}} + \mathbb{K} \vec{\psi} = 0,$$

con

$$\mathbb{C} = \begin{bmatrix} c_1 + c_2 & -c_2 \\ -c_2 & c_2 + c_3 \end{bmatrix}.$$

Conviene destacar que los coeficientes de $\mathbb{C}$ no corresponden necesariamente a mecanismos disipativos únicos. Por ejemplo, el término c_1 puede recibir contribuciones tanto de la interacción de la masa m_1 con el medio como de pérdidas internas en el resorte que la conecta con la pared.

Por su parte, el coeficiente c_2 describe la disipación asociada a la deformación relativa del resorte entre las masas. En un resorte real, esta disipación tiene su origen en procesos irreversibles a nivel microscópico, y puede modelarse, en primera aproximación, mediante un término proporcional a la velocidad relativa entre sus extremos.

En el marco del modelo lineal adoptado, estos distintos mecanismos conducen a términos de la misma forma, por lo que sus contribuciones resultan indistinguibles a nivel macroscópico. En este sentido, la matriz $\mathbb{C}$ debe entenderse como una *descripción efectiva* de los procesos disipativos presentes en el sistema.

5.1.2. Estructura de $\mathbb{C}$

En general, no existe razón para que la matriz de disipación $\mathbb{C}$ se diagonalice junto con $\mathbb{M}$ y $\mathbb{K}$. Mientras que las fuerzas elásticas derivan de un potencial y, por lo tanto, están sujetas a restricciones de simetría bien definidas, las fuerzas disipativas no admiten, en general, una descripción análoga. Como consecuencia, las matrices $\mathbb{K}$ y $\mathbb{C}$ pueden presentar estructuras distintas.

Por ejemplo, dos resortes pueden tener la misma constante elástica pero diferir en sus mecanismos de disipación. En tal caso, las simetrías de $\mathbb{K}$ y $\mathbb{C}$ no coinciden, y la transformación que diagonaliza a $\mathbb{M}$ y $\mathbb{K}$ no necesariamente diagonaliza a $\mathbb{C}$.

Si las fuerzas disipativas son lineales en las velocidades, $\vec{F}_{\text{dis}} = -\mathbb{C}\dot{\vec{\psi}}$, la potencia entregada al sistema por estas fuerzas viene dada por

$$P = \vec{F}_{\text{dis}} \cdot \dot{\vec{\psi}} = -\dot{\vec{\psi}}^T \mathbb{C} \dot{\vec{\psi}}.$$

Dado que las fuerzas disipativas extraen energía del sistema, esta potencia es no positiva,

$$P \leq 0.$$

Resulta entonces conveniente definir la potencia disipada como la magnitud positiva

$$P_{\text{dis}} = -P = \dot{\vec{\psi}}^T \mathbb{C} \dot{\vec{\psi}} \geq 0.$$

Si se descompone la matriz $\mathbb{C}$ en su parte simétrica y antisimétrica,

$$\mathbb{C} = \frac{1}{2}(\mathbb{C} + \mathbb{C}^T) + \frac{1}{2}(\mathbb{C} - \mathbb{C}^T) \equiv \mathbb{C}_s + \mathbb{C}_a,$$

se obtiene

$$\dot{\vec{\psi}}^T \mathbb{C} \dot{\vec{\psi}} = \dot{\vec{\psi}}^T \mathbb{C}_s \dot{\vec{\psi}} + \dot{\vec{\psi}}^T \mathbb{C}_a \dot{\vec{\psi}}.$$

El segundo término se anula idénticamente, ya que $\mathbb{C}_a$ es antisimétrica. Por lo tanto, la disipación de energía está determinada únicamente por la parte simétrica de la matriz $\mathbb{C}$,

$$P_{\text{dis}} = \dot{\vec{\psi}}^T \mathbb{C}_s \dot{\vec{\psi}}.$$

La condición de que $P_{\text{dis}} \geq 0$ para todo $\dot{\vec{\psi}}$ implica que $\mathbb{C}_s$ sea semidefinida positiva. Esta condición permite que existan movimientos no triviales para los cuales la disipación se anula, es decir, velocidades no nulas tales que $P_{\text{dis}} = 0$. En sistemas con varios grados de libertad, esta situación se presenta, por ejemplo, cuando los mecanismos disipativos actúan únicamente sobre velocidades relativas: en ese caso, los movimientos en los que todas las partes se desplazan en fase no disipan energía. En términos algebraicos, estos movimientos pertenecen al núcleo de la forma cuadrática definida por $\mathbb{C}_s$.

Por otra parte, la parte antisimétrica $\mathbb{C}_a$ no contribuye a la disipación de energía: si bien puede modificar la dinámica del sistema al introducir acoplamientos entre velocidades, no está asociada a pérdida de energía.

5.1.3. Condiciones de desacople

Como consecuencia de lo discutido anteriormente, las ecuaciones de movimiento pueden permanecer acopladas aun al expresarlas en coordenadas modales. En efecto, al

introducir el cambio de variables $\vec{\psi} = \mathbb{D}\vec{\Phi}$ y multiplicar la ecuación de movimiento por $\mathbb{D}^T$, se obtiene

$$\mathbb{D}^T \mathbb{M} \mathbb{D} \ddot{\vec{\Phi}} + \mathbb{D}^T \mathbb{C} \mathbb{D} \dot{\vec{\Phi}} + \mathbb{D}^T \mathbb{K} \mathbb{D} \vec{\Phi} = 0.$$

Dado que, por construcción, $\mathbb{D}^T \mathbb{M} \mathbb{D} = \mathbb{I}$ y $\mathbb{D}^T \mathbb{K} \mathbb{D}$ es diagonal, la ecuación adopta la forma

$$\ddot{\vec{\Phi}} + (\mathbb{D}^T \mathbb{C} \mathbb{D}) \dot{\vec{\Phi}} + \mathbf{\Omega}^2 \vec{\Phi} = 0,$$

donde

$$\mathbf{\Omega}^2 = \text{diag}(\omega_1^2, \omega_2^2, \dots).$$

En general, la matriz $\mathbb{D}^T \mathbb{C} \mathbb{D}$ no es diagonal. Denotando $\tilde{\mathbb{C}} = \mathbb{D}^T \mathbb{C} \mathbb{D}$, la ecuación anterior puede escribirse como

$$\ddot{\vec{\Phi}} + \tilde{\mathbb{C}} \dot{\vec{\Phi}} + \mathbf{\Omega}^2 \vec{\Phi} = 0.$$

En componentes, resulta

$$\ddot{\Phi}_i + \sum_j \tilde{C}_{ij} \dot{\Phi}_j + \omega_i^2 \Phi_i = 0,$$

lo que muestra que, en general, las ecuaciones permanecen acopladas a través de las velocidades modales. A diferencia del caso conservativo, el acoplamiento no se manifiesta en los desplazamientos, sino en las velocidades.

El desacople completo se recupera únicamente cuando la matriz $\tilde{\mathbb{C}}$ es diagonal, es decir, cuando la transformación $\mathbb{D}$ diagonaliza simultáneamente a $\mathbb{M}$, $\mathbb{K}$ y $\mathbb{C}$.

Una clase particularmente importante de matrices de disipación que cumple esta condición es aquella en la cual $\mathbb{C}$ puede expresarse como combinación lineal de $\mathbb{M}$ y $\mathbb{K}$, es decir,

$$\mathbb{C} = \alpha \mathbb{M} + \beta \mathbb{K}, \quad (5.3)$$

con α y β constantes. En este caso, la matriz $\mathbb{C}$ es diagonalizada por la misma transformación que diagonaliza a $\mathbb{M}$ y $\mathbb{K}$, y el desacople modal se conserva.

En efecto,

$$\mathbb{D}^T \mathbb{C} \mathbb{D} = \alpha \mathbb{D}^T \mathbb{M} \mathbb{D} + \beta \mathbb{D}^T \mathbb{K} \mathbb{D}.$$

Dado que ambas matrices del lado derecho son diagonales, también lo es $\mathbb{D}^T \mathbb{C} \mathbb{D}$.

En consecuencia, las ecuaciones de movimiento se desacoplan en la forma

$$\ddot{\Phi}_i + (\alpha + \beta \omega_i^2) \dot{\Phi}_i + \omega_i^2 \Phi_i = 0,$$

donde ω_i son las frecuencias propias del sistema conservativo.

Comparando con la forma estándar del oscilador armónico amortiguado, se identifica

$$\gamma_i = \alpha + \beta \omega_i^2,$$

lo que muestra que cada modo posee, en general, un coeficiente de amortiguamiento distinto, determinado tanto por los parámetros de disipación como por su frecuencia

propia. En particular, cuando $\beta \neq 0$, los modos de mayor frecuencia resultan más fuertemente amortiguados, un hecho que tendrá consecuencias directas en la respuesta del sistema ante excitaciones externas.

Este tipo de disipación se conoce como *amortiguamiento proporcional*.

5.1.4. Regímenes de amortiguamiento

Bajo las condiciones discutidas anteriormente, cada coordenada modal satisface una ecuación del tipo

$$\ddot{\Phi}_i + \gamma_i \dot{\Phi}_i + \omega_i^2 \Phi_i = 0,$$

formalmente idéntica a la del oscilador armónico amortiguado con un grado de libertad.

En consecuencia, para cada modo se distinguen los mismos regímenes que en dicho caso, determinados por la relación entre γ_i y ω_i :

- **Subamortiguado** ($\gamma_i/2 < \omega_i$): el movimiento es oscilatorio, con frecuencia

$$\omega'_i = \sqrt{\omega_i^2 - \frac{\gamma_i^2}{4}},$$

y con una amplitud que decrece exponencialmente en el tiempo en la forma

$$e^{-\gamma_i t/2}.$$

El tiempo característico de decaimiento es $\tau = 2/\gamma_i$.

- **Amortiguamiento crítico** ($\gamma_i/2 = \omega_i$): el sistema retorna al equilibrio sin oscilar, en el menor tiempo posible. La solución tiene la forma

$$\Phi_i(t) = (A + Bt) e^{-\gamma_i t/2}.$$

- **Sobreamortiguado** ($\gamma_i/2 > \omega_i$): el sistema retorna al equilibrio sin oscilaciones, como superposición de dos exponentes reales decrecientes. La solución tiene la forma

$$\Phi_i(t) = A e^{r_+ t} + B e^{r_- t},$$

donde

$$r_{\pm} = -\frac{\gamma_i}{2} \pm \sqrt{\frac{\gamma_i^2}{4} - \omega_i^2}.$$

En este caso aparecen dos escalas temporales características, asociadas a los dos exponentes, lo que da lugar a una relajación más lenta que en el caso crítico.

A diferencia del caso de un solo grado de libertad, en sistemas con varios grados de libertad cada modo puede encontrarse en un régimen distinto. Como consecuencia, la dinámica global del sistema puede presentar simultáneamente componentes oscilatorias y no oscilatorias, asociadas a diferentes modos. En particular, los modos menos amortiguados tienden a dominar la dinámica en tiempos largos.

Para ilustrar los distintos regímenes de amortiguamiento en un sistema con varios grados de libertad, retomamos el modelo de dos masas y tres resortes introducido anteriormente. Para facilitar la interpretación de los resultados, en las ecuaciones (5.2) consideraremos el caso $m_1 = m_2 = m$ y $k_1 = k_2 = k_3 = k$.

Interacción con el medio Como primer ejemplo analizaremos el caso en que $c_2 = 0$, de modo que la matriz de disipación se reduce a

$$\mathbb{C} = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_1 \end{bmatrix} = \alpha_1 \mathbb{M},$$

donde $c_1 = \alpha_1 m$.

En estas condiciones, y suponiendo despreciables las pérdidas internas en los resortes, las fuerzas disipativas representan únicamente la interacción de cada masa con el medio. En particular, no aparecen términos asociados a velocidades relativas, por lo que no se introducen acoplamientos disipativos entre los grados de libertad.

Se trata de un caso particular del amortiguamiento proporcional considerado en la ec. (5.3), en el cual las matrices $\mathbb{M}$, $\mathbb{C}$ y $\mathbb{K}$ pueden diagonalizarse simultáneamente. Como consecuencia, cada coordenada modal satisface una ecuación equivalente a la del oscilador armónico amortiguado con un grado de libertad, lo que permite identificar directamente los distintos regímenes de amortiguamiento.

En la figura 5.1 se muestra la evolución temporal del sistema para distintos valores del coeficiente α_1 , manteniendo $\alpha_2 = 0$. En ausencia de disipación (caso 1), el movimiento resulta de la superposición de los dos modos normales del sistema, lo que da lugar a un patrón de interferencia entre las coordenadas.

Para valores pequeños de α_1 (caso 2), ambos modos siguen contribuyendo de manera comparable, aunque su amplitud decae progresivamente en el tiempo. En este caso, el coeficiente de amortiguamiento modal es $\gamma = \alpha_1$, que resulta menor que las frecuencias propias $\omega_1 = 1$ y $\omega_2 = \sqrt{3}$, por lo que ambos modos se encuentran en el régimen subamortiguado.

Al aumentar la disipación (caso 3), las oscilaciones se atenúan más rápidamente y el número de ciclos observables disminuye. En este régimen, ambos modos continúan siendo subamortiguados, aunque con una frecuencia efectiva reducida y un decaimiento más rápido. Para valores mayores, como en el caso 4, se tiene $\gamma/2 > \omega_1$ y $\gamma/2 > \omega_2$, por lo que ambos modos se encuentran en el régimen sobreamortiguado y las oscilaciones desaparecen por completo.

Es importante notar que, en todos los casos, la estructura modal del movimiento se conserva: ambos modos experimentan el mismo amortiguamiento y, en consecuencia, la disipación no favorece selectivamente a ninguno de ellos. La evolución temporal resulta así de una atenuación uniforme de la dinámica que ya estaba presente en el sistema conservativo. Esto reproduce, para cada modo, la clasificación habitual del oscilador armónico amortiguado.

Es importante notar que, en todos los casos, la estructura modal del movimiento

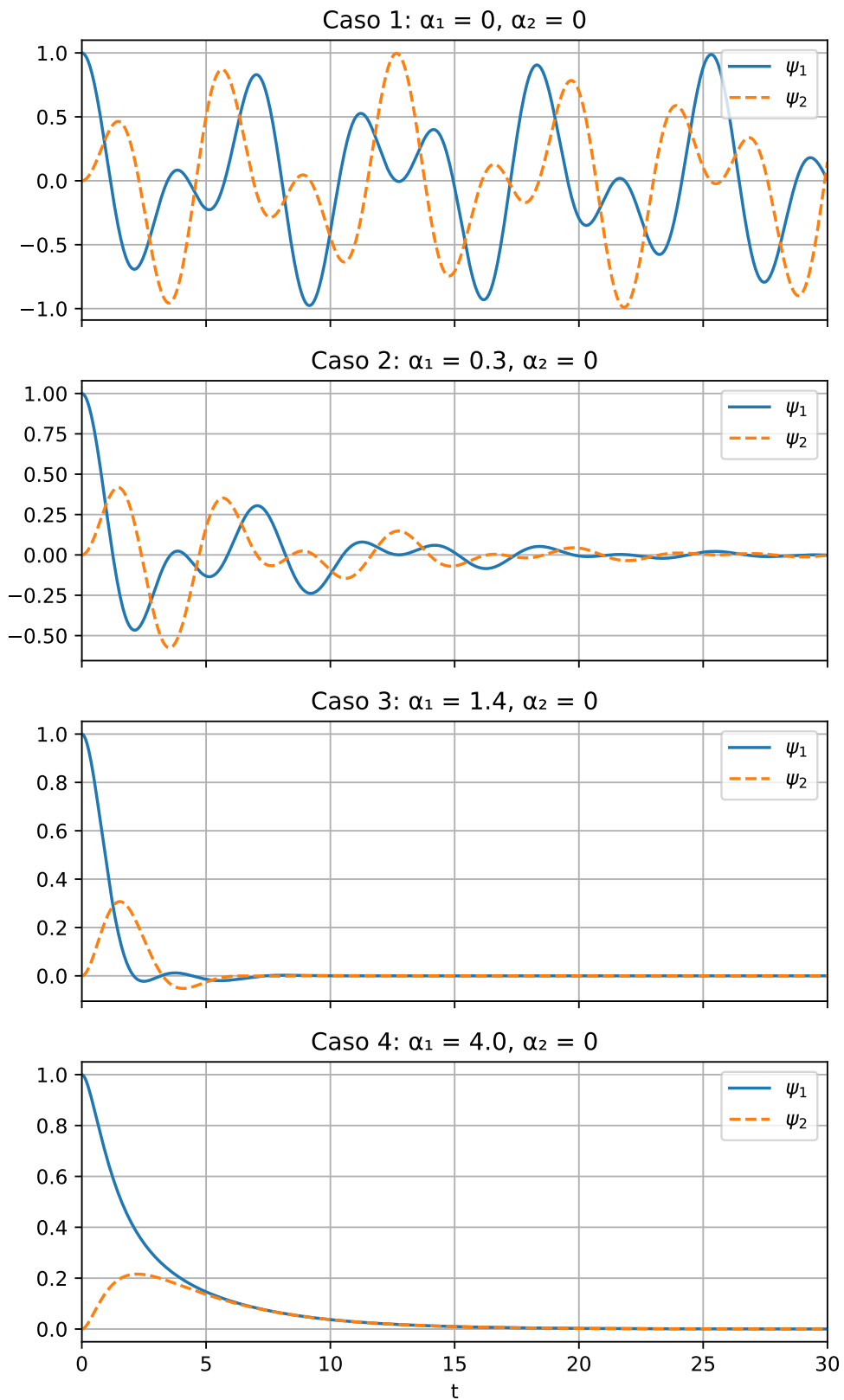


Figura 5.1: Evolución temporal de las coordenadas $\psi_1(t)$ y $\psi_2(t)$ para distintos valores del coeficiente de disipación α_1 , con $\alpha_2 = 0$. A medida que α_1 aumenta desde cero, el sistema pasa de un régimen conservativo, con batidos característicos, a un régimen fuertemente amortiguado en el que las oscilaciones desaparecen.

se conserva: ambos modos experimentan el mismo amortiguamiento y, en consecuencia, la disipación no favorece selectivamente a ninguno de ellos. La evolución temporal resulta así de una atenuación uniforme de la dinámica presente en el sistema conservativo, reproduciendo para cada modo la clasificación habitual del oscilador armónico amortiguado.

Este comportamiento contrasta con el que se obtiene cuando se introduce disipación asociada a la velocidad relativa entre las masas. Como se verá a continuación, en ese caso la disipación no actúa de manera uniforme sobre los modos, sino que introduce un mecanismo de selección modal que modifica cualitativamente la dinámica del sistema.

Disipación interna y selección modal A diferencia del caso anterior, consideramos ahora la situación en la que $\alpha_2 \neq 0$, manteniendo fijo α_1 . En estas condiciones, definimos $c_1 = \alpha_1 m$ y $c_2 = \alpha_2 m$, de modo que la matriz de disipación puede escribirse como

$$\mathbb{C} = \begin{bmatrix} \alpha_1 + \alpha_2 & -\alpha_2 \\ -\alpha_2 & \alpha_1 + \alpha_2 \end{bmatrix} m.$$

La presencia de términos proporcionales a las velocidades relativas entre las masas introduce un mecanismo de disipación interna asociado a los elementos que acoplan ambos grados de libertad.

En estas condiciones, $\mathbb{C}$ y $\mathbb{K}$ presentan la misma estructura de acoplamiento. Como consecuencia, conmutan entre sí y comparten los mismos autovectores, lo que garantiza que pueden diagonalizarse simultáneamente mediante la misma transformación. Esta propiedad refleja el hecho de que los mecanismos disipativos considerados reproducen la misma estructura de interacción que las fuerzas elásticas.

Los coeficientes de amortiguamiento asociados a cada modo resultan, en general, distintos. En particular, se obtiene

$$\gamma_1 = \alpha_1, \quad \gamma_2 = \alpha_1 + 2\alpha_2,$$

lo que pone de manifiesto que la disipación no actúa de manera uniforme sobre los modos normales.

En la figura 5.2 se muestra la evolución temporal del sistema para distintos valores de α_2 , manteniendo fijo α_1 . En el caso $\alpha_2 = 0$ (caso 1), se recupera el comportamiento discutido previamente, en el que ambos modos experimentan el mismo amortiguamiento.

Al aumentar α_2 (casos 2 y 3), el modo asociado a la deformación relativa entre las masas resulta más fuertemente amortiguado, mientras que el modo en fase se ve afectado en menor medida. Como consecuencia, la contribución del primero se atenúa rápidamente y la dinámica queda dominada por el segundo.

Este efecto se vuelve más pronunciado a medida que aumenta α_2 (caso 4), donde el modo más amortiguado desaparece prácticamente en los primeros instantes y la dinámica efectiva queda gobernada por un único modo.

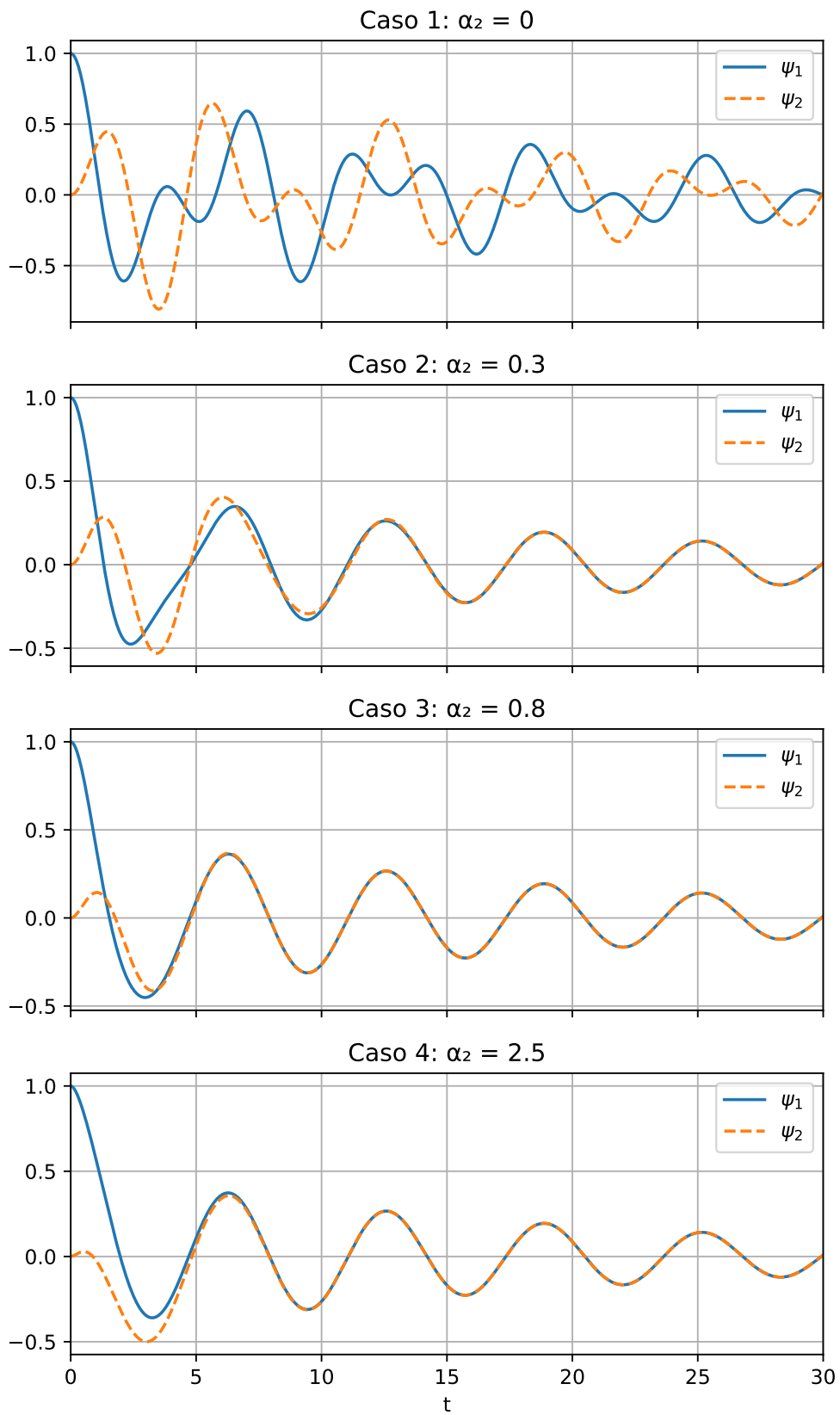


Figura 5.2: Evolución temporal de $\psi_1(t)$ y $\psi_2(t)$ para distintos valores del coeficiente de disipación interna α_2 , con $\alpha_1 = 0.1$ fijo. De arriba hacia abajo se observa que la disipación interna afecta de manera creciente al modo asociado a la deformación relativa entre las masas, dando lugar a una dinámica progresivamente dominada por el modo en fase.

A diferencia del caso anterior, la disipación no actúa de manera uniforme sobre los modos del sistema, sino que introduce un mecanismo de selección modal que favorece aquellos modos que disipan energía más lentamente.

Este mecanismo de selección modal reaparece en sistemas forzados, donde la interacción con fuerzas externas puede excitar preferentemente ciertos modos, en función de su frecuencia y de la estructura temporal de la excitación.

5.2. Fuerzas externas y respuesta del sistema

Consideramos ahora la acción de fuerzas externas dependientes del tiempo, introducidas previamente a través del término $\vec{f}(t)$ en las ecuaciones de movimiento 5.1. En este caso, la evolución del sistema está determinada no solo por las condiciones iniciales, sino también por la forma temporal de la excitación aplicada.

5.2.1. Proyección modal de las fuerzas externas

Para analizar la respuesta del sistema en presencia de fuerzas externas, resulta natural utilizar las coordenadas modales introducidas anteriormente. Consideremos, entonces, el cambio de variables

$$\vec{\psi} = \mathbb{D} \vec{\Phi},$$

donde $\mathbb{D}$ es la matriz que diagonaliza simultáneamente a $\mathbb{M}$, $\mathbb{C}$ y $\mathbb{K}$ en los casos considerados.

Sustituyendo en la ecuación de movimiento y multiplicando por $\mathbb{D}^T$, se obtiene

$$\mathbb{D}^T \mathbb{M} \mathbb{D} \ddot{\vec{\Phi}} + \mathbb{D}^T \mathbb{C} \mathbb{D} \dot{\vec{\Phi}} + \mathbb{D}^T \mathbb{K} \mathbb{D} \vec{\Phi} = \mathbb{D}^T \vec{f}(t).$$

Dado que, por construcción, $\mathbb{D}^T \mathbb{M} \mathbb{D} = \mathbb{I}$, $\mathbb{D}^T \mathbb{K} \mathbb{D} = \mathbf{\Omega}^2$ y, en los casos considerados, $\mathbb{D}^T \mathbb{C} \mathbb{D}$ es diagonal, la ecuación anterior se reduce a

$$\ddot{\vec{\Phi}} + \tilde{\mathbb{C}} \dot{\vec{\Phi}} + \mathbf{\Omega}^2 \vec{\Phi} = \tilde{\vec{f}}(t),$$

donde $\tilde{\vec{f}}(t) = \mathbb{D}^T \vec{f}(t)$ representa las fuerzas externas expresadas en coordenadas modales.

En estas condiciones, las ecuaciones de movimiento se desacoplan en un conjunto de ecuaciones independientes,

$$\ddot{\Phi}_i + \gamma_i \dot{\Phi}_i + \omega_i^2 \Phi_i = \tilde{f}_i(t),$$

cada una de las cuales corresponde a un oscilador armónico amortiguado sometido a una fuerza externa.

Este resultado muestra que el efecto de las fuerzas externas puede analizarse de manera independiente en cada modo normal. En particular, cada modo responde únicamente a la componente de la fuerza externa proyectada sobre su correspondiente dirección modal.

La respuesta del sistema resulta así de la superposición de las respuestas individuales de cada modo, ponderadas por dichas proyecciones. De este modo, la transformación modal no solo permite desacoplar la dinámica libre del sistema, sino también organizar la interacción con el exterior en términos de contribuciones modales independientes.

En lo que sigue, analizaremos la solución de estas ecuaciones para fuerzas externas armónicas. Esta elección no implica una pérdida de generalidad: una excitación temporal arbitraria puede descomponerse en una superposición de contribuciones armónicas mediante herramientas del análisis de Fourier, y, debido a la linealidad del sistema, la respuesta correspondiente puede obtenerse como superposición de las respuestas a cada una de ellas. Este análisis permitirá, en particular, identificar los mecanismos de resonancia del sistema.

5.2.2. Fuerzas externas armónicas

Para fijar ideas, consideremos en primer lugar la ecuación de un modo desacoplado sometido a una fuerza externa armónica,

$$\ddot{\Phi} + \gamma\dot{\Phi} + \omega_0^2\Phi = f_0 \cos(\omega t),$$

donde ω_0 es la frecuencia propia del modo y γ su coeficiente de amortiguamiento.

La solución general de esta ecuación puede escribirse como suma de dos contribuciones,

$$\Phi(t) = \Phi_{\text{trans}}(t) + \Phi_{\text{est}}(t),$$

donde $\Phi_{\text{trans}}(t)$ es la solución de la ecuación homogénea asociada y describe un régimen transitorio que decae en el tiempo, mientras que $\Phi_{\text{est}}(t)$ es una solución particular que persiste en el tiempo y define el régimen estacionario.

Dado que el régimen transitorio desaparece para tiempos suficientemente grandes, nos concentraremos en la solución estacionaria. Buscando una solución de la forma

$$\Phi_{\text{est}}(t) = A \cos(\omega t) + B \sin(\omega t),$$

se obtiene

$$A = \frac{f_0(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2}, \quad B = \frac{f_0 \gamma \omega}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2}.$$

La descomposición anterior tiene una interpretación física clara. La contribución proporcional a $\cos(\omega t)$ está en fase con la fuerza externa y describe una respuesta de tipo elástico: la fuerza realiza trabajo alternadamente positivo y negativo a lo largo del ciclo, sin transferencia neta de energía en promedio.

En cambio, la contribución proporcional a $\sin(\omega t)$ está desfasada en $\pi/2$ respecto de la excitación y da lugar a una transferencia neta de energía hacia el sistema. En efecto, la potencia instantánea $P = f(t)\dot{\Phi}(t)$ tiene un valor medio no nulo únicamente debido a esta componente, por lo que B está directamente asociado a la disipación de energía.

La amplitud total de la respuesta viene dada por

$$\mathcal{A}(\omega) = \sqrt{A^2 + B^2} = \frac{f_0}{\sqrt{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2}}.$$

Por otra parte, la componente en cuadratura determina la potencia media absorbida por el sistema, que resulta proporcional a

$$\langle P \rangle \propto \gamma\omega^2 \mathcal{A}^2(\omega),$$

y presenta un máximo en las cercanías de la frecuencia de resonancia.

La forma de la curva de resonancia está caracterizada por el coeficiente de amortiguamiento. En particular, es habitual introducir el factor de calidad

$$Q = \frac{\omega_0}{\gamma},$$

que mide la relación entre la energía almacenada en el sistema y la energía disipada por ciclo.

Valores grandes de Q corresponden a sistemas débilmente amortiguados, con picos de resonancia altos y estrechos, mientras que valores pequeños de Q describen sistemas fuertemente disipativos, con resonancias más anchas y menos pronunciadas.

En el régimen de amortiguamiento débil, la frecuencia de resonancia se aproxima a ω_0 , y el ancho del pico es del orden de $\Delta\omega \sim \gamma$.

Este resultado muestra que la respuesta del sistema depende de manera crítica de la relación entre la frecuencia de excitación y la frecuencia propia del modo. En particular, la amplitud presenta un máximo característico en una frecuencia próxima a ω_0 , cuyo valor y anchura dependen del amortiguamiento.

5.2.3. Respuesta modal de sistemas con varios grados de libertad

Retomamos ahora el sistema general. En coordenadas modales, cada grado de libertad satisface una ecuación del tipo

$$\ddot{\Phi}_i + \gamma_i \dot{\Phi}_i + \omega_i^2 \Phi_i = \tilde{f}_i(t).$$

Para una excitación armónica, cada componente modal de la fuerza puede escribirse como

$$\tilde{f}_i(t) = f_i \cos(\omega t),$$

donde f_i es la proyección de la fuerza externa sobre el modo i .

En consecuencia, la respuesta estacionaria de cada modo viene dada por una expresión análoga a la obtenida en el caso de un grado de libertad, con $\omega_0 \rightarrow \omega_i$ y $\gamma \rightarrow \gamma_i$. La respuesta total del sistema se obtiene como superposición de las contribuciones modales,

$$\vec{\psi}(t) = \sum_i \Phi_i(t) \vec{v}_i,$$

donde $\vec{v}_i$ son los vectores modales.

Este resultado muestra que la respuesta del sistema puede interpretarse como superposición de resonancias modales independientes, cada una caracterizada por su propia frecuencia y coeficiente de amortiguamiento.

5.2.4. Ejemplos: respuesta en frecuencia y selección modal

Para ilustrar estos resultados, retomamos el sistema de dos masas y tres resortes analizado anteriormente. Al incorporar una fuerza externa armónica en la parte móvil 1 y tomar $m_1 = m_2 = m$, $k_1 = k_2 = k_3 = k$, las ecuaciones (5.2) quedan

$$\begin{aligned} m\ddot{\psi}_1 &= -k\psi_1 - k(\psi_1 - \psi_2) - c_1\dot{\psi}_1 - c_2(\dot{\psi}_1 - \dot{\psi}_2) + f_0 \cos(\omega t), \\ m\ddot{\psi}_2 &= -k\psi_2 - k(\psi_2 - \psi_1) - c_1\dot{\psi}_2 - c_2(\dot{\psi}_2 - \dot{\psi}_1). \end{aligned} \quad (5.4)$$

Definiendo las coordenadas normales

$$\Phi_1 = \frac{\psi_1 + \psi_2}{\sqrt{2}}, \quad \Phi_2 = \frac{\psi_1 - \psi_2}{\sqrt{2}},$$

las ecuaciones de movimiento se desacoplan en

$$\ddot{\Phi}_i + \gamma_i \dot{\Phi}_i + \omega_i^2 \Phi_i = f_i \cos(\omega t),$$

donde f_i representa la proyección de la fuerza externa sobre el modo i . Dado que la fuerza actúa sobre una sola masa, su proyección sobre ambos modos es no nula, resultando $f_1 = f_2 = f_0/\sqrt{2}$.

En la figura 5.3 se muestra la respuesta en frecuencia del sistema para disipación uniforme ($\alpha_2 = 0$) y distintos valores de c_1 . Se observan dos resonancias asociadas a los modos normales, ubicadas en $\omega_1 = 1$ y $\omega_2 = \sqrt{3}$, con anchuras comparables debido a que ambos modos experimentan el mismo amortiguamiento.

La descomposición en componentes en fase y en cuadratura permite identificar el papel de la disipación: la componente en cuadratura, asociada a la absorción de energía, presenta máximos en las cercanías de las frecuencias de resonancia, mientras que la componente en fase cambia de signo al atravesarlas. Al aumentar c_1 , los picos se ensanchan y disminuyen su amplitud, en acuerdo con la reducción del factor de calidad del sistema.

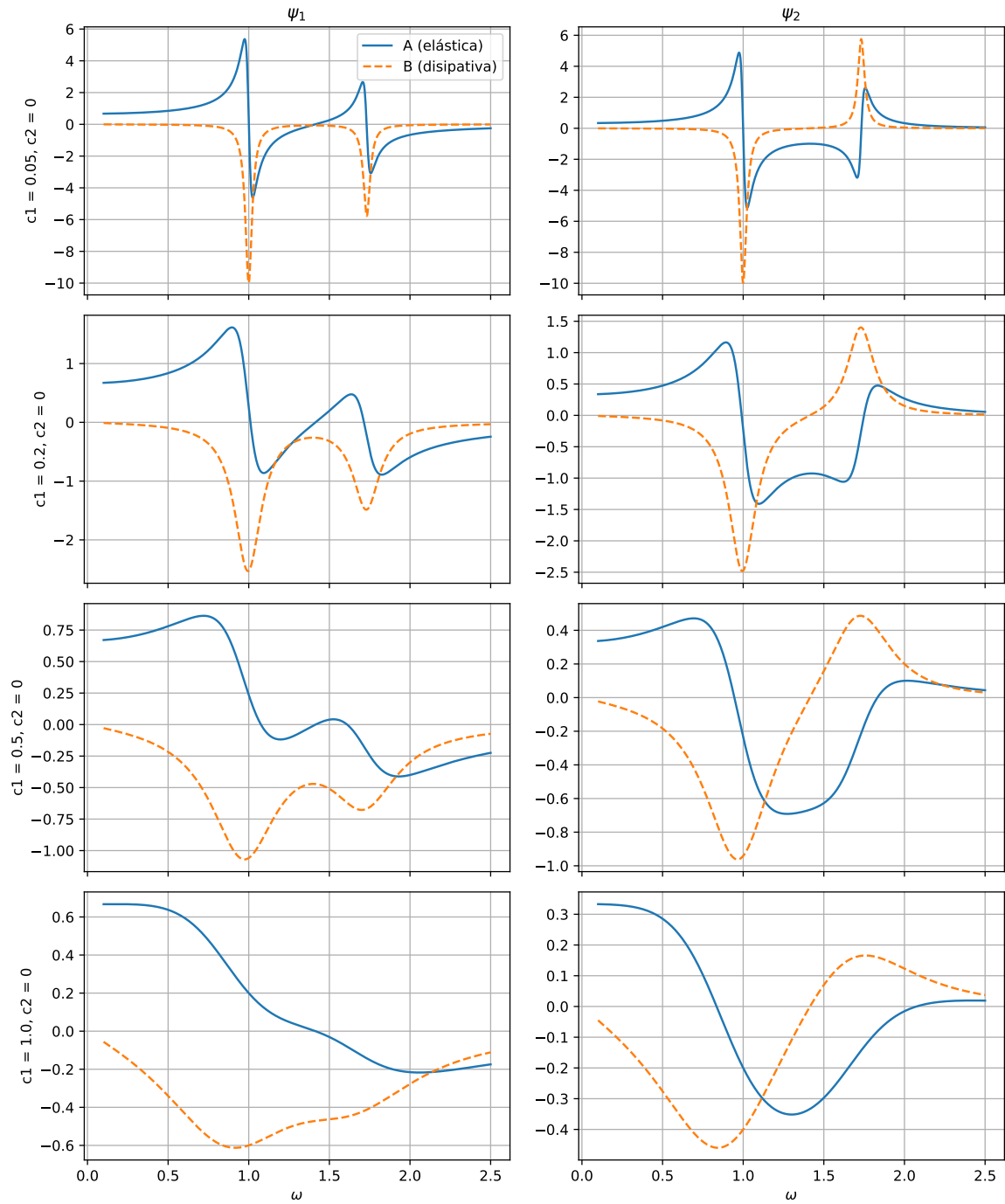


Figura 5.3: Respuesta estacionara vs frecuencia impulsora para el sistema de dos masas acopladas con distintos valores del coeficiente de amortiguamiento c_1 y $c_2 = 0$. Para cada parte móvil se muestran: i) las componentes en fase (línea continua) y ii) en cuadratura (línea punteada).

Este comportamiento contrasta con el que se obtiene cuando $\alpha_2 \neq 0$, donde la disipación introduce diferencias entre los modos y modifica cualitativamente la estructura de las resonancias.

En la figura 5.4 se muestra la respuesta en frecuencia del sistema para distintos valores de α_2 , manteniendo fijo α_1 .

En este caso, los modos presentan coeficientes de amortiguamiento distintos, lo que se traduce en resonancias con anchuras y alturas diferentes. En particular, cuando aumenta α_2 , el modo asociado a la deformación relativa entre las masas se amortigua progresivamente mientras que el modo en fase se ve afectado en menor medida.

Como consecuencia, el pico del modo más disipativo se ensancha y su amplitud se reduce, hasta volverse prácticamente imperceptible para valores grandes de α_2 . En ese régimen, la respuesta queda dominada por el modo en fase.

En el dominio de frecuencias se manifiesta así el mecanismo de selección modal discutido anteriormente: la disipación no solo atenúa la respuesta, sino que favorece aquellos modos que disipan energía más lentamente. Este efecto se refleja también en la componente en cuadratura, cuya contribución se concentra progresivamente en torno a la resonancia del modo menos amortiguado.

La comparación entre las figuras 5.3 y 5.4 muestra que la estructura de la disipación determina no solo la evolución temporal del sistema, sino también las características de su respuesta resonante.

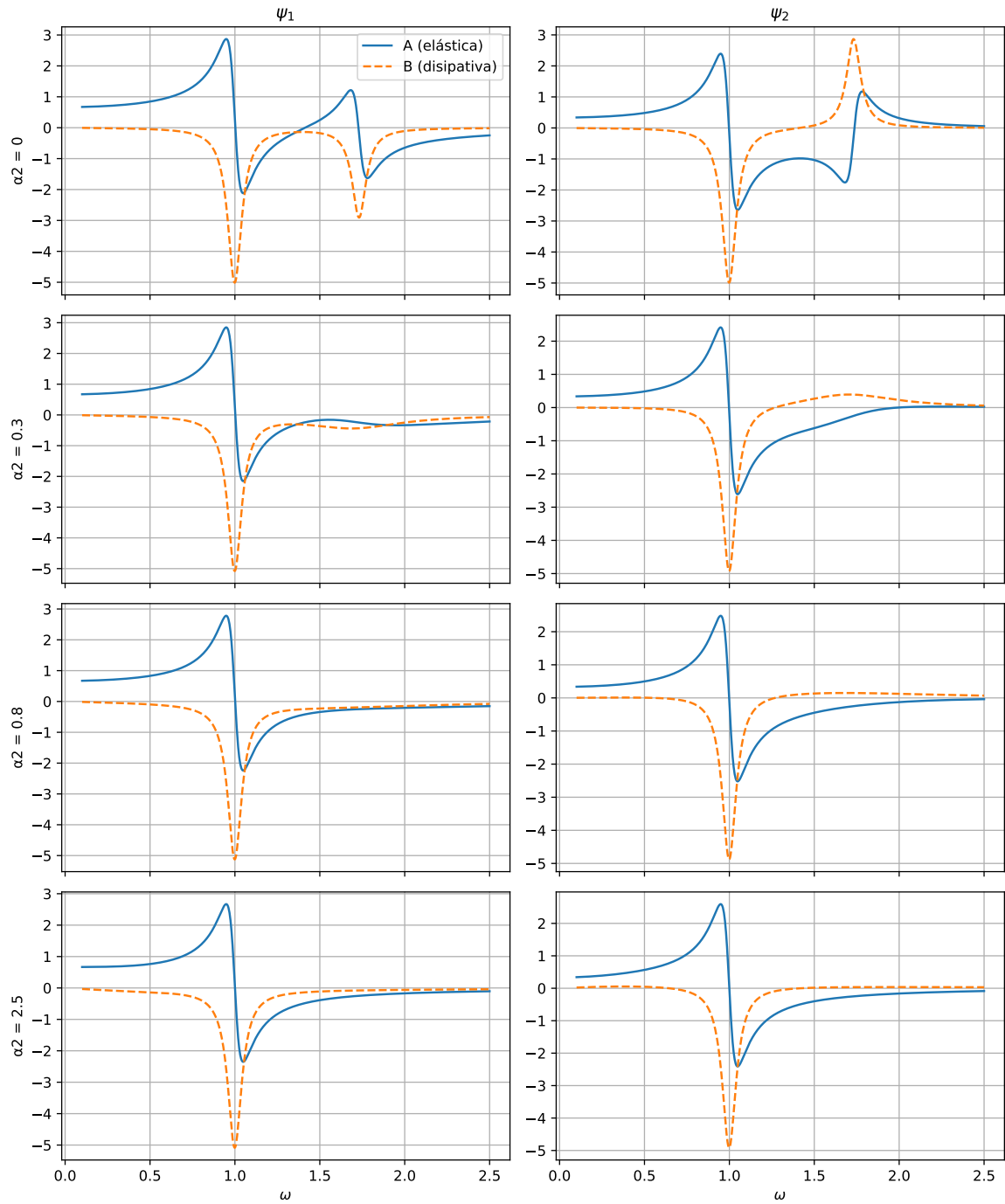


Figura 5.4: Respuesta estacionaria en función de la frecuencia de la fuerza impulsora para el sistema de dos masas acopladas para distintos valores de α_2 , manteniendo fijo $\alpha_1 = 0.1$. Se muestran, para cada parte móvil, las componentes en fase (línea continua) y en cuadratura (línea punteada) de la respuesta estacionaria.

Capítulo 6

Muchas partes móviles

Prestaremos especial atención a las *cadena lineales* de partes móviles, un tipo especial de sistemas donde las propiedades físicas permiten mostrar analíticamente muchas características de los modos propios, sin necesidad de diagonalizar de manera explícita la matriz del sistema (que para estos sistemas es tridiagonal). Conviene destacar los siguientes puntos, que muestran la importancia del estudio que haremos a continuación:

- i) bajo ciertas circunstancias, es posible explicitar soluciones para un valor arbitrario de N , mientras que en el método general no es fácil exhibir el comportamiento de los autovalores y autovectores asociados a los modos sin decir cuánto vale N ;
- ii) abre un camino para explorar el **límite continuo**, cuando hay tantas partes móviles tan juntas que no se ven sus tamaños. En este límite nos encontraremos por primera vez con la **ecuación de ondas clásica** y veremos que los modos del sistema son las **ondas estacionarias**;
- iii) gracias a la inclusión de partes móviles “virtuales”, es posible avanzar en el tratamiento analítico sin necesidad de explicitar qué pasa en el comienzo y el final de la cadena;
- históricamente fue el primer intento de estudiar las **propiedades emergentes** de un sistema físico a partir de la dinámica de sus partes. Las oscilaciones *longitudinales* de las cadenas, es decir las oscilaciones en la dirección en que se extiende el sistema en el equilibrio (el eje x en la figura 6.1) fueron estudiadas por Newton en el libro II de sus Principia (1686) en sus intentos de obtener una expresión para la velocidad del sonido. Si parece extraño que Newton usara cadenas de resortes para estudiar el sonido en el aire, más adelante nos quedará claro que era el único camino que tenía, porque para tratar el aire como medio continuo habría necesitado las ecuaciones diferenciales en derivadas parciales, que todavía no existían. Más tarde, John Bernoulli desde Basilea y Daniel Bernoulli (su hijo) desde San Petersburgo, iniciaron en 1727 una conmovedora correspondencia familiar sobre las propiedades de las oscilaciones *transversales* (es decir, en la dirección perpendicular al eje x) de cadenas con N cuentas equiespaciadas. Entre carta y

carta demostraron que hay N modos con N frecuencias propias. Más tarde Daniel Bernoulli demostró lo que ya sabemos, que el movimiento más general está dado por la superposición de N modos y así estableció un principio de superposición que fue el detonante de lo que luego serían las **series de Fourier**. Por eso, no debe extrañar (y hasta resulta lógico cuando se lo piensa), que haya que encarar intuitivamente algunas herramientas que luego se estudiarán con más rigor en materias de matemática más avanzadas, tal como tuvieron que hacer científicos como los Bernoulli, Taylor o Euler.

6.1. Cadena con N cuentas idénticas equiespaciadas

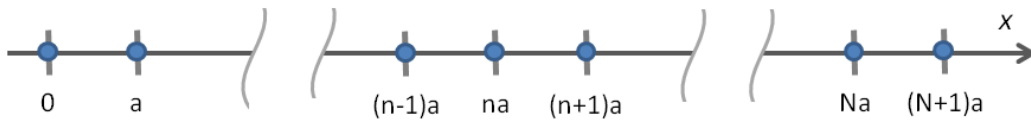


Figura 6.1: N cuentas idénticas equidistantes (separación a) engarzadas en una cuerda elástica sin masa (o unidas por resortes idénticos). Excepto las dos masas de los extremos, el resto de las masas tienen todas dos vecinos cercanos.

Veremos un método especial para resolver el problema de modos normales de un sistema como el bosquejado en la figura 6.1: N cuentas idénticas de masa m , equidistantes (separación a) y engarzadas en una cuerda elástica sin masa o unidas por resortes idénticos. En el equilibrio, todos los tramos de cuerda tienen igual tensión T_0 y si imaginamos que en vez de cuerdas hay resortes de constante elástica k y longitud natural ℓ_0 , resulta $T_0 = k(a - \ell_0)$. La cuenta del comienzo, con $n = 0$ en $x = 0$, y la cuenta del final, con $n = N + 1$ en $x = (N + 1)a = L$, son cuentas virtuales a las que reservaremos un tratamiento aparte cuyos detalles dependerán de cada situación. Así por ejemplo se puede hacer que estos extremos tengan masa infinita para representar una cuerda con extremos fijos, como en una guitarra, o se pueden dejar los extremos libres como si la cuerda fuera una serpiente. Si bien elegimos estudiar apartamientos del equilibrio en la dirección perpendicular a la dirección de equilibrio (problema de oscilaciones transversales), es fácil convencerse de que el mismo tratamiento es válido para el problema de oscilaciones longitudinales.

6.1.1. Ecuaciones de movimiento para oscilaciones transversales

La situación fuera del equilibrio está bosquejada en la figura 6.2. Si llamamos T_I y T_D a las tensiones en las sogas a la izquierda y derecha de la parte móvil n ($n = 1, 2, \dots, N$), ℓ_I y ℓ_D a las longitudes de estas sogas y α_I y α_D a los ángulos que estas sogas forman con la horizontal, la ecuación de movimiento de la parte móvil n resulta

$$m \ddot{\psi}_n = -T_I \sin \alpha_I + T_D \sin \alpha_D. \quad (6.1)$$

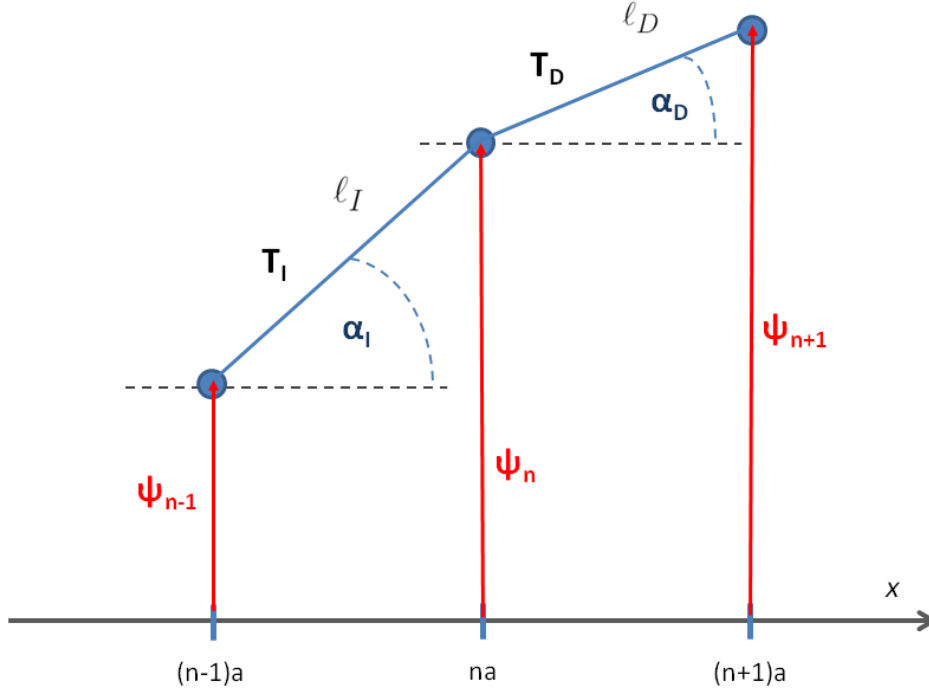


Figura 6.2: Desplazamientos transversales de la cadena con cuentas idénticas equidistantes

Este es un sistema de ecuaciones diferenciales *no lineales* en las funciones incógnita $\psi_n(t)$, porque T_I y T_D

$$T_I = k(\ell_I - \ell_0), \quad T_D = k(\ell_D - \ell_0) \quad (6.2)$$

dependen de cuánto se estiraron las sogas ($\ell_I - \ell_0$ y $\ell_D - \ell_0$) y tanto estos estiramientos como los ángulos α_I y α_D dependen de la posición de la parte móvil n y de sus vecinas cercanas, es decir de $\psi_{n-1}(t)$, $\psi_n(t)$ y $\psi_{n+1}(t)$.

Queda claro entonces que el sistema (6.1) no tiene solamente primeras potencias de los desplazamientos, es decir, es no lineal. Para linealizarlo, usamos la hipótesis de pequeños apartamientos del equilibrio. En este caso aproximamos los senos por las tangentes, desarrollamos en serie de Taylor las raíces cuadradas que aparecen en (??) y conservamos solamente términos lineales en los desplazamientos. Se obtiene así el siguiente sistema

$$\ddot{\psi}_n = \frac{T_0}{m} \frac{\psi_{n+1} - \psi_n}{a} - \frac{T_0}{m} \frac{\psi_n - \psi_{n-1}}{a}, \quad (6.3)$$

una forma que conviene recordar cuando encontremos que en el límite continuo la aceleración de una parte móvil es proporcional a una derivada segunda con respecto a la posición de la parte móvil dentro del sistema (etiquetada por el índice discreto n).

Por conveniencia vamos a distinguir el caso de resortes muy estirables (*slinkies*), donde

$$\ell_0 \ll a, \quad T_I \approx T_D \approx T_0 = ka$$

del caso en que se tiene un resorte común, o una soga elástica, donde

$$\ell_I \approx \ell_D \approx a \quad \text{y} \quad T_I \approx T_D \approx T_0 = k(a - \ell_0).$$

Agrupando términos en (6.3), el sistema de ecuaciones diferenciales linealizado (y homogéneo), para toda masa con dos vecinos cercanos queda

$$\ddot{\psi}_n = \frac{T_0}{ma} (\psi_{n+1} - 2\psi_n + \psi_{n-1}) . \quad (6.4)$$

Hay dos maneras equivalentes de pensar teóricamente el tema de los vecinos cercanos. La primera es imaginando que estamos considerando un sistema infinito, que no empieza ni termina nunca, de esta manera queda asegurado que cualquier parte móvil tiene dos vecinos cercanos y entonces (6.4) vale para cualquier valor de n . La segunda manera es explicitando que el sistema es finito y entonces (6.4) solo vale para $n = 1, 2, \dots, N$, pero no para $n = 0$ (extremo izquierdo) ni para $n = N + 1$ (extremo derecho).

6.1.2. Modos normales

En un modo normal las funciones $\psi_n(t)$ ($n = 1, 2, \dots, N$) deben representar movimientos donde todas las partes móviles oscilen con la misma frecuencia y pasen simultáneamente por la posición de equilibrio. No queda otra posibilidad que estas funciones adopten la forma $\psi_n(t) = A_n \cos(\omega t + \varphi)$. Y como además tienen que ser solución del sistema (6.4), se obtiene la siguiente *relación de recurrencia*¹ para los A_n

$$A_{n+1} - \left(2 - \frac{ma}{T_0} \omega^2\right) A_n + A_{n-1} = 0 . \quad (6.5)$$

La ec. (6.5) se puede reescribir como

$$\frac{A_{n+1} + A_{n-1}}{2 A_n} = 1 - \frac{ma}{2 T_0} \omega^2 . \quad (6.6)$$

Un caso fácil de resolver sería cuando se dan como datos **dos** elementos adyacentes de la sucesión A_n **y la frecuencia** ω , porque en este caso con la ayuda de la relación de recurrencia (6.5) podrían obtenerse todos los elementos A_n . Otro caso fácil de resolver sería cuando se dan como datos **tres** elementos adyacentes, porque (6.6) daría la frecuencia ω y también todos los elementos A_n . Por lo general no estaremos en estos casos fáciles, por ejemplo, en el caso de extremos fijos no se dan ni elementos adyacentes (se dan los elementos A_0 y A_{N+1}) ni la frecuencia. Entonces, hay que aguzar el ingenio para resolver la relación de recurrencia y encontrar la dependencia con n de los elementos A_n de la sucesión buscada. La ec. (6.6) dice que la dependencia adecuada es aquella para la cual el cociente $\frac{A_{n+1} + A_{n-1}}{2 A_n}$ no depende de n . Podríamos ponernos a probar, o preguntarle a un matemático, que seguramente nos va a decir que la dependencia con n más general de recurrencias como la (6.5) tiene la forma

$$A_n = B \sin(n \Theta) + C \cos(n \Theta) , \quad (6.7)$$

¹Una relación de recurrencia vendría a ser la forma discreta de una ecuación diferencial.

donde B , C y Θ son tres constantes a determinar (como en los casos fáciles mencionados anteriormente, donde también necesitábamos tres datos). Es fácil ver que la receta (6.7) realmente funciona, es decir, que el cociente $\frac{A_{n+1}+A_{n-1}}{2A_n}$ no depende de n . Efectivamente,

$$A_{n\pm 1} = B \sin[(n \pm 1)\Theta] + C \cos[(n \pm 1)\Theta],$$

y usando las identidades trigonométricas que dan el seno y el coseno de una suma, es fácil ver que

$$\frac{A_{n+1} + A_{n-1}}{2A_n} = \cos \Theta, \quad (6.8)$$

es decir, que la receta (6.7) funciona independientemente de los valores de B y C . Ahora solo resta ajustar el valor de Θ para que se cumpla la ec. (6.6). Así obtenemos la siguiente relación entre Θ , ω y las magnitudes físicas relevantes del sistema

$$\cos \Theta = 1 - \frac{ma}{2T_0}\omega^2. \quad (6.9)$$

Conviene notar que el valor de Θ correspondiente a cada modo está asociado con un ordenamiento espacial, que se manifiesta como una periodicidad espacial observable en el conjunto de partes móviles. Esto es así porque la secuencia A_n de cada modo se obtiene mediante una expresión (la (6.7)) que satisface

$$A_{n+2\pi/\Theta} = A_n,$$

es decir, que se comporta periódicamente, con período $2\pi/\Theta$, en la variable n , lo que equivale a decir que se comporta periódicamente, con período $2\pi a/\Theta$, a lo largo del eje x . Vemos así que la ec. (6.9) es una relación entre el período temporal asociado con ω y el período espacial asociado con Θ . La relación entre el período temporal (es decir, ω) y el período espacial de un sistema se conoce con el nombre de *relación de dispersión*.

Técnicamente la ec. (6.9) es la *relación de dispersión para los modos transversales de una cadena con N cuentas idénticas equiespaciadas*. Notar que la forma de la relación de dispersión *siempre* viene dada por (6.9), para cualquier cadena como esta, *independientemente de las condiciones de los extremos*. Las condiciones de contorno en los extremos discretizan los valores de Θ y de ω permitidos, es decir, determinan las frecuencias de los modos (los valores permitidos para ω) y la distribución espacial de las partes móviles en cada modo (asociada al valor de Θ que corresponde a cada valor de ω). Sabemos que tiene que haber N modos, aunque hasta ahora no hemos logrado explicitarlos, porque hasta ahora no hemos explicitado las condiciones en los extremos. Desde un punto de vista físico queda claro que tienen que ser las condiciones de contorno las que determinen los valores permitidos para Θ y para ω , porque las maneras que el sistema limitado en el espacio “prefiere” para oscilar colectivamente cuando se lo aparta del equilibrio, tienen que depender de cómo se atan (o no se atan) las fronteras del sistema (en este caso, los extremos de la cadena).

6.1.3. Relación de dispersión

Si reescribimos la relación (6.9) entre Θ , ω y las magnitudes físicas relevantes del sistema usando una identidad trigonométrica que da el coseno en función del argumento

mitad, obtenemos esta forma más linda:

$$\omega^2 = \frac{4T_0}{ma} \sin^2 \frac{\Theta}{2} \implies \omega = \underbrace{2\sqrt{\frac{T_0}{ma}}}_{\omega_{Max}} \left| \sin \frac{\Theta}{2} \right| \quad (6.10)$$

que muestra que aunque todavía no sepamos el valor de Θ , la frecuencia ω de un modo está comprendida en el intervalo $[0, \omega_{Max} = 2\sqrt{\frac{T_0}{ma}}]$. Por ejemplo, en la figura 6.3 se muestra la curva $\omega(\Theta)$ y tres valores de frecuencia obtenidos para un sistema hipotético con tres partes móviles. Como $\omega(\Theta)$ es periódica, a cada valor de frecuencia ω_p sobre el tramo de curva azul (con Θ_p en el intervalo $0 \leq \Theta_p \leq \pi$) le corresponde una doble infinitud de valores de Θ fuera de dicho intervalo: una infinitud de la forma $\Theta_p + 2\pi q$

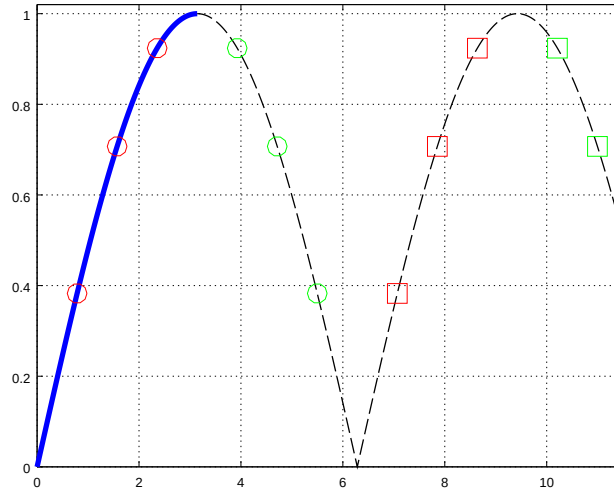


Figura 6.3: Forma de la ecuación de dispersión $\omega(\Theta)$, ec. (6.10), Θ en el eje horizontal y ω/ω_{Max} en el eje vertical.

(en los tramos con pendiente positiva, los símbolos rojos) y otra infinitud de la forma $-\Theta_p + 2\pi q$ (en los tramos con pendiente negativa, los símbolos verdes), q un número entero. Es fácil ver que todos los valores de la infinitud $\Theta_p + 2\pi q$ representan físicamente el mismo movimiento colectivo, porque según la ec. (6.7) todos tienen exactamente los mismos valores de A_n . Dicho de otra manera, $A_n(\Theta_p + 2\pi q) = A_n(\Theta_p)$. Los valores de la infinitud $-\Theta_p + 2\pi q$ también representan la misma situación física, porque el lado derecho de la igualdad (6.8) toma el mismo valor tanto para Θ_p como para $-\Theta_p + 2\pi q$, definiendo así la misma recurrencia, es decir, la misma sucesión de amplitudes modales A_n .

Para ver cómo en qué puntos de la curva de la figura 6.3 caen los simbolitos rojos (y los verdes), es decir, para ver cómo funciona la discretización de los valores de Θ_p (y en consecuencia, de ω_p), especifiquemos las condiciones en los extremos, que es la información que hasta ahora hemos evitado explicitar. Las idealizaciones más comunes son los casos de extremo fijo y de extremo libre.

6.1.4. Modos para dos extremos fijos

Consideramos primero el caso de ambos extremos fijos. En este caso los extremos en $x = 0$ y en $x = L$, figura 6.1, están atados a una pared, lo que equivale a decir que las masas virtuales 0 y $N + 1$ tienen masa tan grande que no se pueden mover. Para los modos esto se traduce en que se anulen las amplitudes de las partes móviles virtuales, es decir, $A_0 = A_{N+1} = 0$. Y de la ec. (6.7) y para que $A_0 = 0$ debe ser $C = 0$. Y si $C = 0$, entonces $B \sin[(N + 1) \Theta] = 0$, una condición que se cumple cuando $B = 0$ (que no nos interesa, porque no representa movimiento) o cuando $\sin[(N + 1) \Theta] = 0$. En resumen, para que en la cadena con extremos fijos haya un modo, la cantidad $(N + 1) \Theta$ debe ser una raíz de la función seno, es decir, que debe ser $(N + 1) \Theta = p \pi$, con p un número entero arbitrario. Los modos quedan así etiquetados por este índice p . En cada modo los desplazamientos de las partes móviles son

$$\psi_n^{(p)}(t) = A_n^{(p)} \cos(\omega_p t + \varphi_p) = \overbrace{B^{(p)} \sin\left[\underbrace{\frac{p \pi}{(N + 1)} n}_{\Theta^{(p)}}\right]}^{A_n^{(p)}} \cos(\omega_p t + \varphi_p), \quad (6.11)$$

con

$$\omega_p = \omega_{Max} \left| \sin \frac{p \pi}{2(N + 1)} \right|, \quad (6.12)$$

En el método matricial los modos se visualizan a través de dos cantidades: los autovalores, que dan el comportamiento temporal del conjunto, y los autovectores, que dan el comportamiento espacial de cada parte móvil en la oscilación colectiva. Análogamente, en el método de recurrencias que estamos estudiando, cada modo queda caracterizado por dos cantidades que podemos calcular en términos de p y de las magnitudes físicas relevantes del sistema (T_0 , a , m , N y $L = (N + 1)a$). Estas cantidades son

$$\Theta^{(p)} = \frac{p \pi}{(N + 1)} = \frac{p \pi a}{L},$$

que da el *comportamiento espacial* de cada parte móvil en la oscilación colectiva, y

$$\omega_p(\Theta^{(p)}),$$

que se obtiene de la ecuación de dispersión (6.10) y que rige el *comportamiento temporal* del conjunto.

Cuando hay muchas partes móviles, puede ser conveniente etiquetar la n -ésima parte móvil con la coordenada $x = na$ (ver figura 6.1). Esta coordenada permite indentificar a la parte móvil como aquella que en el equilibrio está ubicada sobre el eje x en la posición na respecto a un origen. En esta descripción, y entendiendo que x toma solamente valores discretos, podemos reemplazar en el argumento del primer seno en la ec. (6.11)

$$\Theta^{(p)} n \quad \text{por} \quad \underbrace{\frac{\Theta^{(p)}}{a}}_{k_p} x. \quad (6.13)$$

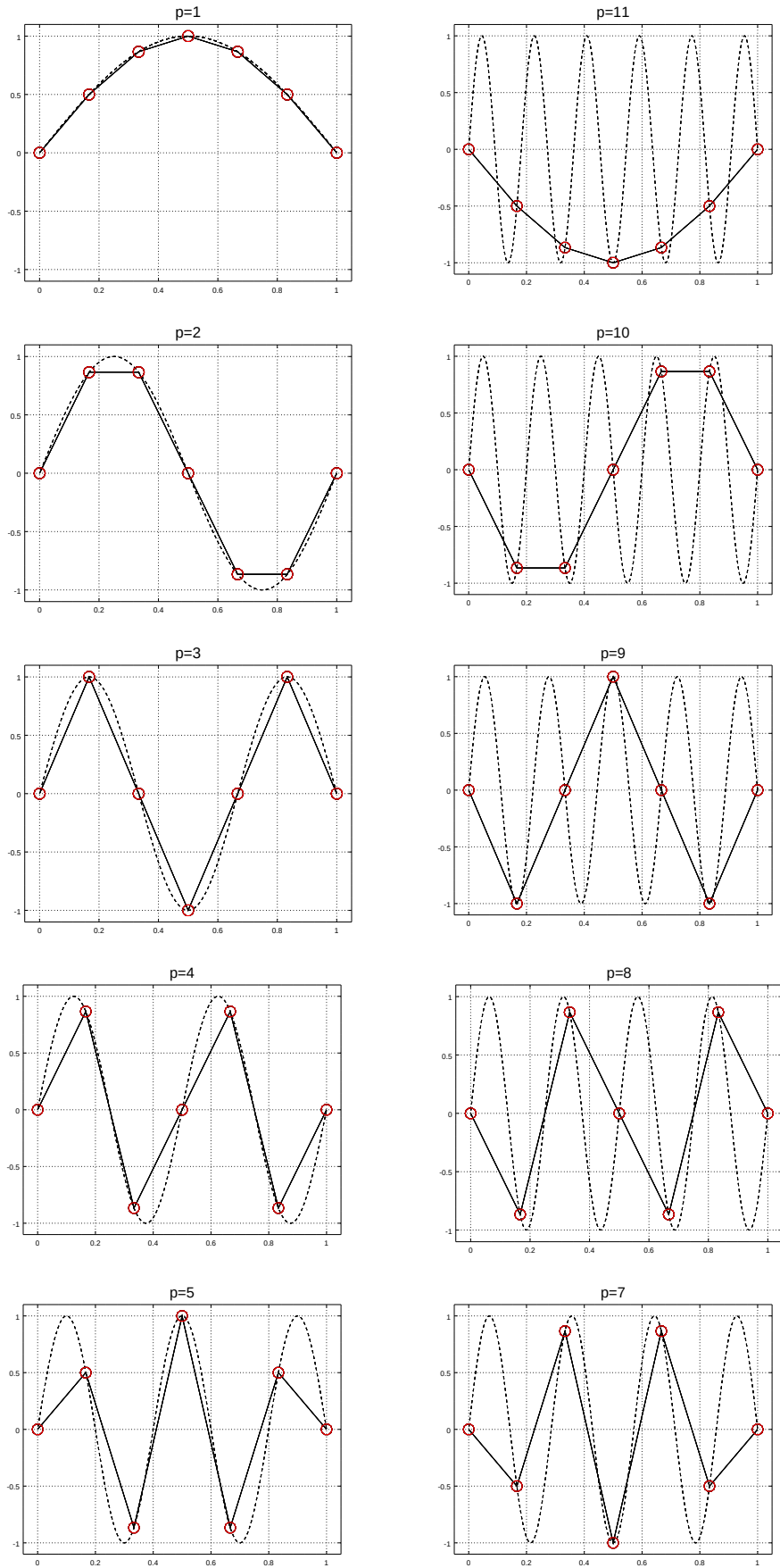


Figura 6.4: Forma de los modos para distintos valores de p en la ec. (6.11).

Este reemplazo pone en evidencia que en los modos de la ec. (6.11) la cantidad $k_p = \Theta^{(p)}/a$ juega el papel de una frecuencia espacial y que entonces el período espacial de las funciones de x (de n) asociadas con los modos es

$$\lambda_p = \frac{2\pi}{k_p} = \frac{2L}{p}. \quad (6.14)$$

Al período espacial asociado con movimientos armónicos colectivos de sistemas (limitados o no) se lo llama **longitud de onda** y se denota con la letra griega λ (lambda minúscula) y a la frecuencia angular espacial asociada con este período se lo llama **número de onda** y se denota con la letra k (una elección muy desafortunada cuando hay resortes involucrados).

Notemos que en el caso de extremos fijos se verifican las propiedades que ya habíamos anunciado en la figura 6.3 para el caso general de extremos con condiciones arbitrarias. En primer lugar vemos que hay solamente N frecuencias normales en el tramo azul de la curva de la figura 6.3 y son las que tienen $\Theta^{(p)} = p\pi/(N+1)$ con $p = 1, 2, \dots, N-1, N$ (los circulitos rojos). El valor $p = 0$ está excluido porque con frecuencia cero no hay movimiento colectivo y el valor $p = N+1$ también porque para $\Theta^{(N+1)} = \pi$, $\omega_{(N+1)} = \omega_{Max}$ y tampoco hay movimiento colectivo. En segundo lugar, las otras N soluciones en el siguiente tramo de la curva de dispersión de la ec. (6.10), el tramo de trazos discontinuos y con pendiente negativa (los circulitos verdes en la figura 6.3), corresponden a $p = N+2, N+3, \dots, 2N+1, 2N+2$ y guardan la siguiente correspondencia con las frecuencias del tramo anterior

$$\underbrace{\omega_{N+1+j}}_{\text{circulitos verdes}} = \underbrace{\omega_{N+1-j}}_{\text{circulitos rojos}}, \quad j = 1, 2, \dots, N-1, N. \quad (6.15)$$

Por lo discutido en la sección 6.1.3, las soluciones con $p = N+2, N+3, \dots, 2N+1, 2N+2$, describen configuraciones de modos que ya han sido tenidas en cuenta por las amplitudes relativas de las partes móviles obtenidas para las soluciones con $p = 1, 2, \dots, N-1, N$. Esto está ilustrado en la figura 6.4, donde se muestra la secuencia A_n^p en una cadena con $N = 5$ cuentas y con extremos fijos. Las curvas de trazos representan las funciones $A_n^{(p)}$ de la ec. (6.12) y solamente sirven como una guía para ubicar las partes móviles. No representan la forma de la cuerda con cuentas, porque las sogas (o los resortes) entre cuentas están tensas y entonces corresponden a tramos rectos de las figuras. En la columna de la izquierda se muestran las secuencias A_n^p para $1 \leq p \leq N = 5$ (circulitos rojos en la figura 6.3), mientras que en la columna de la izquierda se muestran las secuencias A_n^p para $N+1 = 6 \leq p \leq 2N = 10$ (circulitos verdes en la figura 6.3). Observar que la secuencia A_n para las soluciones con $p = N+1+j$ (verdes), difiere en un factor constante de la secuencia A_n para soluciones con $p = N+1-j$ (rojas), con $j = 1, 2, \dots, N-1, N$. Nótese que las configuraciones espaciales para los casos $p = 0$ y $p = 6$ no se muestran, ya que corresponden a todas las cuentas en su posición de equilibrio.

Los modos más bajos tienen $\Theta^{(p)}$ más chico, es decir que en el dibujo de la figura 6.3 están ubicados sobre la parte de la curva de dispersión donde el seno se puede aproximar por una función lineal. Entonces, para los modos más bajos, ω depende

linealmente de k , es decir que los modos más bajos están sobre el tramo casi lineal de la parte azul de la figura 6.3. Notar que en este caso, los valores de las frecuencias de los modos son múltiplos de la frecuencia más baja o fundamental

$$\omega_p \approx \omega_{Max} \frac{p \pi}{2(N+1)} = p \sqrt{\frac{T_0 a}{m}} \frac{\pi}{L} = p \omega_1, \quad (6.16)$$

con

$$\omega_1 = \sqrt{\frac{T_0 a}{m}} \frac{\pi}{L}.$$

Resumiendo, siempre que se satisfagan las hipótesis necesarias para que se cumpla la aproximación lineal, el movimiento más general de cada parte móvil de la cadena de masas idénticas y equiespaciadas con extremos fijos y condiciones iniciales arbitrarias, se escribe como superposición de los modos dados por la ec. (6.11)

$$\psi_n(t) = \sum_{p=1}^N B^{(p)} \sin\left[\frac{p \pi}{(N+1)} n\right] \cos(\omega_p t + \varphi_p), \quad (6.17)$$

con $n = 1, 2, \dots, N$ (etiqueta de cada parte móvil) y ω_p dado por la ec. (6.12).

Condiciones iniciales

En nuestra solución general obtenida en (6.17), las N amplitudes $B^{(p)}$ y las N fases iniciales φ_p son incógnitas que se obtienen a partir de las $2N$ condiciones iniciales: N posiciones iniciales de las partes móviles

$$\psi_n(0) = \sum_{p=1}^N B^{(p)} \sin\left[\frac{p \pi}{(N+1)} n\right] \cos(\varphi_p), \quad (6.18)$$

y otras N velocidades iniciales de las partes móviles

$$\dot{\psi}_n(0) = - \sum_{p=1}^N \omega_p B^{(p)} \sin\left[\frac{p \pi}{(N+1)} n\right] \sin(\varphi_p). \quad (6.19)$$

Teniendo en cuenta que cualquier condición inicial se puede representar con un dibujito como los que están en la columna izquierda de la figura 6.4, pero con las partes móviles en posiciones bastante arbitrarias (como que las sogas/resortitos sigan siendo elásticas y que se comporten linealmente), el resultado de la ec. (6.18) es bastante sorprendente. Porque esta igualdad está diciendo que cualquier dibujito arbitrario para las posiciones iniciales de las N partes móviles equiespaciadas se puede escribir como combinación lineal de los dibujitos “normales”, es decir, los dibujitos que corresponden a los modos. Para pensar ... este resultado está relacionado con el hecho que los autovectores que generan el cambio de base son ortogonales ... a retomar más adelante ². Por ahora, nos podríamos quedar con la demostración de la siguiente identidad

$$\sum_{n=1}^N \sin\left[\frac{p \pi}{(N+1)} n\right] \sin\left[\frac{q \pi}{(N+1)} n\right] = \frac{N+1}{2} \delta_{pq}, \quad (6.20)$$

²Prestar atención, actualizaciones pronto

con δ_{pq} la delta de Kronecker (igual a 1 si $p = q$, o igual a 0 cuando $p \neq q$).

6.1.5. Condición para extremo libre

Supongamos que un extremo de la cadena, por ejemplo el de la derecha, ver figura 6.1, no está fijo sino que tiene un anillo de masa $m_{(N+1)}$ que desliza sin rozamiento por una varilla vertical. Para este anillo, la ecuación de movimiento es parecida a la (6.3), pero con M_{N+1} en vez de m y sin el aporte del vecino derecho

$$M_{N+1} \ddot{\psi}_{N+1} = -T_0 \frac{\psi_{N+1} - \psi_N}{a}. \quad (6.21)$$

Vemos que esta condición de extremo es bastante general, en el sentido que incluye al caso anterior de extremo fijo, que correspondería a un anillo de masa muy grande ($M_{N+1} \rightarrow \infty$). Es común idealizar un extremo libre como el límite opuesto, es decir, un anillo sin masa ($M_{N+1} \rightarrow 0$). En este caso, la condición (6.21) exige

$$\psi_{N+1} = \psi_N$$

y esta igualdad provee una relación entre las constantes B , C y Θ que determinan las amplitudes A_n según la ec. (6.7). Otra relación entre las mismas constantes se obtiene cuando se explicita la condición sobre la masa virtual del extremo izquierdo. Procediendo de una manera completamente análoga a la empleada para extremos fijos, se obtiene ahora la discretización de los modos, que de nuevo quedan etiquetados por un índice discreto p .

6.1.6. Oscilaciones longitudinales

Hasta aquí hemos estudiado las oscilaciones **transversales** de la cadena con N cuentas idénticas equiespaciadas. Es fácil comprobar que las oscilaciones **longitudinales** (es decir, los apartamientos del equilibrio en la dirección que ocupa el sistema cuando está en reposo) admiten un tratamiento formalmente idéntico. Si las funciones $\psi_n(t)$ representan ahora los apartamientos longitudinales, como en la figura 6.5, en vez de los apartamientos transversales como en la figura 6.2, las ecuaciones de movimiento también tienen la forma (6.4), con T_0 la tensión de las sogas en el equilibrio (caso sogas elásticas) o $T_0 = ka$ (caso de resortes muy estirables). Luego, los modos normales se obtienen siguiendo el tratamiento ya visto en 6.1.2, la relación de dispersión sigue siendo (6.10), las frecuencias de resonancia para extremos fijos vienen dadas por la ec. (6.12) y las coordenadas normales para los distintos modos están representadas por gráficos como los de la figura 6.4, con la única diferencia de que las funciones seno asociadas con cada modo no se visualizan tan directamente como en el caso de las oscilaciones transversales.

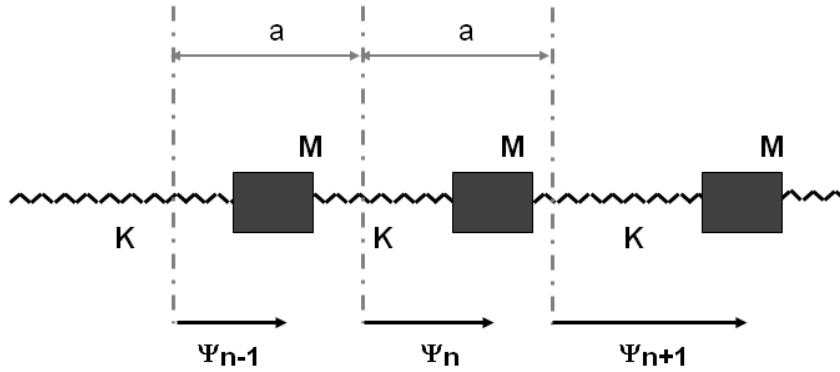


Figura 6.5: Desplazamientos longitudinales de la cadena con cuentas idénticas equidistantes

6.2. Péndulos idénticos acoplados

En las cadenas lineales vistas hasta ahora, las ecuaciones de movimiento en la aproximación lineal tenían la forma (6.3), o (6.4), tanto para oscilaciones longitudinales como para transversales. Es claro que tal sistema de ecuaciones diferenciales acopladas conducirá siempre a relaciones de dispersión de la misma forma que (6.10). Esto quiere decir que, independientemente de las condiciones en los extremos, y tal como se representa en la figura 6.3, las frecuencias de los modos normales estarán comprendidas en el intervalo $0 \leq \omega \leq \omega_{Max}$. Que las frecuencias de resonancia del sistema tengan como cota mínima el valor nulo, está claramente relacionado con el hecho de que la fuerza resititutiva sobre una parte móvil, proveniente solo de los vecinos cercanos, son proporcionales a los desplazamientos relativos entre vecinos cercanos y, en consecuencia, la fuerza resititutiva es cero cuando las posiciones relativas entre partes móviles no cambian. A continuación veremos un sistema donde, además de estas características, aparece también una nueva fuerza resititutiva que depende solamente del desplazamiento de la parte móvil de su posición de equilibrio. Y entonces, la fuerza resititutiva total ya no es cero, ni siquiera en el caso en que las posiciones relativas entre partes móviles no cambien. El sistema en cuestión consiste en un conjunto de péndulos idénticos (masa M , longitud ℓ), acoplados por resortes también idénticos (constante elástica K), esquematizado en la figura 6.6. En este caso, y de acuerdo a lo discutido anteriormente, el sistema de ecuaciones de Newton se escribe igual que el sistema (6.4), excepto por un término adicional que representa la fuerza resititutiva gravitatoria

$$\ddot{\psi}_n = -\omega_0^2 \psi_n + \frac{K}{M} (\psi_{n+1} - 2\psi_n + \psi_{n-1}) , \quad (6.22)$$

con $\omega_0^2 = g/\ell$.

Para analizar los modos de este sistema de péndulos acoplados, partimos de la ec. (6.22) y repetimos de manera formalmente idéntica los procedimientos realizados para analizar los modos en la eq. (6.4), es decir el caso $\omega_0 = 0$, página 88. Es claro que si

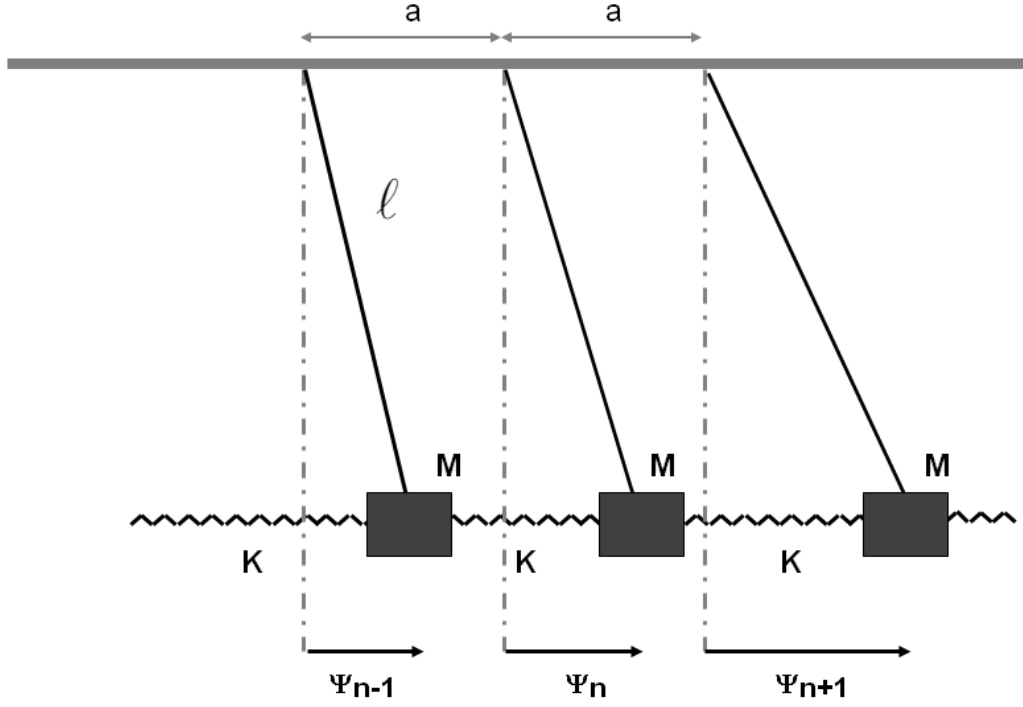


Figura 6.6: Esquema para analizar un conjunto de péndulos idénticos, acoplados mediante resortes también idénticos, un modelo de sistema donde actúan dos tipos de fuerzas restitutivas, i) las elásticas, que dependen de la posición relativa entre vecinos cercanos y ii) la gravitatoria, que no depende de la posición relativa entre vecinos cercanos, sino sólo del apartamiento local del equilibrio.

antes llegábamos a una eq. en diferencias que se puede reescribir como

$$\omega^2 = 2 \frac{K}{M} \left(1 - \frac{A_{n+1} + A_{n-1}}{2 A_n} \right), \quad (6.23)$$

(ver (6.5)), ahora vamos a llegar a una ecuación en diferencias similar, pero con el nuevo término que corresponde a $\omega_0 \neq 0$

$$\omega^2 = \omega_0^2 + 2 \frac{K}{M} \left(1 - \frac{A_{n+1} + A_{n-1}}{2 A_n} \right), \quad (6.24)$$

que se puede interpretar recordando que el lado derecho de (8.12) da la fuerza restitutiva por unidad de desplazamiento y por unidad de masa cuando $\omega_0 = 0$ (no hay gravedad o péndulos muy largos). Cuando $\omega_0 \neq 0$, la fuerza restitutiva aumenta de la misma manera para todas las masas y se suma a la la fuerza restitutiva por unidad de desplazamiento y por unidad de masa que ya tenían las partes móviles sin gravedad.

Y como ya sabemos que el cociente entre amplitudes del lado derecho de la ec. (8.12) se puede hacer independiente de n con sucesiones de la forma

$$A_n = B \sin(n \Theta) + C \cos(n \Theta), \quad (6.25)$$

llegamos así a una nueva relación de dispersión (ver eq. (6.10))

$$\omega^2 = \omega_0^2 + \underbrace{\frac{4K}{M}}_{\omega_{MAX}^2} \sin^2 \frac{\Theta}{2}, \quad (6.26)$$

válida para la relación entre el período temporal y el período espacial en un modo. Como sucedía en el caso $\omega_0 = 0$, las frecuencias de los modos quedan discretizadas cuando se imponen las condiciones en los extremos, aunque siempre estarán ubicadas sobre la curva $\omega(\Theta)$ definida por la ec. y las situaciones de interés físico serán las que correspondan al intervalo $0 \leq \Theta \leq \pi$ (o cualquier otro intervalo, aunque por simplicidad conviene elegir $[0, \pi]$).

En el caso en que los péndulos virtuales de los extremos estén fijos a una pared, el tratamiento es completamente similar al que se hizo para la cadena con cuentas, donde $\omega_0 = 0$, adaptando el significado de $\psi_n(t)$. De esta manera, vemos que el movimiento más general de cada parte móvil del sistema esquematizado en la figura 6.6, con extremos fijos y condiciones iniciales arbitrarias, se escribe también como (6.12), es decir como superposición de modos dados por la ec. (6.11), que solo difieren en la relación de dispersión, es decir en la manera en que ω_p se obtiene a partir de Θ_p .

6.3. Aproximación continua

Veremos cómo tratar el caso límite en que el número de partes móviles del sistema es muy grande. La idea general se aplica a muchos sistemas reales y para fijar ideas podríamos pensar en las oscilaciones transversales de una cadena tensionada con N cuentas idénticas equiespaciadas, como en las figuras 6.1 o 6.2, con cuentas cada vez más chiquitas y más juntas, tan chiquitas y tan juntas que no seremos capaces de ver la separación entre cuentas a simple vista. Es decir que no podremos decir si se trata de las oscilaciones transversales de un collar con cuentas o de una cuerda. Análogamente, también podríamos pensar en las oscilaciones longitudinales de una cadena similar, como en la figura 6.5, o de sistemas con otras fuerzas restitutivas, como el sistema de péndulos acoplados de la figura 6.6. En todos estos casos esperamos que, a partir de un valor suficientemente grande de N , el número de partes móviles del sistema deje de importar y que solamente importe la inercia o la elasticidad de una cierta longitud del sistema. Es decir, esperamos que a partir de un dado valor de N , el movimiento colectivo de un metro de sistema cuando hay 10^{23} partes móviles con masa m , sea casi indistinguible del movimiento colectivo de un metro de sistema cuando hay el doble de partes móviles con masa $m/2$, o cuando hay el cuádruple de partes móviles con masa $m/4$, o el óctuple con masa $m/8$, etc. En estos casos es conveniente cambiar la notación y dejar de etiquetar a las partes móviles con un índice discreto, porque cada vez nos interesa menos la individualidad de cada parte móvil, dicho esto sin intención de ofender a ninguna parte móvil.

Una alternativa al índice discreto para etiquetar partes móviles sería pintar el sistema de distintos colores y describir el movimiento diciendo “ahí está la parte azul, ahí está la

parte verde, etc.”. No vamos a usar esta alternativa, porque no siempre se puede pintar lo que uno quiere estudiar. La alternativa que vamos a usar para etiquetar las partes móviles es relacionarlas con las posiciones que tenían en el equilibrio. Por ejemplo, para sistemas como los de las figuras 6.1, 6.5, o 6.6, la etiqueta sería la coordenada x . Entonces, en vez de describir los desplazamientos mediante un conjunto numerable de funciones de una variable $\psi_n(t)$, usaremos una función de dos variables $\psi(x, t)$. La primera variable, x , juega el papel del índice discreto n y hace referencia a la parte móvil que estamos mirando, que es aquella que en el equilibrio estaba en la posición $x = na$. La segunda variable, t , sigue siendo el tiempo. En síntesis: para pasar en cada caso el sistema de ecuaciones diferenciales a la nueva notación, hay que hacer los siguientes reemplazos

$$\begin{aligned}\psi_n(t) &\longrightarrow \psi(x = na, t) \\ \psi_{n+1}(t) &\longrightarrow \psi(x + a, t) \\ \psi_{n-1}(t) &\longrightarrow \psi(x - a, t) \\ \ddot{\psi}_n(t) &\longrightarrow \partial^2 \psi(x = na, t) / \partial t^2.\end{aligned}\tag{6.27}$$

Notar que para sistemas bidimensionales, por ejemplo la tela tensa de una cama elástica, deberíamos usar las dos coordenadas cartesianas x e y , o las coordenadas polares ρ y θ . Y para un sistema tridimensional, por ejemplo un gas en un recipiente, deberíamos usar las tres coordenadas cartesianas x , y y z , o coordenadas esféricas.

6.3.1. Cuerda $\implies$ ecuación de ondas clásica

En la nueva notación y en el caso de oscilaciones transversales de la cadena con cuentas, el sistema de ecuaciones diferenciales (6.4) queda

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{T_0}{m a} \left[\psi(x + a, t) - 2\psi(x, t) + \psi(x - a, t) \right].\tag{6.28}$$

Observar que hasta acá solamente se trata de un cambio de notación, porque x toma solamente los valores discretos na . Pero si ahora pensamos en el límite físico de muchas partes móviles muy juntas, entonces podemos aproximar x por una variable continua, que es en el fondo lo que hacemos implícitamente en la ferretería cuando pedimos una cierta longitud de cable, en vez de especificar el número de partes móviles. En este caso, $\psi(x, t)$ es una función de x que da la foto del sistema en cada instante t . Y como a es muy chiquito y si la foto no es muy “patológica”³, entonces podemos desarrollar $\psi(x \pm a, t)$ en serie de Taylor alrededor de x

$$\psi(x \pm a, t) = \psi(x, t) \pm a \frac{\partial \psi}{\partial x}(x, t) + a^2 \frac{1}{2!} \frac{\partial^2 \psi}{\partial x^2} \pm a^3 \frac{1}{3!} \frac{\partial^3 \psi}{\partial x^3} + a^4 \frac{1}{4!} \frac{\partial^4 \psi}{\partial x^4} + \dots.\tag{6.29}$$

Introduciendo estos desarrollos en (6.28), todos los términos con derivadas impares se cancelan, mientras que los términos con derivadas pares se refuerzan. Así obtenemos

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{T_0}{m a} \left[a^2 \frac{\partial^2 \psi}{\partial x^2} + a^4 \frac{2}{4!} \frac{\partial^4 \psi}{\partial x^4} + \dots \right]\tag{6.30}$$

³es decir, suponemos que en las fotos de sistemas reales las derivadas que aparecen en los desarrollos de Taylor están bien definidas

una expresión que diría que la aceleración vertical de este sistema, en el límite matemático $a \rightarrow 0$, es nula. Un resultado feo pero consistente con el hecho de que cuando pusimos más partes en el sistema, dejamos constante la masa de cada parte móvil. Y en el límite de $N \rightarrow \infty$ y $a \rightarrow 0$, llegamos a un sistema con masa infinita, por eso la aceleración vertical resulta cero, porque no lo podemos mover.

Pero el procedimiento anterior, si bien correcto matemáticamente, es inconsistente con la motivación física inicial, donde esperábamos que a partir de un dado valor de N no importara el número de partes móviles contenidas en una longitud fija del sistema. Para que la masa del sistema con longitud L sea independiente del número de partes móviles, tenemos que pedir que en este límite la cantidad m/a permanezca constante. Vemos así que en nuestra descripción aparece naturalmente una magnitud ρ

$$\rho = \lim_{a \rightarrow 0} \frac{m}{a}, \quad (6.31)$$

llamada densidad lineal de masa, o masa por unidad de longitud, que en el caso de cuentas idénticas equiespaciadas toma un valor independiente de x . Así, el lado izquierdo de (6.28) no tiende a cero cuando $a \rightarrow 0$ y el sistema de ecuaciones para el movimiento del sistema se transforma en una proporcionalidad entre dos derivadas segundas parciales

$$\frac{\partial^2 \psi}{\partial t^2} = \underbrace{\frac{T_0}{\rho}}_{v^2} \frac{\partial^2 \psi}{\partial x^2}, \quad (6.32)$$

la derivada segunda temporal (la aceleración) y la derivada segunda espacial (relacionada con la fuerza restitutiva). Esta ecuación es la **ecuación de ondas clásicas** en su versión unidimensional para un sistema que ocupa una dirección fija en el espacio (o sea, unidimensional en cartesianas, no en esféricas). El factor de proporcionalidad entre las dos derivadas segundas parciales en (6.32)

$$v^2 = \frac{T_0}{\rho}, \quad (6.33)$$

tiene unidades de velocidad al cuadrado, v recibe el nombre de **velocidad de fase** y ya podemos intuir que es una velocidad asociada con la onda.

A primera vista es sorprendente que el conjunto de ecuaciones diferenciales ordinarias acopladas que describía la dinámica del conjunto de partes móviles del sistema, haya colapsado en esta sencilla ecuación diferencial en derivadas parciales. Aunque veremos que no tan sencilla, en particular cuando vayamos descubriendo la gran diversidad de soluciones que admite y la gran cantidad de fenómenos físicos que estas soluciones describen.

6.3.2. Slinky $\implies$ ecuación de ondas clásica

Como se discutió en 6.1.6, las oscilaciones longitudinales de una cadena con N cuentas idénticas equiespaciadas unidas por resortes idénticos, también tienen ecuaciones de

movimiento de la misma forma que (6.4), excepto que ahora $\psi_n(t)$ representan apartamentos longitudinales, y que $T_0 = ka$, porque en vez de sogas hay resortes. La ecuación análoga a (6.4) se escribe como

$$\ddot{\psi}_n = \frac{K}{M} (\psi_{n+1} - 2\psi_n + \psi_{n-1}) . \quad (6.34)$$

Aparte de estos detalles, sigue siendo válido el cambio de notación (6.28) y también podemos repetir los desarrollos (6.29) para prepararnos para el límite $a \rightarrow 0$. Al reemplazar en (6.34), se siguen cancelando los términos con derivadas impares y reforzando los términos con derivadas pares y así obtenemos

$$\frac{\partial^2 \psi}{\partial t^2} = \frac{k a^2}{m} \left[\frac{\partial^2 \psi}{\partial x^2} + a^2 \frac{2}{4!} \frac{\partial^4 \psi}{\partial x^4} + \dots \right] \quad (6.35)$$

La aceleración en el límite matemático $a \rightarrow 0$ de nuevo parecería nula, pero en nuestro límite físico queremos que las propiedades del sistema con longitud L sean independientes del número de partes móviles. Aplicado a la masa, y al igual que en el caso de las sogas con cuentas, esto impone que la cantidad m/a permanezca constante cuando $a \rightarrow 0$. Y aplicado a la elasticidad equivalente, esto impone que a medida que se agregan resortes, se cambie la constante k de los mismos, de manera tal que el resorte equivalente de longitud L formado por resortes en serie con constantes k , tenga una elasticidad que sea independiente de N , es decir de a . Como la constante equivalente K_e de N resortes en serie con la misma constante k es

$$\frac{1}{K_e} = \frac{1}{k} + \frac{1}{k} + \dots + \frac{1}{k} = \frac{N}{k} = \frac{L}{ak}$$

y teniendo en cuenta que la propiedad extensiva (la que se suma cuando se aumenta el número de partes del sistema) es la inversa de las constantes de los resortes, en el límite físico tenemos que pedir que la constante inversa del resorte por unidad de longitud,

$$\frac{1}{K_e L}$$

sea una propiedad del sistema, independientemente de la longitud. Entonces, la cantidad ka^2/m que aparece en (6.35) tiende a una constante

$$\frac{ka^2}{m} = \frac{K_e L}{\rho}$$

y en el límite $a \rightarrow 0$, (6.35) se transforma, igual que en (6.32), en una proporcionalidad entre las dos derivadas segundas parciales

$$\frac{\partial^2 \psi}{\partial t^2} = \underbrace{\frac{K_e L}{\rho}}_{v^2} \frac{\partial^2 \psi}{\partial x^2} . \quad (6.36)$$

De nuevo encontramos la **ecuación de ondas clásicas** en su versión unidimensional para un sistema que ocupa una dirección fija en el espacio, pero ahora la velocidad de fase, el factor de proporcionalidad entre las dos derivadas segundas parciales con unidades de velocidad al cuadrado, es

$$v^2 = \frac{K_e L}{\rho} . \quad (6.37)$$

6.3.3. Péndulos acoplados $\implies$ ecuación de Klein-Gordon

De manera completamente análoga a lo hecho en 6.3.1 y 6.3.2, empleemos ahora la descripción continua para estudiar el caso límite de $a \rightarrow 0$ en un sistema de péndulos idénticos acoplados como el visto en 6.2. Empezamos por cambiar la notación en el sistema de ecuaciones diferenciales, originalmente dado por (6.22). Seguimos empleando los reemplazos (6.28) y también usamos los desarrollos de Taylor (6.29) para prepararnos para el límite $a \rightarrow 0$. Al reemplazar en (6.22), se siguen cancelando los términos con derivadas impares y reforzando los términos con derivadas pares y luego de tener en cuenta las consideraciones sobre las propiedades intensivas de una longitud L de sistema de péndulos acoplados, obtenemos la siguiente ecuación en derivadas parciales

$$\frac{\partial^2 \psi}{\partial t^2} = -\omega_0^2 \psi + v^2 \frac{\partial^2 \psi}{\partial x^2}, \quad (6.38)$$

con v^2 dado por (6.37). Esta ecuación se parece mucho a la ecuación de ondas clásica, excepto que ahora difiere en un nuevo término que contiene a la función ψ multiplicada por ω_0^2 y se conoce como la **ecuación de ondas de Klein-Gordon** unidimensional (se reduce a la ecuación de ondas clásicas (6.36) en el caso $\omega_0 = 0$).

6.4. La ecuación de ondas

Como se discutió en el contexto de las ecs. (6.13) y (6.14), la ec. de dispersión (6.10) para el sistema discreto se puede reescribir como una relación entre ω y la frecuencia espacial $k = \Theta/a$. En el límite continuo ($a \rightarrow 0$) el valor del seno en (6.10) se puede aproximar por su argumento

$$\omega = 2\sqrt{\frac{T_0}{m a}} \left| \sin \frac{k a}{2} \right| \approx \sqrt{\frac{T_0 a}{m}} k, \quad (6.39)$$

y la ec. de dispersión $\omega(k)$ para una cuerda homogénea es una función lineal, de pendiente

$$v = \sqrt{\frac{T_0}{\rho}}. \quad (6.40)$$

Esta pendiente tiene unidades de velocidad y se llama **velocidad de fase**. En el caso de la cuerda homogénea la velocidad de fase depende solamente de T_0 y de ρ , pero no de la frecuencia, se dice que es un **caso no dispersivo**. En cambio, cuando la velocidad de fase depende de la frecuencia, se dice que es un **caso dispersivo**.

Ecuaciones de ondas como la (6.32), de la forma

$$\frac{\partial^2 \psi}{\partial t^2} = v^2 \frac{\partial^2 \psi}{\partial x^2}, \quad (6.41)$$

no solamente rigen la evolución del apartamiento del equilibrio de una soga elástica homogénea, sino también la perturbación de magnitudes físicas en fenomenologías muy

distintas a la de la soga, como la presión en un gas (ondas acústicas, donde v depende del tipo de gas y de la presión en el equilibrio) o el campo eléctrico en el vacío (ondas electromagnéticas, donde v es una constante universal, la velocidad de la luz en el vacío).

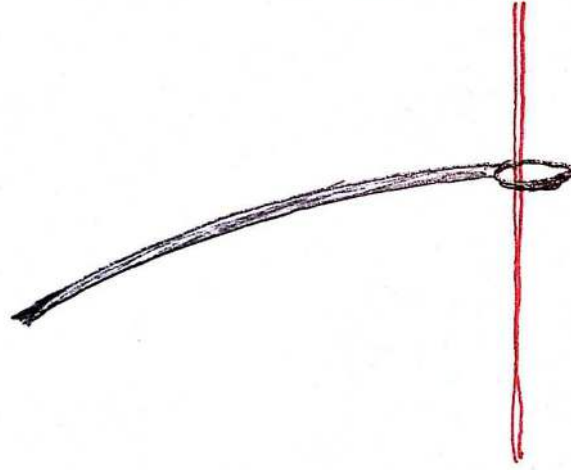


Figura 6.7: Extremo libre: la cuerda con anillo no ejerce fuerza transversal sobre la varilla $\Rightarrow$ cuerda horizontal $\Rightarrow$ pendiente cero.

6.5. Modos normales en el medio continuo

Si en vez de buscar la solución más general de la ec. (6.41) buscamos soluciones que correspondan a modos (es decir, soluciones donde las partes blablabla, blablabla, blablabla) necesariamente tenemos que restringir la búsqueda a funciones de onda de la forma

$$\psi(x, t) = A(x) \cos(\omega t + \varphi), \quad (6.42)$$

donde la función $A(x)$ juega un papel análogo a la sucesión A_n en la versión discreta (ver 4.1.2, página 88). Los modos normales de la soga o de otros medios continuos limitados en el espacio y que responden a la ecuación de ondas clásicas se llaman *ondas estacionarias*. Para encontrar $A(x)$, reemplazamos (6.42) en (6.41) y queda

$$\frac{d^2 A}{dx^2} + \underbrace{\frac{\omega^2}{v^2}}_{k^2} A = 0, \quad (6.43)$$

que es como la ecuación del oscilador armónico, pero en el dominio espacial en vez del temporal. Las soluciones tienen la forma

$$A(x) = B \sin(kx) + C \cos(kx), \quad (6.44)$$

donde B y C son constantes y

$$k = \frac{\omega}{v}. \quad (6.45)$$

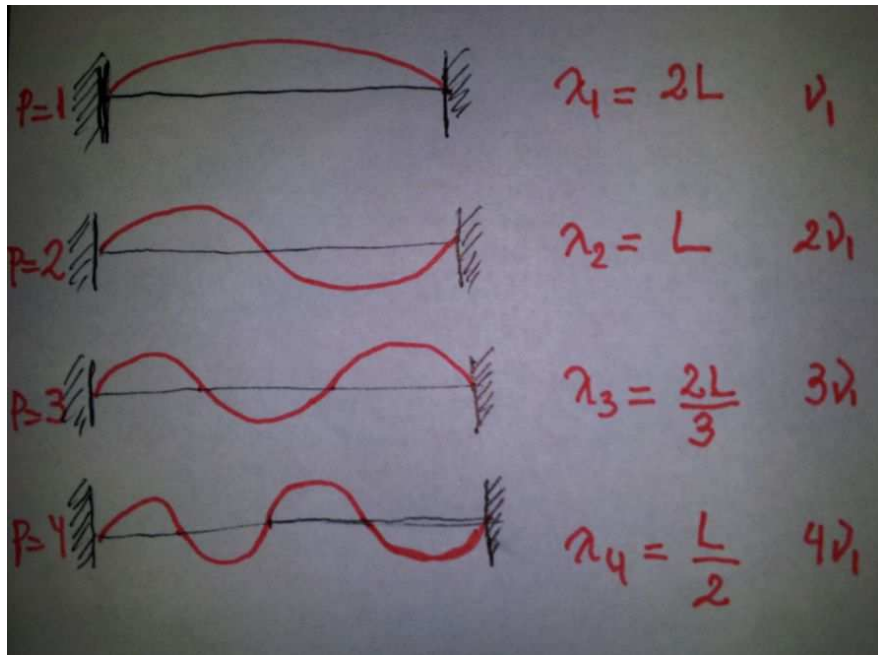


Figura 6.8: Formas de los primeros modos

Comparar (6.42) con la ec. (6.7) y (6.45) con (6.39). Resumen: un modo normal (u onda estacionaria) de la cuerda tiene la forma

$$\psi(x, t) = \left[B \sin(kx) + C \cos(kx) \right] \cos(\omega t + \varphi), \quad (6.46)$$

donde determinaremos k a partir de las condiciones en los extremos ($x = 0, L$) y determinaremos $\omega(k)$ con la relación de dispersión.

6.6. Condiciones en los extremos

En muchos problemas ondulatorios aparecen dos condiciones de extremos que matemáticamente se manifiestan como “función igual a cero” o como “derivada igual a cero”. La condición “función igual a cero” se debe aplicar en el caso de la cuerda para un extremo fijo y es análoga a la condición usada para las amplitudes de la cadena discreta en la sección 6.1.4. Si los dos extremos de la cuerda están fijos, entonces

$$\psi(0, t) = \psi(L, t) = 0. \quad (6.47)$$

En cambio, la condición “derivada igual a cero” aparece en la cuerda homogénea en la situación de extremo libre, que podemos idealizar siguiendo los lineamientos de la sección 6.1.5, pensando que el extremo de la soga está unido a un anillo sin masa que desliza por una varilla vertical sin rozamiento (ver figura 6.9). En estas condiciones, la varilla no puede ejercer fuerza vertical sobre el sistema cuerda-anillo (porque no hay rozamiento). Entonces, el sistema cuerda-anillo no ejerce fuerza transversal sobre la

varilla y la única manera para que esto suceda es que la cuerda quede horizontal, es decir, que su pendiente sea cero y así no hay proyección de la tensión en la dirección vertical. Observar que la misma condición se puede obtener a partir de la ec. (6.21) en el límite de $a \rightarrow 0$. Para una cuerda con dos extremos libres, se debe pedir

$$\frac{\partial \psi(0, t)}{\partial x} = \frac{\partial \psi(L, t)}{\partial x} = 0. \quad (6.48)$$

Por supuesto puede haber situaciones con un extremo fijo y el otro libre. Como en el caso discreto, las condiciones en los extremos no cambian la relación entre ω y k , solamente seleccionan los valores de k permitidos.

6.7. Cuerda con extremos fijos

6.7.1. Discretización de los modos normales

Para una cuerda uniforme de longitud L con extremos fijos en $x = 0$ y en $x = L$, se debe pedir que

$$C = 0 \quad \text{y} \quad B \sin(kL) = 0, \quad (6.49)$$

por lo tanto

$$kL = p\pi, \quad p = 1, 2, \dots, \infty \quad (6.50)$$

y se discretizan los valores de k y las frecuencias de cada modo

$$k_p = p \frac{\pi}{L}, \quad \lambda_p = \frac{2L}{p}, \quad \omega_p = p \frac{\pi}{L} \sqrt{\frac{T_0}{\rho}}, \quad p = 1, 2, \dots, \infty. \quad (6.51)$$

El movimiento más general de la cuerda continua para vibraciones transversales es una superposición de todos los modos

$$\psi(x, t) = \sum_{p=1}^{\infty} B_p \sin(k_p x) \cos(\omega_p t + \varphi_p), \quad (6.52)$$

y esperamos que esta expresión siga describiendo el comportamiento físico de un sistema discreto con muchas partes móviles siempre y cuando las condiciones iniciales exciten modos con longitudes de onda $\lambda_p \gg a$.

En la figura 6.8 están dibujadas las amplitudes $A_p(x)$ (ver eq. (6.44)) con k_p y $\omega_p = 2\pi\nu_p$ dados por las ecs. (6.51).

6.7.2. Condiciones iniciales

Imponemos las condiciones iniciales en la expresión de la ec. (6.52)

$$\psi(x, 0) = \sum_{p=1}^{\infty} B_p \sin(k_p x) \cos \varphi_p, \quad (6.53)$$

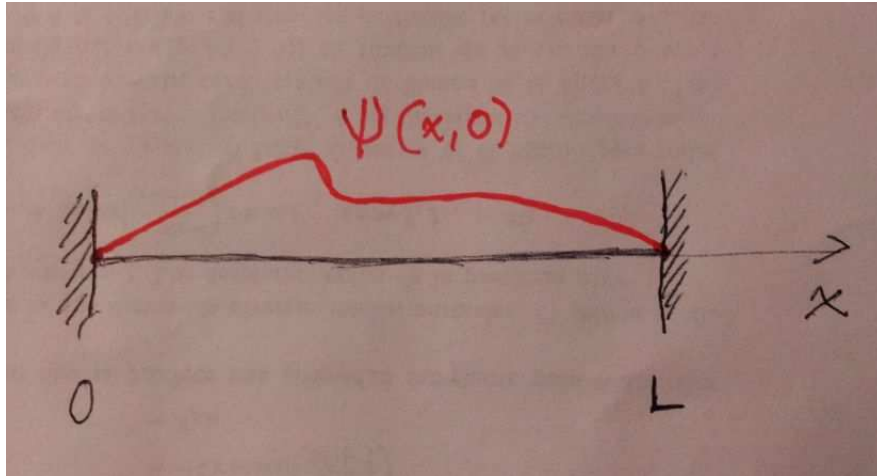


Figura 6.9: Problema de condiciones iniciales, encontrar B_p y φ_p en la ec. (6.51) a partir de $\psi(x, 0)$ y de $\frac{\partial \psi(x, 0)}{\partial t} \forall x \in [0, L]$

y

$$\frac{\partial \psi(x, 0)}{\partial t} = - \sum_{p=1}^{\infty} \omega_p B_p \sin(k_p x) \sin \varphi_p . \quad (6.54)$$

Las funciones a la izquierda de las igualdades (6.53) y (6.54) son conocidas, la primera función es la foto de la forma de la soga en el instante inicial y la otra es la velocidad inicial en cada punto de la soga. ¿Cómo obtener la doble infinitud de amplitudes B_p y fases φ_p ?

Cuando los físicos queremos una solución ya mismo, pero no sabemos matemática ni tenemos tiempo de aprender, recurrimos a métodos FB (llamados así por las iniciales de Fuerza Bruta). Para nuestro problema de determinar B_p y φ_p , un método FB podría proceder de la siguiente manera. Primero, truncar las series y quedarse con 20 términos en cada sumatoria, 20 en (6.53) y 20 en (6.54). Se tendrían así 40 incógnitas, 20 amplitudes B_p y 20 fases φ_p . Y como las igualdades (6.53) y (6.54) deben cumplirse para todo $0 < x < L$, podría forzarse el cumplimiento de estas igualdades en tantos valores de x distintos como hiciera falta. Por ejemplo, en una grilla equiespaciada de 20 puntos $0 < x_q < L$, con $q = 1, 2, \dots, 20$. Se tendría entonces un sistema de 40 ecuaciones con 40 incógnitas. Y con suerte y con una computadora, es probable que se puedan encontrar valores de B_p y φ_p , $p = 1, 2, \dots, 20$ para que las 40 ecuaciones se cumplan lo más ajustadamente posible. Para comprobar si la propuesta FB funciona o no, el proceso se puede repetir truncando las series con 40, en vez de 20 términos, con lo cual ahora habría 80 incógnitas, 40 amplitudes B_p y 40 fases φ_p , que pueden obtenerse numéricamente como antes, pero forzando el cumplimiento de las igualdades (6.53) y (6.54) en una grilla equiespaciada de 40 (en vez de 20) puntos $0 < x_q < L$, con $q = 1, 2, \dots, 40$. Si el proceso FB funciona, las 20 primeras incógnitas B_p y φ_p obtenidas en la segunda truncación, deberían parecerse bastante a las obtenidas en la primera truncación, además de observarse que las últimas 20 obtenidas en la segunda truncación van tomando valores menos significativos. Si esto no ocurre, se puede seguir aumentando el parámetro de truncación y ver si este proceso converge. Así tendríamos un método numérico, que habría que ver bajo que condiciones converge ... todo esto

pertenece al área de la física computacional. Afortunadamente, hay un método más fácil, el método de las series de Fourier, que permite encontrar de manera exacta las infinitas incógnitas B_p y φ_p que aseguran el cumplimiento de (6.53) y (6.54) en todo el intervalo $0 < x < L$. Veremos a continuación el método de las series de Fourier para el caso de la cuerda con extremos fijos, aunque este método se puede aplicar, con pequeñas variantes, para otras soluciones de la ecuación de onda unidimensional y para otras condiciones de extremos.

6.7.3. Series de Fourier

A la derecha de las igualdades (6.53) y (6.54) tenemos dos series que sabemos cuánto dan para todo valor de $x \in [0, L]$ (sobre la sogá): la serie en (6.53) tiene que dar la forma de la sogá en el instante inicial y la serie en (6.54) tiene que dar la velocidad inicial de cada punto de la sogá. Nuestro problema es usar esta información para obtener las amplitudes B_p y las fases φ_p , $p = 1, 2, \dots, \infty$, porque estas cantidades son lo único que nos falta para describir completamente las vibraciones transversales de la cuerda (ver ec. (6.52)).

Algo realmente sorprendente y extrañamente antiintuitivo de la ec. (6.52) es que dice que la forma de la sogá en cualquier instante de tiempo t_0 (en particular también en el instante inicial, como en (6.53))

$$\psi(x, t_0) = \sum_{p=1}^{\infty} \underbrace{B_p \cos(\omega_p t_0 + \varphi_p)}_{F_p} \sin(k_p x), \quad (6.55)$$

se puede escribir como una suma infinita de funciones sinusoidales de amplitudes F_p y con frecuencias espaciales k_p que son múltiplos enteros de la frecuencia espacial fundamental $k_1 = \pi/L$. Puede sonar raro que una función de x , definida para $x \in [0, L]$ y tan arbitraria como puede ser la forma de la cuerda *con extremos fijos* en un instante dado, se pueda escribir de esta manera. A Jean-Baptiste Joseph Fourier también le sonaba raro y gracias a que se salvó de ser guillotinado en la Revolución Francesa fue el primero que pudo demostrar esta propiedad en 1807. Por eso hoy decimos que en la ec. (6.55) estamos haciendo un **análisis de Fourier espacial**. Otra consecuencia sorprendente y antiintuitiva de la ec. (6.52) es que el movimiento más general de un punto arbitrario de la cuerda (por ejemplo, el punto $x_0 \in [0, L]$)

$$\psi(x_0, t) = \sum_{p=1}^{\infty} \underbrace{B_p \sin(k_p x_0)}_{G_p} \cos(\omega_p t + \varphi_p), \quad (6.56)$$

siempre viene dado por una función del tiempo que se puede escribir como una suma infinita de funciones armónicas de amplitudes G_p y frecuencias ω_p que son múltiplos enteros de una frecuencia fundamental $\omega_1 = \pi v/L$ (ver (6.51)). Observar que todos los términos de la serie (6.52) tienen períodos que entran un número entero de veces en el período asociado con la frecuencia fundamental ($T_p = T_1/p$). Esto quiere decir que la función $\psi(x_0, t)$ es periódica, de período T_1 . Y de nuevo, es raro que una función

periódica de período T_1 , tan arbitraria como puede ser el desplazamiento de un punto cualquiera de la cuerda, se pueda escribir como una combinación de funciones armónicas puras, con amplitudes y fases convenientes. De nuevo estamos ante un **análisis de Fourier**, pero ahora **temporal** en vez de espacial.

Por lo discutido en el párrafo anterior, queda claro que para resolver el problema de condiciones iniciales representado por las ecs. (6.53) y (6.54) hay que hacer un análisis de Fourier espacial y encontrar los coeficientes que multiplican a las funciones seno. En la ec. (6.53) estos coeficientes son directamente $B_p \cos \varphi_p$, mientras que en la ec. (6.54) los coeficientes son $-\omega_p B_p \sin \varphi_p$. Si sabemos encontrar estos coeficientes, luego podremos hallar B_p y φ_p tal como hicimos tantas veces (ver (3.9)), mediante el procedimiento de elevar al cuadrado y sumar miembro a miembro las ecuaciones obtenidas (para obtener B_p) y dividir miembro a miembro (para obtener la tangente de φ_p).

Coeficientes de la serie de Fourier impar

Las ecs. (6.53) y (6.54) tienen la forma

$$F(x) = \sum_{p=1}^{\infty} F_p \sin k_p x, \quad 0 \leq x \leq L, \quad (6.57)$$

donde $F(x)$ es dato *y satisface las condiciones de extremos fijos*, es decir, $F(0) = F(L) = 0$. Para encontrar F_p (los coeficientes del desarrollo en serie de Fourier de la función $F(x)$), el procedimiento siempre es el mismo y consiste en multiplicar miembro a miembro la igualdad por una función y luego integrar. Si la función se elige “bien”, al integrar se logra que la serie colapse a un solo término. Las funciones que permiten obtener este efecto tan conveniente son las mismas funciones asociadas a los modos, es decir las funciones $\sin k_p x$, $p = 1, 2, \dots, \infty$. Para ver que esto es así, calculemos las integrales

$$I_{pq} = \int_0^L \sin k_p x \sin k_q x \, dx. \quad (6.58)$$

Usando la identidad

$$2 \sin a \sin b = \cos(a - b) - \cos(a + b),$$

el integrando se puede poner como una suma de cosenos. Y como la integral de cada coseno es un seno, queda

$$I_{pq} = \frac{1}{2} \left[\frac{\sin(k_p - k_q)x}{k_p - k_q} - \frac{\sin(k_p + k_q)x}{k_p + k_q} \right] \Big|_0^L. \quad (6.59)$$

La evaluación en $x = 0$ da cero. Y la evaluación en $x = L$ también, porque según (6.51), $(k_p \pm k_q)L$ es siempre múltiplo entero de π . O sea, que hemos demostrado que $I_{pq} = 0$. ¿Seguro? ¿Y qué pasa con el denominador en (6.59) cuando $k_p = k_q$? Ah, no, en ese caso resulta que I_{pq} no da cero, sino que da $L/2$. En definitiva, hemos demostrado que

$$I_{pq} = \int_0^L \sin k_p x \sin k_q x \, dx = \frac{L}{2} \delta_{pq}, \quad (6.60)$$

con δ_{pq} la delta de Kronecker (igual a 1 si $p = q$, o igual a 0 cuando $p \neq q$). Con este resultado encaramos el procedimiento que mencionamos después de la ec. (6.57): multiplicamos miembro a miembro la igualdad por $\sin k_q x$ y luego integramos en el intervalo $0 \leq x \leq L$. En el lado derecho la integral de la suma es la suma de las integrales (excepto en casos muy patológicos), y entonces queda

$$\int_0^L F(x) \sin k_q x \, dx = \sum_{p=1}^{\infty} F_p I_{pq} = \frac{L}{2} F_q, \quad (6.61)$$

es decir que

$$F_p = \frac{2}{L} \int_0^L F(x) \sin k_p x \, dx, \quad (6.62)$$

para todo valor de $q = 1, 2, \dots, \infty$. Y si todo converge bien, esta “receta” (6.62) nos autorizaría entonces a hacer una identificación entre la función $F(x)$, definida en el intervalo finito $[0, L]$ y nula en los extremos del intervalo (el lado izquierdo de (6.57)), con la serie de funciones sinusoidales de amplitudes F_p y frecuencias espaciales múltiplos enteros de la frecuencia fundamental π/L (el lado derecho de (6.57)). Notar que si bien en nuestro problema físico sólo interesa el valor de $F(x)$ en el intervalo $0 \leq x \leq L$, la serie a la derecha de la igualdad (6.57) puede ser evaluada fuera de este intervalo. Es claro que la serie representa una función periódica, con el período $(2L)$ del armónico

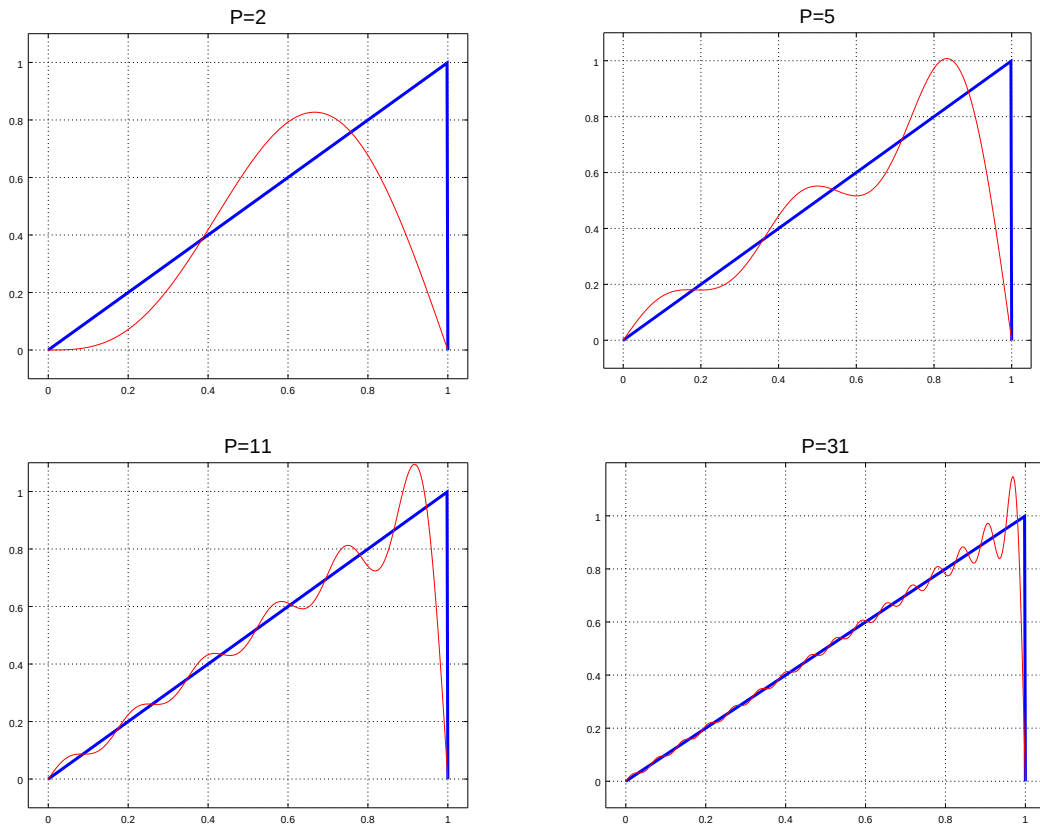


Figura 6.10: La función diente de sierra triangular definida por (6.63) (con $r = 1$ y $L = 1$) y su síntesis parcial sumando solamente los P primeros términos de su serie de Fourier (6.57)

más bajo, ya que

$$\sin \left[k_p \left(x + \frac{2\pi}{k_1} \right) \right] = \sin (k_p x + 2\pi p) = \sin k_p x.$$

¿Qué sucede en el intervalo $[-L, 0]$ (y en todos los intervalos equivalentes, separados del intervalo $[-L, 0]$ por múltiplos de $2L$), donde la información sobre $F(x)$ es irrelevante para el problema físico que estamos tratando? Debido a que los senos son funciones impares, $F(-x) = -F(x)$ y entonces la serie representa una función impar. Y revisando los pasos anteriores, es fácil convencerse que bajo las mismas condiciones de convergencia, toda función periódica impar admite un desarrollo en serie de Fourier dado por (6.61) con amplitudes dadas por (6.62).

Ejemplo: diente de sierra

Una manera infalible de poner a prueba nuestro entendimiento del significado de la igualdad (6.57), es llevarlo a la práctica numérica, por ejemplo comparando una $F(x)$ muy sencilla, definida en el intervalo $[0, L]$, con distintas aproximaciones obtenidas sumando solamente un número finito de términos de la serie de Fourier que corresponde a esta misma función prolongada periódicamente por imparidad al intervalo $[-L, 0]$. Para experimentar numéricamente usemos la función en forma de triángulo que se muestra en color azul en la figura 6.10 y definida por

$$F(x) = \begin{cases} r x & \text{si } 0 \leq x < L, \\ 0 & \text{cuando } x = L. \end{cases} \quad (6.63)$$

Esta función tiene la ventaja de que las integrales (6.62) se obtienen analíticamente de forma muy sencilla, aunque no sería la más adecuada para representar una condición inicial realista (recordemos que fueron las condiciones iniciales las que nos trajeron por este camino). Con cuentas muy fáciles se demuestra que

$$F_q = (-1)^{q+1} \frac{2rL}{q\pi}, \quad q = 1, 2, \dots, \infty. \quad (6.64)$$

En los gráficos de la figura 6.10 se muestran las síntesis parciales obtenidas en el intervalo $[0, L]$, para el caso $r = 1$, $L = 1$ y distintos truncamientos de la serie. Observar que el número de crestas y valles en las líneas rojas da directamente el número de términos empleados y que la amplitud de crestas y valles disminuye a medida que la suma de senos va aproximando mejor a la función. En la figura 6.13 se muestran más detalles sobre la convergencia del cálculo numérico y el comportamiento de la serie fuera del intervalo $[0, L]$. ¿No es asombroso que una función radicalmente tan diferente de una senoide se pueda sintentizar usando solamente funciones sinusoidales?

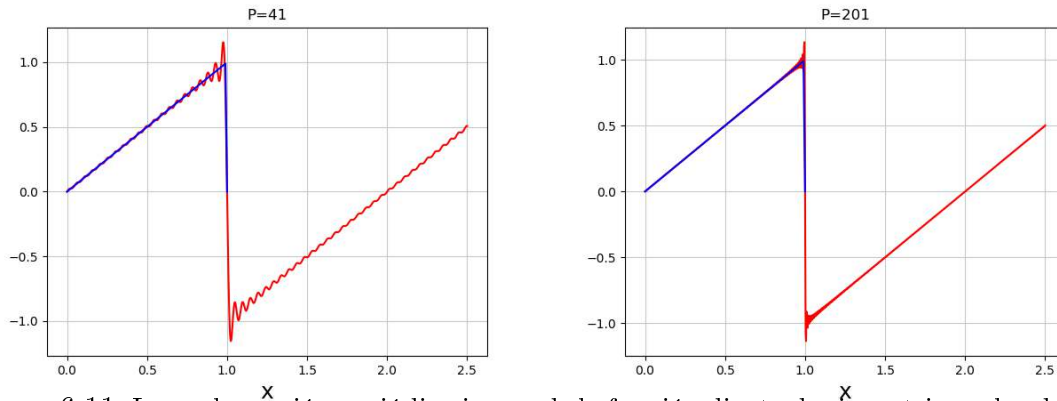


Figura 6.11: La prolongación periódica impar de la función diente de sierra triangular definida por (6.63) (con $r = 1$ y $L = 1$), sumas parciales con 41 y 201 primeros términos en (6.57).

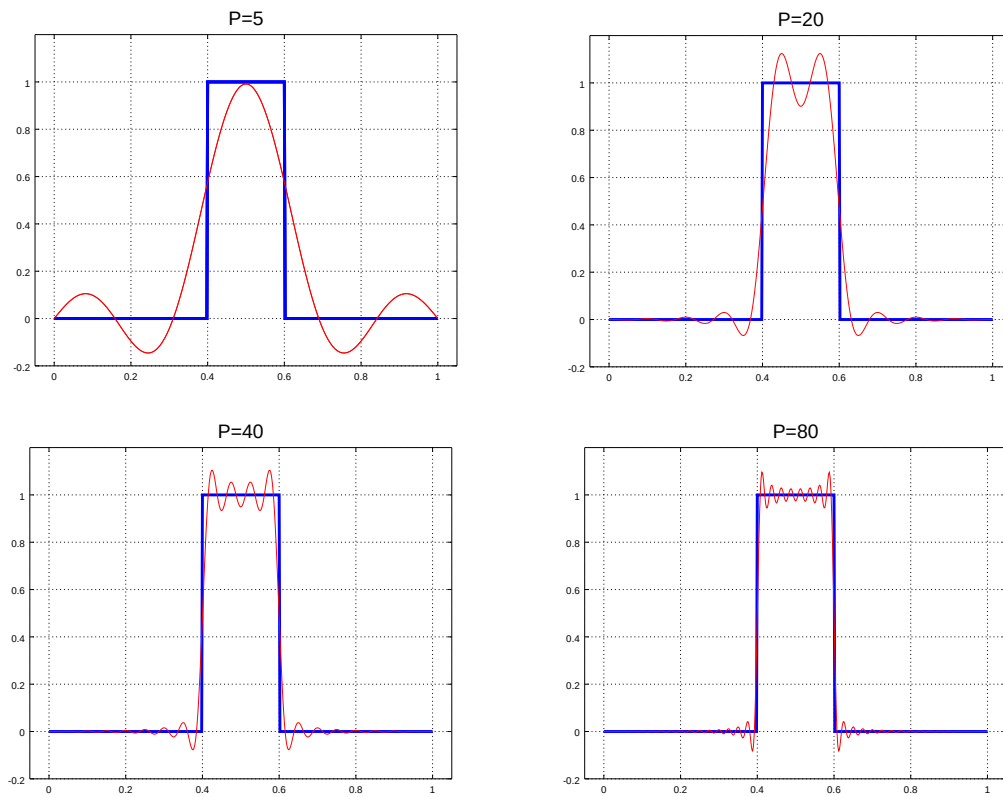


Figura 6.12: El pulso rectangular (6.65) (azul) y su síntesis parcial con P armónicos (rojo).

Ejemplo: pulso rectangular

Como nuevo ejemplo, probemos con un pulso rectangular de ancho c y altura a , localizado entre x_0 y $x_0 + c$ y definida por

$$F(x) = \begin{cases} a & x \in [x_0, x_0 + c], \\ 0 & x \notin [x_0, x_0 + c]. \end{cases} \quad (6.65)$$

Ahora las amplitudes de los armónicos son

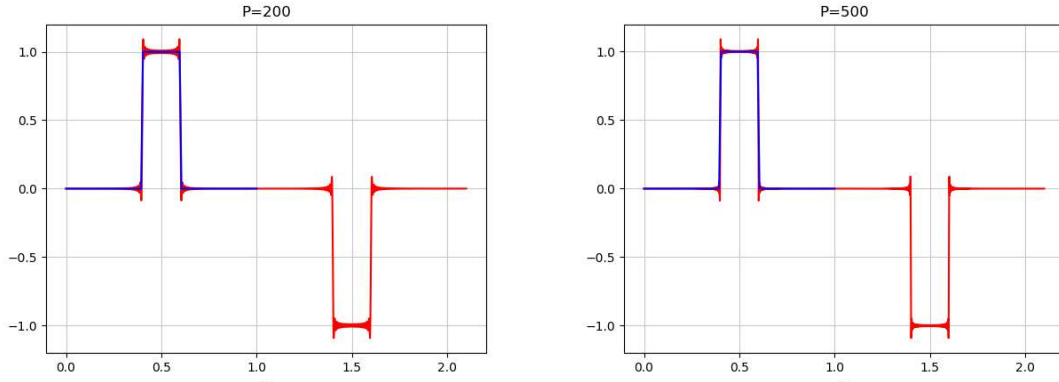


Figura 6.13: La prolongación periódica impar del pulso rectangular definido por (6.65) aproximada por las sumas parciales con 200 y 500 primeros términos en (6.57), parámetros $a = 1$, $x_0 = 0.4$, $c = 0.2$, $L = 1$.

$$F_q = \frac{2a}{q\pi} \left[\cos k_q x_0 - \cos k_q (x_0 + c) \right], \quad q = 1, 2, \dots, \infty. \quad (6.66)$$

En la figura 6.12 se muestran síntesis parciales obtenidas con distintos truncamientos en el caso $a = 1$, $x_0 = 0.4$, $c = 0.2$ y $L = 1$. El script de Python empleado para obtener las figuras (6.10) y (6.13) ⁴ se obtuvo haciendo pequeñas modificaciones al script de la página ??, de manera de incluir la nueva función, (6.65), y los nuevos coeficientes, (6.66).

Condiciones iniciales $\implies B_p$ y φ_p

Ahora ya estamos en condiciones de resolver completamente el problema de condiciones iniciales representado por las ecs. (6.53) y (6.54). Para eso recordemos que las dos series tienen la forma de la ec. (6.57) y que entonces sus amplitudes de Fourier F_p están determinadas por la expresión (6.62). De esta manera, resulta que

$$\begin{aligned} B_p \cos \varphi_p &= \frac{2}{L} \int_0^L \psi(x, 0) \sin k_p x \, dx, \\ -\omega_p B_p \sin \varphi_p &= \frac{2}{L} \int_0^L \frac{\partial \psi(x, 0)}{\partial t} \sin k_p x \, dx. \end{aligned} \quad (6.67)$$

Así obtenemos B_p

$$B_p = \frac{2}{L} \left\{ \left[\int_0^L \psi(x, 0) \sin k_p x \, dx \right]^2 + \frac{1}{\omega_p^2} \left[\int_0^L \frac{\partial \psi(x, 0)}{\partial t} \sin k_p x \, dx \right]^2 \right\}^{1/2}, \quad (6.68)$$

⁴hay una copia en <http://bit.ly/Fourier0102>

y $\tan \varphi_p$

$$\tan \varphi_p = -\frac{1}{\omega_p} \frac{\int_0^L \frac{\partial \psi(x, 0)}{\partial t} \sin k_p x \, dx}{\int_0^L \psi(x, 0) \sin k_p x \, dx}, \quad (6.69)$$

y como las ecs. (6.53) y (6.54) dan los signos de los senos y cosenos, las fases φ_p queda bien determinadas en el intervalo $[0, 2\pi]$.

Notar que:

- Si la velocidad inicial de todos los puntos de la sog a es cero, $\varphi_p = 0 \, \forall p$;
- Si la velocidad inicial de todos los puntos de la sog a es cero y si la forma inicial tiene la forma de uno de los modos (ver figura 6.8), la sog a permanece oscilando en dicho modo;
- Si la velocidad inicial de todos los puntos de la sog a es cero y si la forma inicial tiene un **nodo** (un punto de desplazamiento nulo) que coincide con el nodo de un modo, la sog a permanece oscilando en una superposición de modos que también tengan un nodo en el mismo lugar. Ejemplo de esta situación son los “sonidos armónicos” de la guitarra y otros instrumentos de cuerda, que se obtienen cuando el ejecutante fuerza un nodo en la deformación inicial
- Análogamente, cuando la forma inicial tiene desplazamiento no nulo en el lugar que corresponde al nodo de un armónico, el efecto resultante es suprimir dicho armónico. Por este motivo, los pianos acústicos se diseñan para que los martillos que golpean las cuerdas peguen aproximadamente a una distancia que sea la séptima parte de la longitud de la cuerda. De esta manera, logran atenuar muchísimo la amplitud del séptimo armónico, cuyo sonido no es “agradable” (No entramos a discutir qué se entiende por agradable. Nos pueden decir que significa “disonante”. Pero tampoco vamos a entrar a discutir qué se entiende por disonante).

6.7.4. Evolución temporal

Una vez hallados B_p y φ_p ($p = 1, 2, \dots, \infty$), el movimiento de la sog a queda determinado para todo instante. Por ejemplo, en la figura 6.14 se muestran imágenes de los primeros instantes de la evolución de la deformación para una sog a con $L = 1$ m y condición inicial en forma de pulso rectangular, como en (6.65), con $a = 1$ cm y con $x_0 = 0.45$ m y $c = 0.1$ m (pulso centrado en la mitad de la sog a). La densidad y la tensión se eligen para que la velocidad de fase sea $v = 1$ cm/s y la velocidad inicial de todos los puntos de la sog a es nula. En este caso todos los φ_p son nulos y los B_p están dados por (6.66). Para seguir la deformación de la sog a tenemos que usar (6.52), truncando la serie en un número finito de términos N_s . Para evaluar esta serie en $t \neq 0$ modificamos el script usado para evaluar la serie en $t = 0$.

En la figura 6.14 se observa que la deformación, localizada inicialmente alrededor de $x_0 = 0.5$ m, se empieza a “desparramar” y se propaga hacia las zonas vecinas. En $t = 5$ s la deformación se “ensanchó”, pero se hizo menos intensa en cada punto (bajó su amplitud).

Para $5\text{ s} < t < 50\text{ s}$, ver figura 6.15, la deformación se sigue extendiendo a las zonas vecinas, en la forma de dos paquetes que parecen muy independientes. Si alguien nos preguntara la razón de este comportamiento, podríamos decir que porque las cuentas dan así. ¡Pero seguramente no nos quedaríamos conformes! Sería atrayente poder explicar cosas sin tener que seguir sumando la serie en cada instante. ¿Podríamos predecir la distancia entre los centros de los dos paquetes para cada instante? Estamos en un punto crítico, porque con estos ejemplos queda claro que para entender los comportamientos ilustrados en estas figuras, la descripción que venimos haciendo con los modos normales no parece muy adecuada. Para hablar del rasgueo de la guitarra parecía que los modos normales estaban muy bien, pero acá vemos que no explica claramente lo que sucede con estos paquetitos. Si bien las cuentas dan, nos gustaría descubrir los mecanismos físicos que nos ayuden a comprender mejor lo que estamos viendo.

Llegados a este punto, conviene reflexionar y darse cuenta que estamos cumpliendo el plan que nos habíamos propuesto inicialmente: partiendo de sistemas mecánicos con pocas partes móviles, fuimos agregando partes móviles y en el límite continuo llegamos a una ecuación llamada “de ondas”. Y aunque todavía faltan cosas, el ejemplo que acabamos de ver muestra claramente que estamos en presencia de perturbaciones que se propagan, es decir, estamos poniendo en evidencia los comportamientos ondulatorios. La evolución de la deformación en este ejemplo se puede ver en un gif animado (fácil de hacer con Python) en este enlace <http://bit.ly/1LOZxY0>.

¿Qué sentido físico tienen en esta simulación tiempos “especiales”, como $t = 5$ s, donde la altura del pulso disminuyó a la mitad y su ancho se duplicó respecto al pulso original? ¿O $t = 50$ s (ver figura 6.16), donde la deformación de la sogá es nula en todo punto? ¿Estamos en condiciones de encontrar otros tiempos críticos y de relacionarlos con mecanismos sencillos, que se puedan explicar sin tener que sumar una serie? A continuación veremos que estos tiempos “especiales” están relacionados con mecanismos ondulatorios escondidos detrás de los modos normales y que estos mecanismos se pueden poner en evidencia sin necesidad de sumar las series, no solamente en el ejemplo que estamos viendo de sogá con extremos fijos, sino también en cualquier otro sistema unidimensional gobernado por la ecuación de ondas clásicas (6.32).

6.8. ondas estacionarias = suma de ondas viajeras

Según (6.46) (página 104), la onda estacionaria de la cuerda elástica homogénea, o de cualquier otro sistema continuo gobernado por (6.32), tiene la forma

$$\psi(x, t) = \underbrace{\left[B \sin(kx) + C \cos(kx) \right]}_{D \sin(kx + \alpha)} \cos(\omega t + \varphi), \quad (6.70)$$

con $k = \omega/v$. Este tipo de solución puede representar un modo normal como en la ec. (6.46), en cuyo caso ω queda determinada por las condiciones en los extremos. O también puede representar la respuesta forzada estacionaria de un sistema impulsado armónicamente desde alguno de sus extremos, en cuyo caso ω es impuesta por la fuerza impulsora.

Usando identidades trigonométricas, el producto de un seno por un coseno puede transformarse en una suma de dos senos y así la onda estacionaria puede reescribirse como

$$\psi(x, t) = \frac{D}{2} \left[\sin(\underbrace{kx - \omega t + \alpha - \varphi}_{\text{fase1}}) + \sin(\underbrace{kx + \omega t + \alpha + \varphi}_{\text{fase2}}) \right]. \quad (6.71)$$

Prestemos atención al primer seno, el que depende de la variable fase1: vale cero cuando fase1 es cero, vale 0.5 cuando fase1 es $\pi/6$, vale uno cuando fase1 es $\pi/2$, etc. Supogamos que estoy obsesionado en moverme con esta senoide siempre al lado del valor 0.5. Para asegurarme de estar en el lugar x y en el momento t adecuados lo único que tengo que pedir es que el argumento fase1 valga siempre $\pi/6$. Porque si me quedo parado en donde estoy voy a perder el valor 0.5, porque el argumento fase1 cambia a medida que pasa el tiempo. Dicho de otra manera, si en el instante $t + \Delta t$ quiero seguir al lado del valor 0.5 del seno, entonces tendría que moverme a una posición $x + \Delta x$ que me asegure de que fase1 continúa valiendo $\pi/6$. Es decir, que los valores Δx y Δt no son arbitrarios: tienen que preservar el valor de **fase1** que me daba 0.5 para el seno, o sea

$$\begin{aligned} kx - \omega t + \alpha - \varphi &= k(x + \Delta x) - \omega(t + \Delta t) + \alpha - \varphi = \pi/6, \\ \implies k \Delta x - \omega \Delta t &= 0, \\ \implies \frac{\Delta x}{\Delta t} &= \frac{\omega}{k} = v. \end{aligned} \quad (6.72)$$

En resumen: para seguir al valor 0.5 (o a cualquier otro valor del primer seno en el que estemos interesados), tenemos que mantener fase1 constante y para eso hay que moverse hacia mayores valores de x con una velocidad igual a la velocidad v que aparece en la ecuación de ondas. Ahora queda claro el nombre de **velocidad de fase** para v y el nombre **onda progresiva** para el primer seno en la ec. (6.71) y para cualquier otra función que dependa de la combinación $x - vt$. Un análisis similar para el segundo seno en (6.71) muestra que para seguir un valor determinado de la función que depende del argumento fase2, es necesario desplazarse hacia menores valores de x a medida que pasa el tiempo, es decir hay que desplazarse con velocidad $-v$. Por eso el nombre de **onda regresiva** para el segundo seno en la ec. (6.71) y para cualquier otra función que dependa de la combinación $x + vt$.

Si estamos interesados en cómo evoluciona la deformación inicial del sistema, usamos la expresión general, con modos de la forma (6.70), análoga a (6.52), pero con $C \neq 0$. En términos de la representación progresiva-regresiva (6.71), la evolución se escribe

$$\psi(x, t) = \frac{1}{2} \sum_{p=1}^{\infty} D_p \sin(k_p x - \omega_p t + \alpha_p - \varphi_p) + \frac{1}{2} \sum_{p=1}^{\infty} D_p \sin(k_p x + \omega_p t + \alpha_p + \varphi_p). \quad (6.73)$$

Apliquemos este resultado al ejemplo de las figuras 6.14, 6.15 y 6.16. Como en este caso la velocidad inicial de todos los puntos de la soga es nula, $\varphi_p = 0$ para todo valor de p y recordando que $\omega_p = vk_p$, resulta

$$\psi(x, t) = \frac{1}{2} \sum_{p=1}^{\infty} D_p \sin [k_p (x - vt) + \alpha_p] + \frac{1}{2} \sum_{p=1}^{\infty} D_p \sin [k_p (x + vt) + \alpha_p], \quad (6.74)$$

para todo $0 \leq x \leq L$. En esta representación progresiva-regresiva, la condición inicial se expresa como suma de dos series idénticas

$$\psi(x, 0) = \frac{1}{2} \sum_{p=1}^{\infty} D_p \sin [k_p x + \alpha_p] + \frac{1}{2} \sum_{p=1}^{\infty} D_p \sin [k_p x + \alpha_p], \quad (6.75)$$

de donde deducimos el resultado, bastante trivial, de que cada serie en (6.75) representa a la función $\psi(x, 0)/2$.

La cosa empieza a parecer menos trivial cuando usamos esta información para reinterpretar la representación (6.74) de $\psi(x, t)$ y observamos que la primera sumatoria no es otra cosa que la deformación inicial dividida por el factor 2, pero evaluada en $x - vt$, es decir, desplazada vt hacia la derecha. Y que análogamente, la segunda sumatoria en (6.74) no es otra cosa que la deformación inicial dividida por el factor 2, pero ahora evaluada en $x + vt$, es decir, desplazada vt hacia la izquierda.

Esta separación en ondas progresivas y regresivas nos brinda otra manera de interpretar los resultados que hasta ahora obteníamos sumando ondas estacionarias, es decir, modos. Por ejemplo, consideremos el instante $t = 5$ s, donde se ha duplicado el ancho del pulso (ver figura 6.14). Según (6.74), la deformación en este instante se puede interpretar como resultado de un desplazamiento $vt = 5$ cm, del pulso progresivo hacia la derecha y del pulso regresivo hacia la izquierda. Si en $t=0$ ambos pulsos ocupaban la misma región $0.45 \text{ cm} \leq x \leq 0.55 \text{ cm}$, entonces en $t = 5$ s el pulso regresivo ocupa la región $0.4 \text{ cm} \leq x \leq 0.5 \text{ cm}$, mientras que el pulso progresivo ocupa la región $0.5 \text{ cm} \leq x \leq 0.6 \text{ cm}$. Vemos así que la suma de los dos términos en (6.74) ocupan una región de longitud doble que la región ocupada inicialmente por la deformación y que la altura de la deformación resultante es la mitad de la altura de la deformación inicial.

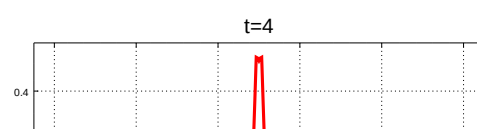
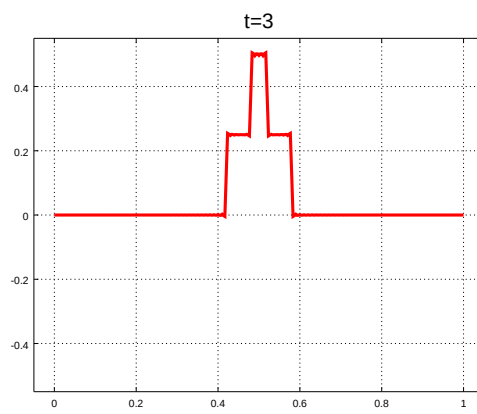
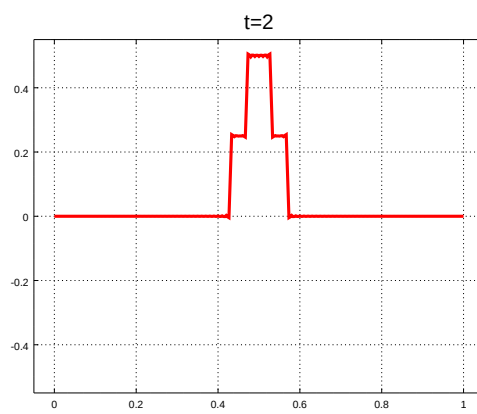
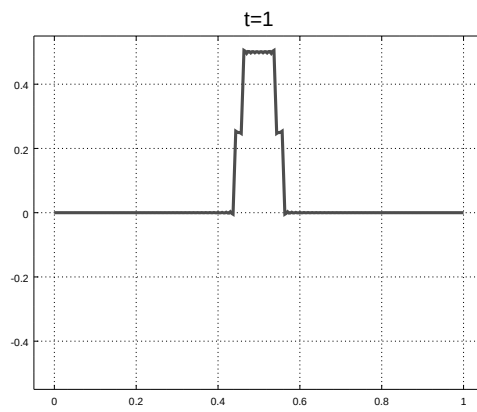
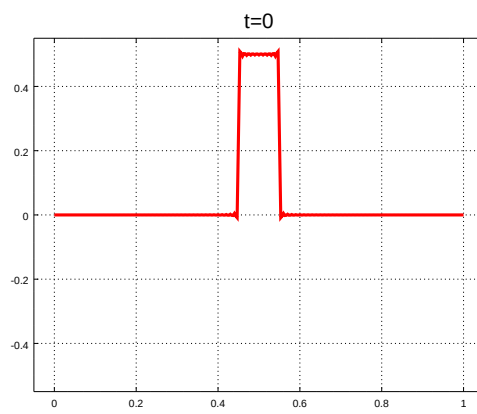
La separación progresiva-regresiva de (6.73) nos provee de una explicación alternativa para los resultados de la figura 6.14, que hasta ahora solamente podíamos explicar “porque las cuentas dan así”. Veremos más adelante (7.1.3) que la separación es válida no solamente para el caso considerado en el ejemplo, de velocidad inicial nula, sino para cualquier solución de la ecuación de ondas clásica 1D.

6.9. Y en otros sistemas ...

Si la soga tiene otras condiciones en los extremos, pasa algo similar a lo que hemos visto, cambia la forma de los modos, pero los modos siguen siendo “ortogonales” entre sí, en el sentido de que sus integrales son deltas de Kronecker, tal como ocurre para

sogas con extremos fijos, ver ec. (6.60). Esta característica se repite para otros sistemas que pueden ser muy distintos a la sogá elástica, pero que también se extienden en una dirección fija del espacio, están limitados espacialmente (empiezan y terminan) y responden a la ecuación de ondas clásica unidimensional.

En la práctica se verán ejemplos no solamente de sogas con otras condiciones de contorno, sino también de otros sistemas muy distintos, como tubos con gases, que exhiben comportamientos (y admiten tratamientos) completamente análogos a los que vimos para la sogá con extremos fijos. Es importante observar que en todos estos sistemas resulta que la forma de los modos permitidos solo depende de la forma de las ecuaciones de movimiento (que terminan siendo siempre la misma, la ecuación de ondas) y además los modos adoptan formas particularmente sencillas, por ejemplo, siempre aparece una relación supersencilla entre la longitud de onda del modo y la longitud total de sistema y esta relación no tiene nada que ver con magnitudes como elasticidad, inercia, densidad, temperatura, presión, etc.



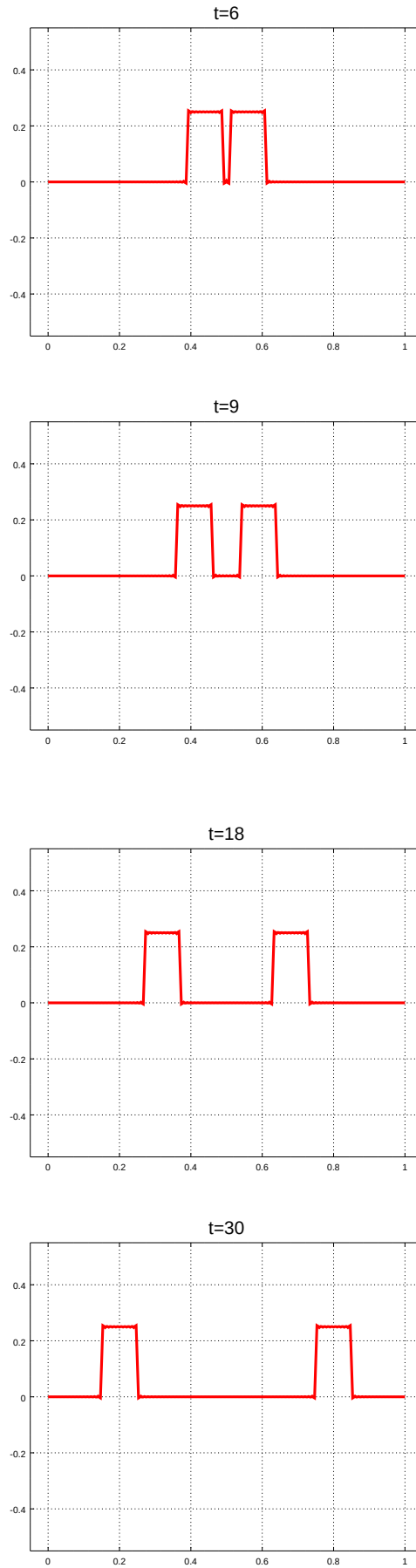


Figura 6.15: $\psi(x, t)$ para $5s < t < 50s$. La deformación inicial tiene la forma de onda cuadrada indicada en la eq. (6.65), con $x_0 = 0.45$ m, $c = 0.1$ m, $a = 1$ cm y $v = 1$ cm/s. La velocidad

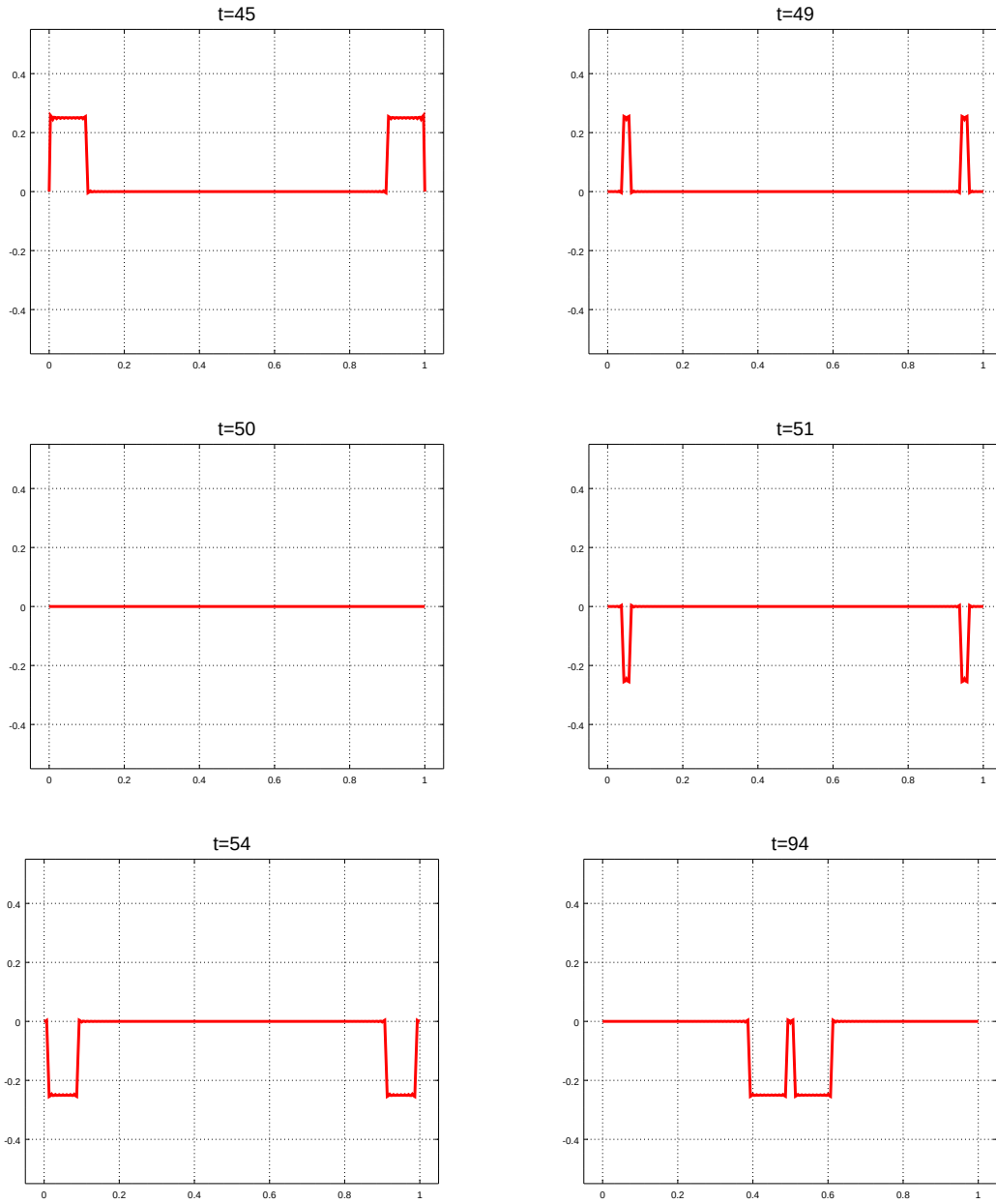


Figura 6.16: $\psi(x, t)$ para $45s < t < 94s$. La deformación inicial tiene la forma de onda cuadrada indicada en la eq. (6.65), con $x_0 = 0.45$ m, $c = 0.1$ m, $a = 1$ cm y $v = 1$ cm/s. La velocidad inicial de todos los puntos de la soga es nula.

Capítulo 7

Ondas propagantes

7.1. Ecuación clásica 1D

7.1.1. Progresivas y regresivas 1D

Según las ecs. (6.70) y (6.71), toda onda estacionaria armónica es superposición de dos ondas viajeras, una progresiva y otra regresiva. Veremos a continuación una propiedad de alguna manera inversa, aunque bastante más general porque no está restringida a dependencias armónicas: si una magnitud física $\psi(x, t)$ depende solamente de la combinación $\eta_1 = x - vt$, entonces $\psi(x, t) = f(\eta_1)$ es solución de la ecuación de ondas (6.41). Análogamente, si la magnitud física $\psi(x, t)$ depende solamente de la combinación $\eta_2 = x + vt$, entonces $\psi(x, t) = g(\eta_2)$ también es solución de la ecuación de ondas (6.41). Dicho de otra manera, veremos que toda función de una sola variable, dos veces derivable, puede representar una posible onda clásica progresiva, cuando su argumento se evalúa en $x - vt$, o regresiva, cuando su argumento se evalúa en $x + vt$ (el carácter progresivo o regresivo está referido a una dirección fija en el espacio, que sin pérdida de generalidad podemos considerar como eje x).

Supongamos primero que $f(\eta_1)$ es una función de la única variable η_1 . Para fijar ideas, asociamos $f(\eta_1)$ con su representación gráfica, f en el eje vertical versus η_1 en el eje horizontal. Sea $\eta_1 = x - vt$: es claro que $f(x - vt)$ representa una función progresiva, porque el gráfico de $f(x - vt)$ vs x para $t > 0$ está desplazado una distancia vt en el sentido positivo del eje x con respecto al gráfico de $f(x - vt)$ vs x a $t = 0$. Notemos que esto es así siempre, independientemente de la forma que tenga $f(\eta_1)$, no importa si se trata de un pulso rectangular, una senoide o una función gaussiana. Y con tanta generalidad, no parecería haber argumentos físicos para asegurar que la función $\psi(x, t)$ elegida idénticamente igual a $f(x - vt)$ sea obligatoriamente solución de la ecuación de ondas clásica 1D (6.41). Sin embargo, lo es, como se puede ver fácilmente teniendo en cuenta que con esta elección

$$\frac{\partial^2 \psi(x, t)}{\partial x^2} = f'' \quad \text{y} \quad \frac{\partial^2 \psi(x, t)}{\partial t^2} = v^2 f'' , \quad (7.1)$$

y entonces se cumple

$$\frac{\partial^2 \psi}{\partial t^2} = v^2 \frac{\partial^2 \psi}{\partial x^2}.$$

De manera completamente análoga podemos mostrar que la función $\psi(x, t)$ elegida idénticamente igual a $g(x + vt)$, con g función de una única variable (η_2), también satisface automáticamente la ecuación de ondas clásica 1D (6.41), puesto que

$$\frac{\partial^2 \psi(x, t)}{\partial x^2} = g'' \quad \text{y} \quad \frac{\partial^2 \psi(x, t)}{\partial t^2} = v^2 g''. \quad (7.2)$$

Ahora, $\psi(x, t) = g(x + vt)$ representa una función regresiva, porque el gráfico de $g(x + vt)$ vs x para $t > 0$ está desplazado una distancia vt en el sentido negativo del eje x con respecto al gráfico de $g(x + vt)$ vs x a $t = 0$.

7.1.2. La solución más general 1D

Hemos visto que las soluciones progresivas y regresivas son siempre solución de la ecuación de ondas clásica 1D. Y como esta ecuación es lineal, resulta que cualquier suma de funciones progresivas es también solución. Ahora vamos por más: veremos a continuación que **la solución más general** de la ec. (6.41) es **siempre** combinación lineal de una solución progresiva y de una regresiva. Con este fin, pasamos a las variables η_1 y η_2 , respectivamente asociadas a comportamientos progresivos y regresivos

$$\begin{aligned} \eta_1 &= x - vt \\ \eta_2 &= x + vt \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} x &= (\eta_1 + \eta_2)/2 \\ t &= (\eta_2 - \eta_1)/(2v). \end{aligned} \quad (7.3)$$

Para ver cómo se escribe (6.41) en las nuevas variables, usamos la regla de la cadena una vez para transformar las derivadas primeras

$$\frac{\partial \psi(x, t)}{\partial x} = \frac{\partial \psi(x, t)}{\partial \eta_1} \frac{\partial \eta_1}{\partial x} + \frac{\partial \psi(x, t)}{\partial \eta_2} \frac{\partial \eta_2}{\partial x}, \quad (7.4)$$

$$\frac{\partial \psi(x, t)}{\partial t} = \frac{\partial \psi(x, t)}{\partial \eta_1} \frac{\partial \eta_1}{\partial t} + \frac{\partial \psi(x, t)}{\partial \eta_2} \frac{\partial \eta_2}{\partial t}, \quad (7.5)$$

y otra vez más para transformar las derivadas segundas

$$\frac{\partial^2 \psi(x, t)}{\partial x^2} = \frac{\partial^2 \psi(x, t)}{\partial \eta_1^2} + \frac{\partial^2 \psi(x, t)}{\partial \eta_2^2} + 2 \frac{\partial^2 \psi(x, t)}{\partial \eta_1 \partial \eta_2}, \quad (7.6)$$

$$\frac{\partial^2 \psi(x, t)}{\partial t^2} = v^2 \left[\frac{\partial^2 \psi(x, t)}{\partial \eta_1^2} + \frac{\partial^2 \psi(x, t)}{\partial \eta_2^2} - 2 \frac{\partial^2 \psi(x, t)}{\partial \eta_1 \partial \eta_2} \right]. \quad (7.7)$$

Vemos así que en las nuevas variables η_1 y η_2 , (6.41) se escribe de una manera muy sencilla

$$\frac{\partial^2 \psi}{\partial \eta_2 \partial \eta_1} = 0, \quad (7.8)$$

que dice que la primera derivada parcial con respecto a η_1 no depende de η_2 . Por lo tanto, la primera derivada parcial con respecto a η_1 es solamente una función F de η_1

$$\frac{\partial \psi}{\partial \eta_1} = F(\eta_1). \quad (7.9)$$

La solución de esta ecuación diferencial en derivadas parciales es muy fácil, la función incógnita ψ es la primitiva de la función $F(\eta_1)$ (llamémosla $f(\eta_1)$), más una constante que puede depender de η_2 (llamémosla $g(\eta_2)$), es decir que hemos demostrado que la solución más general de la ec. (6.41) es

$$\psi(x, t) = f(x - vt) + g(x + vt), \quad (7.10)$$

que es lo que queríamos demostrar.

Si en un medio continuo infinito hubiera una única onda progresiva, seguiría propagándose sin modificaciones. Pero si la onda progresiva se encontrara con una discontinuidad, necesariamente debería aparecer una onda regresiva, para poder ajustar las condiciones de contorno en la discontinuidad. Por este motivo en sistemas limitados en el espacio, como en la soga con extremos fijos, sobreviven sólo ondas estacionarias, que pueden ser interpretadas como superposición de ondas viajeras con amplitudes y periodicidades espaciales adecuadas para ajustar las condiciones en ambos extremos.

7.1.3. Fórmula de d'Alembert

Según (6.70) y (6.71), toda onda estacionaria es superposición de dos ondas viajeras, una progresiva y otra regresiva. Y según (7.10), toda onda unidimensional que depende de una coordenada fija en el espacio es siempre combinación lineal de dos ondas: una progresiva y otra regresiva. Si bien el resultado (7.10) es muy fuerte, porque no toda función de dos variables se escribe como suma de dos funciones de una sola variable, su demostración no dice nada sobre la forma que tienen las funciones f y g .

En la sección 6.8, cuando interpretamos en términos de ondas progresivas y regresivas la simulación de la cuerda con extremos fijos y condiciones iniciales de onda cuadrada, quedaba claro que eran las condiciones iniciales $\psi(x, 0)$ y $\partial\psi(x, 0)/\partial t$ las que fijaban la forma de las funciones f y g . En dicha simulación (ver figuras (6.14), (6.15) y (6.16)), la velocidad inicial de todos los puntos de la soga era nula y con esta condición, y reescribiendo los modos como combinación de ondas progresivas y regresivas, resultó fácil encontrar que las funciones f y g eran pulsitos cuadrados, idénticos entre sí e idénticos a la forma inicial, pero con altura mitad. También vimos en la sección 6.8 que con la separación progresivas-regresivas, en vez de usar directamente los modos, era posible explicar muy satisfactoriamente la evolución del apartamiento del equilibrio en base al desplazamiento de los pulsitos. ¿Será siempre posible encontrar las funciones f y g para condiciones iniciales arbitrarias? La respuesta es positiva y veremos a continuación la demostración dada en 1747 por Jean le Rond d'Alembert, uno de los grandes genios del Iluminismo.

El procedimiento de d'Alembert parte de imponer las condiciones iniciales en la solución general dada por la ec. (7.10), primero sobre la posición inicial

$$\psi(x, 0) = f(x) + g(x), \quad (7.11)$$

y luego sobre la velocidad inicial

$$\frac{\partial \psi(x, 0)}{\partial t} = v \left[-f'(x) + g'(x) \right]. \quad (7.12)$$

Integrando miembro a miembro la última ecuación, queda

$$h(x) = v \left[-f(x) + g(x) \right], \quad (7.13)$$

donde $h(x)$ es una primitiva de la función de x que da la velocidad inicial de la soga

$$h(x) = \int \frac{\partial \psi(x, 0)}{\partial t} dx + C, \quad (7.14)$$

definida a menos de una constante C . Despejando $f(x)$ y $g(x)$ del sistema de ecs. (7.11) y (7.14), se obtiene

$$f(x) = \frac{1}{2} \left[\psi(x, 0) - \frac{1}{v} h(x) \right], \quad (7.15)$$

$$g(x) = \frac{1}{2} \left[\psi(x, 0) + \frac{1}{v} h(x) \right]. \quad (7.16)$$

Las ecs. (7.15) y (7.16) dan la forma funcional de las funciones de una variable asociadas con las partes progresiva y regresiva de la solución más general de la ecuación de ondas. Para obtener la dependencia funcional correcta de la función de onda con las variables x y t usamos (7.10):

$$\psi(x, t) = \frac{1}{2} \left[\psi(x - vt, 0) + \psi(x + vt, 0) \right] + \frac{1}{2v} \left[h(x + vt) - h(x - vt) \right],$$

que si se usa (7.14) se puede reescribir como

$$\psi(x, t) = \frac{1}{2} \left[\psi(x - vt, 0) + \psi(x + vt, 0) \right] + \frac{1}{2v} \int_{x-vt}^{x+vt} \frac{\partial \psi(x, 0)}{\partial t} dx. \quad (7.17)$$

Notar que cuando la velocidad inicial de todos los puntos de la soga es cero, se anula el segundo término de (7.17) y el resultado se reduce al obtenido en (6.75).

A diferencia de la demostración en 6.8, la demostración de d'Alembert no hace referencia a dependencias temporales armónicas ni a series de Fourier. A pesar de ser descripciones equivalentes, es obvio que la descripción de la evolución de una cuerda elástica *à la d'Alembert* (ec. (7.17)) es muy diferente a la descripción *à la Fourier* (ecs. (6.52), (6.68) y (6.69)). En Física se habla de *dualidad* para describir modelos teóricos que parecen ser diferentes pero que son equivalentes, porque predicen exactamente las

mismas propiedades físicas. En los ejemplos “triviales”, dos teorías parecen ser diferentes debido sólo al modo en que se presentan. Sería como si las teorías se enunciaran en español y en chino: un físico que dominara ambas lenguas podría realizar fácilmente una traducción entre una y otra lengua y demostraría así su equivalencia. En un caso trivial, con la traducción no se gana nada desde el punto de vista de la Física, un cambio de lenguaje no aporta nuevas ideas. En cambio, en los ejemplos no triviales, distintas descripciones de la misma situación física producen ideas físicas y métodos matemáticos de análisis diferentes y complementarios. Es lo que pasa en este caso entre las descripciones de d’Alambert y de modos, cada descripción tiene su ventaja, a diferencia del caso del español y el chino, hay ideas físicas muy importantes que toman protagonismo en cada una de estas descripciones duales, tales como el concepto de frecuencia de resonancia o de λ_p en la descripción de modos, o el concepto de velocidad de propagación o de ondas viajeras en la descripción de d’Alambert. Mientras que la descripción de modos parece ser la más adecuada para explicar el timbre de un instrumento de cuerdas, o efectos como los sonidos armónicos de la guitarra, la descripción de d’Alambert parece ser la más adecuada para explicar la evolución de la deformación en la simulación mostrada en las figuras 6.14 y 6.15). Ambas descripciones dan el mismo resultado físico, de hecho los resultados en las figuras 6.14 y 6.15), a los que se agregan los de la figura 6.16, fueron obtenidos con la descripción de modos, pero para interpretar los “pulsitos”, necesitamos la descripción de d’Alambert. En el curso de F2 veremos otros ejemplos de descripciones duales.

Con respecto a los cuadros que se muestran en la figura 6.16, es interesante prestar atención a las ondas reflejadas en ambos extremos e interpretar la deformación de la cuerda elástica en términos de módulo y fase del coeficiente de reflexión. ¿Cómo cambiaría la evolución de la misma deformación inicial si la cuerda tuviera el extremo derecho libre?

7.2. Ondas, medios y dimensiones

Hasta ahora hemos visto comportamientos físicos que dependen de una dirección fija en el espacio, esencialmente porque la posición de equilibrio de los sistemas estudiados, como las sogas o las cadenas con cuentas, es una dirección fija en el espacio. Pero en otros sistemas donde la posición de equilibrio se describe con más dimensiones, también puede haber comportamientos ondulatorios dependientes solamente de una dirección fija en el espacio, siempre que esta simetría espacial se introduzca mediante los contornos, las deformaciones iniciales o las fuerzas impulsoras. Esto es lo que sucede por ejemplo en tubos de gas, o en canales con líquido, o en recipientes con gases, cuando las perturbaciones se hacen mediante pistones o compuertas planas. Si bien las ondas son unidimensionales, en el sentido de que se pueden describir con una sola variable espacial, es importante notar que son unidimensionales en cartesianas, no en esféricas ni en cilíndricas. Por tratarse de una dirección fija en el espacio, hemos obtenido ecuaciones como la ec. (6.41) o la ec. (??), donde en cada punto la derivada segunda temporal tiene en cuenta la aceleración de la parte móvil mientras que la derivada segunda espacial está relacionada con la fuerza restitutiva sobre dicha parte móvil producida por vecinos

cercanos.

En las redes unidimensionales a lo largo del eje x la fuerza restitutiva producida por vecinos cercanos da lugar a derivadas segundas con respecto a x . Pero el número de vecinos cercanos aumenta con la dimensión del sistema. Es fácil darse cuenta que cuando las partes móviles interactúan en mallas bidimensionales, cada parte móvil no sólo tiene vecinos a lo largo de la dirección x , sino también vecinos en la dirección perpendicular, digamos y . Por este motivo, en el caso 2D la fuerza restitutiva que produce la aceleración tiene que tener dos contribuciones: i) por un lado la contribución de antes, de los vecinos cercanos en x y que da lugar a la derivada segunda con respecto a x ; y ii) la nueva contribución de los vecinos cercanos en y que en el límite continuo dará lugar a una derivada segunda con respecto a y . Si el sistema es isótropo, es decir si x e y son indistinguibles, las dos fuerzas restitutivas se suman con el mismo factor de peso. Por este motivo, no debe sorprendernos que la ecuación de ondas para una membrana elástica (el análogo 2D de la cuerda), tenga la forma

$$\frac{\partial^2 \psi}{\partial t^2} = v^2 \nabla^2 \psi, \quad (7.18)$$

donde ∇^2 indica la suma de las derivadas segundas espaciales $\nabla^2 \psi = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2}$. Un razonamiento similar es válido cuando las partes móviles interactúan en mallas 3D, donde de nuevo se obtiene la ec. (7.18), pero ahora $\nabla^2 \psi = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2}$. Notar que si los resortes a lo largo de x tuvieran distinta elasticidad que los resortes a lo largo de y (un medio anisótropo) las dos fuerzas restitutivas se sumarían con distinto factor de peso y en tal caso no se podría formar el laplaciano a partir de las derivadas segundas.

7.2.1. La onda plana en un medio 3D

En un medio 3D, las soluciones de la ec. (7.18) que dependen de una dirección fija en el espacio tienen que tener la forma de la ec. (7.10), es decir que siempre son combinación lineal de dos ondas: una progresiva y otra regresiva. Para que esto sea así tenemos que poder elegir el sistema de coordenadas de manera tal que el eje x sea *esa dirección fija en el espacio*. Si no tenemos esta libertad, por ejemplo porque conviene elegir los ejes para que se adapten a los contornos, podría ser problemático. Por ejemplo, quiero mirar una perturbación que avanza por la diagonal que va del techo al piso de una habitación con forma de paralelepípedo rectangular, pero sería deseable que los planos $x - y$, $x - z$ e $y - z$ coincidan con las paredes. Sea $\vec{r}$ el vector posición de cada punto del medio 3D y el versor $\hat{n}$ la dirección de esta diagonal. Estamos de acuerdo en que convendría elegir el eje x a lo largo de esta diagonal y en este caso la onda progresiva sería $f(x - vt)$ y no tendríamos que usar notación vectorial con el vector posición $\vec{r}$. Pero cuando no se puede elegir el eje x a lo largo de la dirección fija en el espacio (en este ejemplo, la diagonal), no hay otro remedio que usar $\vec{r}$. En realidad todo se reduce al caso sencillo si notamos que los puntos con igual perturbación son planos perpendiculares al versor $\hat{n}$, es decir que cada plano se caracteriza porque sus puntos tienen la misma proyección del vector $\vec{r}$ sobre $\hat{n}$. Esto equivale a decir que $\hat{n} \cdot \vec{r}$ juega el papel que nos habría gustado que jugara x si se hubiera podido elegir el eje x

a lo largo de la diagonal. Toda esta discusión muestra que la **onda plana** progresiva en un medio 3D se escribe como

$$\psi(\vec{r}, t) = f(\hat{n} \cdot \vec{r} - vt), \quad (7.19)$$

mientras que la regresiva

$$\psi(\vec{r}, t) = g(\hat{n} \cdot \vec{r} + vt). \quad (7.20)$$

La onda plana armónica en un medio 3D

Si existen motivos para afirmar que la onda, además de ser plana, tiene que tener dependencia temporal armónica de frecuencia ω , entonces se puede encontrar la forma de f y g como ya hicimos muchas veces, es decir haciendo la derivada segunda temporal, integrando la parte espacial y llegando a senos y cosenos espaciales. Otra manera es pensar esto con el resultado de la ec. (7.10). Consideremos la progresiva: decir que sólo es función de $\hat{n} \cdot \vec{r} - vt$ es lo mismo que decir que sólo es función de $\hat{n} \cdot \vec{r}/v - t$, que equivale a decir que sólo es función de $t - \hat{n} \cdot \vec{r}/v$. Y acá agregamos que si sabemos que debe tener dependencia temporal armónica de frecuencia ω (por ejemplo, porque es la respuesta a un forzado estacionario) entonces debe variar como un coseno o un seno de ωt . Pero la función f no depende de t y de $\hat{n} \cdot \vec{r}$ por separado, sino solamente en la combinación $t - \hat{n} \cdot \vec{r}/v$. Entonces, resulta inevitable que toda perturbación progresiva que: i) sea solución de la ec. de ondas (7.18), ii) dependa de una dirección fija en el espacio y iii) sea armónica, se tenga que escribir, a menos de una fase constante, como

$$\psi(\vec{r}, t) = A \cos[\omega(t - \hat{n} \cdot \vec{r}/v)] = A \cos(\omega t - \vec{k} \cdot \vec{r}) = A \cos(\vec{k} \cdot \vec{r} - \omega t), \quad (7.21)$$

donde se ha definido el **vector de ondas** $\vec{k}$

$$\vec{k} = \frac{\omega}{v} \hat{n}, \quad (7.22)$$

cuya magnitud es la del $k = 2\pi/\lambda$ de antes, y su dirección es la de propagación. Obviamente, puede resultar conveniente usar senos en lugar de cosenos, o notación exponencial. De manera análoga se obtienen las representaciones de las ondas planas regresivas armónicas (cambia un signo).

Hemos prestado atención a las ondas planas armónicas porque las soluciones de la forma (7.21) forman base, es decir **cualquier** solución de la ecuación (7.18) se puede escribir como superposición conveniente de estas ondas, con distintos valores de ω , distintas direcciones de $\vec{k}$ y con el módulo del vector $\vec{k}$ determinado por su definición de la ec. (7.22), basada en la ecuación de dispersión.

7.2.2. Fuentes con simetría esférica

Supongamos un medio 3D isótropo y que en este medio la magnitud física ψ satisface la ecuación de ondas clásica (7.18), con $\nabla^2 \psi = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2}$. Si las condiciones iniciales o las fuentes impulsoras o forzantes tienen simetría esférica, la descripción con

coordenadas cartesianas no es lo más conveniente (en realidad es horrible). Por ejemplo, cuando se informa que una explosión “se escuchó a una distancia de cien kilómetros”, se está especificando el valor de la coordenada esférica r y se da por entendido que no es necesario dar los valores de las coordenadas esféricas angulares θ y ϕ , porque los efectos de la explosión poseen simetría esférica o de revolución, es decir, que son iguales en todas las direcciones. La misma simetría aparece en el caso de fuentes pequeñas, que desde lejos se ven como puntuales. Si bien ideales, las fuentes puntuales cobran importancia cuando vale superposición, porque cualquier fuente extensa se puede analizar como una colección de fuentes puntuales.

En todos los casos mencionados, la simetría esférica de la fuente permite afirmar que la función de onda depende solamente de la coordenada r , es decir que debe ser $\psi(r, t)$ y no $\psi(r, \theta, \phi, t)$. Los puntos del espacio que tienen igual perturbación, o sea, los *frentes de onda*, son esferas concéntricas. En estas condiciones el operador ∇^2 no es la derivada segunda respecto a r dos veces, si no que su forma se obtiene a partir del cambio de variables $(x, y, z) \rightarrow (r, \theta, \phi)$. Aplicando la regla de la cadena para pasar las derivadas con respecto a coordenadas cartesianas a derivadas con respecto a coordenadas esféricas, el laplaciano resulta

$$\frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \frac{\partial \psi}{\partial r}) + \text{derivadas angulares} = \frac{1}{r} \frac{\partial^2}{\partial r^2} (r \psi) + \text{derivadas angulares} . \quad (7.23)$$

En el caso en que $\psi(r, t)$, las derivadas con respecto a θ y ϕ son cero y la eq. (7.18) queda

$$\frac{1}{r} \frac{\partial^2}{\partial r^2} (r \psi) = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2} , \quad (7.24)$$

es decir,

$$\frac{\partial^2}{\partial r^2} (r \psi) = \frac{1}{v^2} \frac{\partial^2 (r \psi)}{\partial t^2} . \quad (7.25)$$

Este resultado muestra que la combinación $r \psi$ satisface la ecuación de ondas 1D (6.41). Luego, la solución más general para $r \psi$ viene dada por la ec. (7.10)

$$r \psi = f(r - vt) + g(r + vt) , \quad (7.26)$$

y la función de onda más general para simetría esférica resulta la suma de una perturbación progresiva f/r (que se aleja del origen, explosión, o divergencia) mas una perturbación regresiva g/r (que se acerca al origen, implosión, o convergencia)

$$\psi(r, t) = \frac{f(r - vt)}{r} + \frac{g(r + vt)}{r} . \quad (7.27)$$

Para fuentes armónicas, la onda esférica progresiva que se aleja del origen resulta entonces, a menos de una fase,

$$\psi(r, t) = \frac{A}{r} \cos(kr - \omega t) , \quad (7.28)$$

o en notación exponencial

$$\psi(r, t) = \frac{A}{r} \exp i(kr - \omega t) . \quad (7.29)$$

Si el medio no tiene pérdidas, la energía transportada por un frente de onda particular tiene que ser constante a medida que el frente se aleja del origen. Este resultado es consistente con el hecho de que la amplitud resulta inversamente proporcional a r . Así, en cada elemento de área la densidad de energía por unidad de área es inversamente proporcional a r^2 , mientras que el área total del frente de ondas aumenta como r^2 . Esto explica por qué cuanto más lejos estamos del emisor, más débil se recibe la señal emitida, a menos que se agrande el área del detector, para compensar el efecto de la dependencia inversamente proporcional a r^2 . Justamente “agrandar el área del detector” es lo que se hace en los radiotelescopios de observatorios astronómicos con antenas de gran tamaño, como RATAN-600 o Arecibo, con antenas de diámetro 576 m y 305 m, respectivamente.

Notar que muy lejos de las fuentes, y para detectores pequeños, un frente de onda esférico puede aproximarse por un frente de onda plano.

7.2.3. Fuentes con simetría cilíndrica

Si las condiciones iniciales o las fuentes impulsoras o forzantes tienen simetría cilíndrica, en vez de esférica, se debe obtener la forma del laplaciano con argumentos similares a los empleados para obtener la eq. (7.24), pero ahora se debe hacer el cambio de variables de las coordenadas cartesianas (x, y, z) a las coordenadas cilíndricas (r, θ, z) (no es el mismo r de antes). De esta manera, la ecuación para ondas emitidas por fuentes con simetría cilíndrica (la de un alambre infinito), resulta

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \psi}{\partial r} \right) = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2} . \quad (7.30)$$

Notar que los puntos del espacio con igual perturbación, o sea, los frentes de onda, son cilindros concéntricos.

Las soluciones armónicas de (7.30) tienen una dependencia radial dada por las funciones de Bessel. Para lo que necesitamos en este curso, solamente diremos que el comportamiento asintótico de las soluciones armónicas lejos de las fuentes, $r \gg \lambda$, es

$$\psi(r, t) = \frac{A}{\sqrt{r}} \exp i(kr - \omega t) . \quad (7.31)$$

De nuevo, si el medio no tiene pérdidas, la energía transportada por un frente de onda particular (en forma de cilindro) tiene que ser constante a medida que el frente se aleja del origen. Este resultado es consistente con el hecho de que la amplitud resulta inversamente proporcional a $\sqrt{r}$. Así, en cada elemento de área la densidad de energía por unidad de área y por unidad de tiempo (la intensidad) es inversamente proporcional a r , mientras que el área total del frente de ondas aumenta como r . Para puntos de observación muy lejanos, y para detectores pequeños, un frente de onda cilíndrico puede aproximarse por un frente de onda plano.

7.3. Modulación

Hasta ahora prestamos atención a oscilaciones y ondas con dependencia temporal armónica, es decir “cosenos de ωt ” (o senos o exponenciales). ¿Por qué la fijación con esta dependencia temporal? Por lo visto anteriormente, hay solamente dos respuestas posibles (ok, hay muchas más, pero solamente dos tuvieron protagonismo en nuestro curso hasta ahora). Como ya se habrán respondido mentalmente, una de las respuestas posibles es *porque queríamos ver los modos*. Y la segunda respuesta posible es *porque estábamos interesados en la respuesta forzada estacionaria*. Si el extremo de un sistema se fuerza armónicamente, $\psi(0, t) = A_0 \cos \omega t$, y si la perturbación ψ es solución de la ecuación de ondas clásica 1D, entonces la propagación de la perturbación progresiva se obtiene reemplazando t por $t - x/v(\omega)$ y así se tiene que $\psi(x, t) = A_0 \cos(k(\omega)x - \omega t)$. Pero con este tipo de perturbación, vista como una señal para transmitir información, no se puede transmitir nada, es como querer comunicarse cantando siempre la misma nota. Para transmitir información hay que hacer cambios de algo, ya sea cambiar la intensidad, el ritmo, la frecuencia, la fase ... algo. Esto se llama **modular**: producir cambios que sirvan para codificar información. Si la fuerza impulsora en el extremo de un sistema que transmite ondas responde a estos cambios temporales, el sistema se puede usar para que los cambios viajen, lleguen a un receptor y así transmitir información.

A continuación veremos un tercer motivo para justificar tanto interés por las funciones armónicas. Y este tercer motivo surge de una sorprendente propiedad de una gran clase de funciones que empiezan y terminan (no como los senos y cosenos, que no empiezan ni terminan nunca). Resulta que los elementos de esta gran clase de funciones se pueden escribir como superposición continua de funciones con dependencia temporal armónica. Por este motivo, a partir de este preciso momento centraremos la atención en entender los nuevos aspectos que aparecen cuando se superponen distintas frecuencias. El primer aspecto nuevo es incorporar que la velocidad de fase $v(\omega)$ puede depender de la frecuencia. Si bien para el ejemplo de la cuerda o para ondas electromagnéticas en el vacío la velocidad de fase es una constante y entonces la ecuación de dispersión $k(\omega) = \omega/v$ es una función lineal de ω , en otros casos (como para las ondas electromagnéticas en medios materiales) la velocidad de fase $v(\omega)$ puede depender de la frecuencia y en este caso $k(\omega)$ no es una función lineal. Cuando la velocidad de fase depende de la frecuencia, se dice que el medio es **dispersivo** (otra acepción de dispersivo, la acepción anterior estaba en la página 146).

7.3.1. Caso más sencillo

El ejemplo más sencillo de superposición de funciones con dependencia temporal armónica es el de dos frecuencias distintas con la misma amplitud y fase. El resultado relevante está en la ecuación (??). Supongamos que la fuerza impulsora al comienzo del sistema, en vez de ser monocromática como dijimos al comienzo de 7.3, produce

una deformación $\psi(0, t)$ de la forma

$$\psi(0, t) = A_0(\cos \omega_1 t + \cos \omega_2 t) = \underbrace{2A_0 \cos \omega_m t}_{A_m(t)} \cos \omega_p t = A_m(t) \cos \omega_p t. \quad (7.32)$$

Como antes, si la perturbación ψ es solución de la ecuación de ondas clásica 1D, las perturbaciones progresivas asociadas con ω_1 y con ω_2 se obtienen reemplazando t por $t - x/v(\omega)$. Así se tiene que la perturbación resultante se puede escribir como

$$\psi(x, t) = A_0 \cos(k(\omega_1)x - \omega_1 t) + A_0 \cos(k(\omega_2)x - \omega_2 t) = \underbrace{2A_0 \cos(k_m x - \omega_m t)}_{A_m(x, t)} \cos(k_p x - \omega_p t), \quad (7.33)$$

con $k(\omega_1) = \omega_1/v(\omega_1)$ y $k(\omega_2) = \omega_2/v(\omega_2)$. Es decir, que la perturbación resultante se obtiene como una perturbación portadora, de frecuencia

$$\omega_p = (\omega_1 + \omega_2)/2, \quad (7.34)$$

y número de onda

$$k_p = (k(\omega_1) + k(\omega_2))/2, \quad (7.35)$$

modulada por una perturbación moduladora, de frecuencia

$$\omega_m = (\omega_1 - \omega_2)/2, \quad (7.36)$$

y número de onda

$$k_p = (k(\omega_1) - k(\omega_2))/2. \quad (7.37)$$

Si queremos seguir al coseno que corresponde a la modulación (es decir, si queremos ir siempre con un punto característico, como la cresta, o el valle), entonces hay que mantener el argumento $k_m x - \omega_m t$ en el mismo valor. Dicho de otra manera, para que $A_m(x + \Delta x, t + \Delta t) = A_m(x, t)$, Δx y Δt tienen que estar ligados por la relación

$$\frac{\Delta x}{\Delta t} = \frac{\omega_m}{k_m} = \frac{\omega_1 - \omega_2}{k(\omega_1) - k(\omega_2)}. \quad (7.38)$$

En el límite de frecuencias muy próximas, el lado derecho de esta ecuación tiende a la derivada $d\omega/dk$, la *derivada* de la relación de dispersión, mientras que el lado izquierdo tiende a la velocidad necesaria para ir siempre con un punto característico de la modulación. Esta velocidad se llama **velocidad de grupo** y siempre coincide con la velocidad de propagación de la información, es decir de la energía

$$v_g = \frac{d\omega}{dk}, \quad (7.39)$$

que por lo general no coincide con la velocidad de fase, a menos que la relación de dispersión sea lineal, es decir cuando la velocidad de fase v no depende de la frecuencia.

7.3.2. Modulaciones reales

El caso de pulsaciones es un caso especial de fuerzas impulsoras cuya dependencia temporal puede expresarse como combinación lineal de distintas funciones armónicas,

$$\psi(0, t) = \sum_{\omega} A(\omega) \cos [\omega t + \phi(\omega)] , \quad (7.40)$$

y esta combinación lineal también resulta

$$\psi(0, t) = A_{mod}(t) \cos \omega_p t , \quad (7.41)$$

producto de una perturbación portadora, a la frecuencia promedio, por una perturbación moduladora. En el caso fácil pero superidealizado de pulsaciones, la moduladora es periódica a la frecuencia semidiferencia. Pero en general, en los sistemas reales, la moduladora contiene funciones que codifican la información que se quiere transmitir, ya sea en forma digital (una secuencia de ceros y unos, como la información musical en los DVDs) o en forma analógica (como la información musical en los discos de vinilo). Para elegir la frecuencia promedio (de la portadora) se deben poner de acuerdo el emisor y el receptor, aunque en la práctica esto no es posible para ondas electromagnéticas (la base de la sociedad de la información tal como la conocemos) porque existen en todo el mundo diversos entes de control que gestionan el uso del espectro para cada aplicación. Para ondas de radio la portadora tiene la frecuencia que marca el dial, $\nu \approx 10^3$ kHz para AM, $\nu \approx 10^2$ MHz para FM. En cambio, las frecuencias que componen la moduladora vienen impuestas por las frecuencias características de la información que se quiere transmitir.

Tomemos el ejemplo sencillo de radio AM. La perturbación en el comienzo del sistema (es decir, las corrientes eléctricas en la antena transmisora) tienen la forma (7.41), pero la modulación $A_{mod}(t)$, en vez de ser una simple función armónica a la frecuencia semidiferencia, se sintetiza mediante una suma de funciones armónicas. Esta suma tiene un término constante, que representa la presión atmosférica en el estudio de radio cuando todo está en silencio, más otros términos armónicos con amplitudes A_{Mj} , frecuencias ω_{Mj} y fases ϕ_{Mj} que representan los sonidos que se desea transmitir, ya sea música, un discurso o una conversación. Resumiendo, la modulación $A_{mod}(t)$ se expresa de la siguiente manera

$$A_{mod}(t) = A_{M0} + \sum_j A_{Mj} \cos [\omega_{Mj} t + \phi_{Mj}] , \quad (7.42)$$

donde el primer subíndice es una etiqueta que dice que estamos hablando de una modulación, mientras que el segundo subíndice numera de alguna manera a las frecuencias moduladoras. Como ésta es la forma que modula a la portadora, al introducir (7.42) en (7.41), y luego de usar la identidad trigonométrica que transforma el producto de cosenos en una suma, se llega a que la perturbación en el comienzo del sistema se escribe como

$$\begin{aligned} \psi(0, t) = A_{M0} \cos \omega_p t + \frac{1}{2} \sum_j A_{Mj} \cos [(\omega_{Mj} + \omega_p)t + \phi_{Mj}] \\ + \frac{1}{2} \sum_j A_{Mj} \cos [(\omega_p - \omega_{Mj})t - \phi_{Mj}] , \end{aligned} \quad (7.43)$$

La expresión (7.43) pone en evidencia que para poder irradiar ondas que tengan toda la información sobre los sonidos audibles que ocurren en el estudio de radio, entonces la “fuerza impulsora” en la antena transmisora tiene que estar sintetizada con señales armónicas con frecuencias angulares entre la menor frecuencia contenida en la segunda sumatoria (conocida como banda lateral inferior), y la mayor frecuencia contenida en la primera sumatoria (conocida como banda lateral superior), es decir

$$\omega_p - \omega_M^{Max} \leq \omega \leq \omega_p + \omega_M^{Max} , \quad (7.44)$$

donde ω_M^{Max} es el máximo valor de todas las frecuencias de modulación que intervienen en (7.42). La diferencia entre la mayor frecuencia de la banda lateral superior y la menor frecuencia de la banda lateral inferior se conoce como *ancho de banda* y según (7.44) resulta igual a $2\omega_M^{Max}$. Debido a la existencia del ancho de banda, el número de canales (estaciones de radio, de televisión, etc.) en un dado rango de frecuencias está limitado. Si estamos hablando de frecuencias de audio, entre $\nu \approx 50$ Hz y $\nu \approx 20$ kHz, el ancho de banda debería ser de aproximadamente 40 kHz, pero debido que AM no es muy pretenciosa con la fidelidad, y teniendo en cuenta que la nota más alta del piano tiene frecuencia $\nu \approx 4.2$ kHz, es usual tener anchos de banda del orden de 5 kHz. En cambio para televisión, las frecuencias de modulación vienen impuestas por la persistencia retiniana, el número de píxeles de cada cuadro, el número de colores ... y es usual tener anchos de banda del orden de algunas decenas de MHz (y por este motivo, las frecuencias portadoras de las transmisiones aéreas de televisión están en el rango de las centenas de MHz).

El estudio de la dispersión en un medio es muy importante para las comunicaciones. Esto es así porque si el medio es muy dispersivo, la velocidad de fase depende fuertemente de la frecuencia, y entonces la señal emitida por la fuerza impulsora $\psi(0, t)$ (o la antena emisora, o cualquier perturbación que se imponga en el medio para que llegue a un receptor), se va a ir degradando debido a que las distintas componentes frecuenciales viajan con distintas velocidades de fase y entonces la modulación no mantiene su forma a medida que la onda avanza. En cambio, si el medio es poco dispersivo y si el contenido de frecuencias de la señal no es muy grande, la modulación mantiene su forma y se propaga con la velocidad de grupo v_g .

7.3.3. Síntesis de señales

Ya vimos que cuando el extremo de un sistema que responde a la ecuación de ondas clásica 1D se fuerza armónicamente para que la perturbación inicial tenga la forma $\psi(0, t) = A_0 \cos \omega t$, la perturbación progresiva se propaga y penetra en el sistema. Si bien esto NO sirve como señal para transmitir información, porque una función puramente periódica no empieza ni termina nunca, Sí sirve para sintetizar otras señales que den lugar a perturbaciones progresivas finitas, es decir, con localización temporal y espacial. Esto es así porque se puede probar que muchas funciones que empiezan y terminan son sintetizables a partir de funciones armónicas con distintas frecuencias. Como la prueba rigurosa está fuera del alcance del curso, a continuación estudiaremos la **síntesis** en un caso sencillo, pero no tan supersencillo como el de dos cosenos con

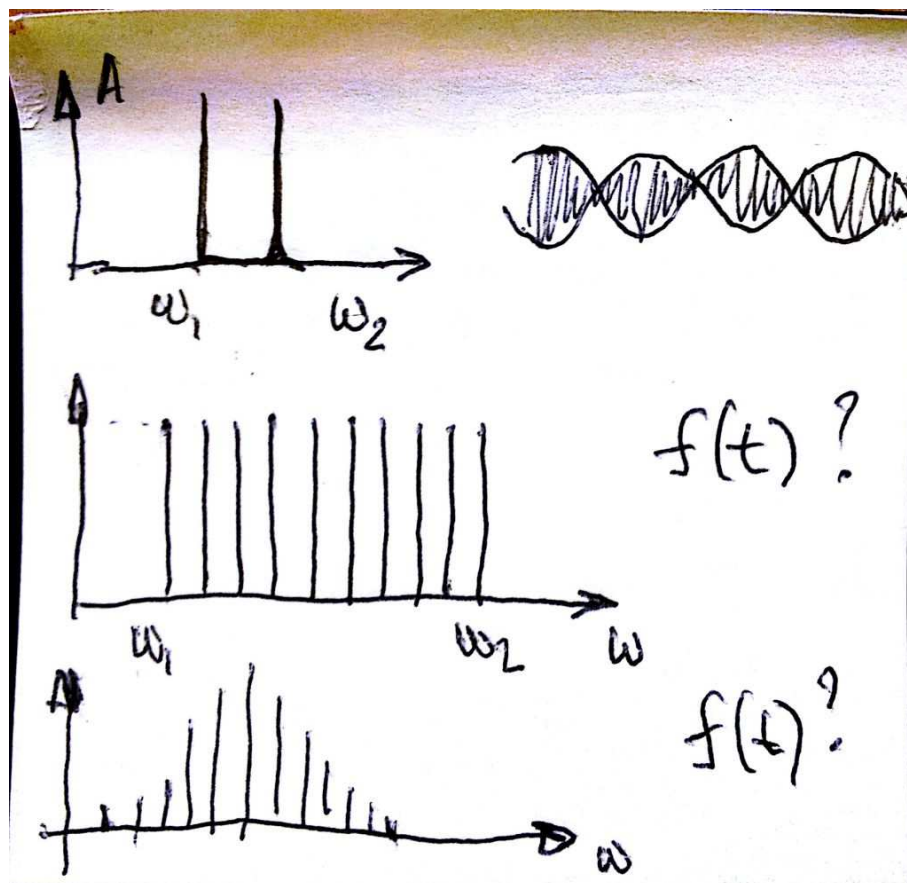


Figura 7.1: Los gráficos de la columna izquierda simbolizan cómo se elige la síntesis de funciones a partir de sumar cosenos. En la primera fila, el ejemplo de los dos cosenos, en la segunda fila, el ejemplo discutido en 7.3.3.

la misma amplitud y fase que vimos antes. Los gráficos de la columna izquierda de la Figura 7.1 representan casos de síntesis de funciones a partir de sumar cosenos. En el eje vertical se grafican las amplitudes y en el eje horizontal las frecuencias de los cosenos usados para la síntesis. En la primera fila, el ejemplo de los dos cosenos con dos frecuencias distintas con la misma amplitud y fase (como un peine con dos dientes de la misma altura), que da lugar a la conocida función del tiempo que corresponde al batido: una portadora a la frecuencia promedio, modulada por una amplitud variable en el tiempo y que viaja a la velocidad de grupo. Las funciones de las amplitudes en función de las frecuencias representadas en los gráficos de la columna izquierda de la Figura 7.1 se conocen como *espectros*. Otros ejemplos que conocemos de funciones que no empiezan ni terminan nunca es el caso de funciones periódicas, sintetizadas a partir de series de Fourier, como ya se discutió en una animación, ver enlace en <http://bit.ly/2cDIVqE>.

Mezclando N componentes armónicos

En la segunda fila de la Figura 7.1 se esquematiza el nuevo ejemplo de la clase de hoy, vamos a sintetizar una función a partir de N cosenos con la misma amplitud (a) y con la misma fase y con frecuencias ω_p equiespaciadas en el intervalo $\omega_1 \leq \omega \leq \omega_2$.

El dibujito ahora es como un peine con N dientes de la misma altura y a diferencia de ejemplo con las series de Fourier, todas las frecuencias están localizadas en un intervalo finito. Primero investigaremos qué tipo de perturbación inicial $\psi(0, t)$ se logra con esta mezcla de cosenos. Y luego, cómo se propaga esta perturbación por el sistema.

Si $\psi(0, t)$ es la suma de estos N cosenos

$$\psi(0, t) = \sum_{n=0}^{N-1} a \cos [(\omega_1 + n \delta \omega)t] , \quad (7.45)$$

con

$$\delta \omega = \frac{\omega_2 - \omega_1}{N - 1} = \frac{\Delta \omega}{N - 1} , \quad \Delta \omega = \omega_2 - \omega_1 . \quad (7.46)$$

Conviene pasar a notación compleja, cada coseno es la parte real de una exponencial compleja y las amplitudes y una parte de las exponenciales sale de factor común de la sumatoria. Sobreentendiendo que luego tomaremos parte real y empleando la expresión para la suma de los N primeros términos de la sucesión geométrica, tenemos que

$$\psi(0, t) = a e^{i \omega_1 t} \sum_{n=0}^{N-1} [e^{i \delta \omega t}]^n = a e^{i \omega_1 t} \frac{e^{i N \delta \omega t} - 1}{e^{i \delta \omega t} - 1} , \quad (7.47)$$

que se puede reescribir como

$$\psi(0, t) = a e^{i[\omega_1 + (N-1) \delta \omega / 2] t} \frac{\sin(\frac{N}{2} \delta \omega t)}{\sin(\frac{1}{2} \delta \omega t)} = a \frac{\sin(\frac{N}{2} \delta \omega t)}{\sin(\frac{1}{2} \delta \omega t)} \cos \omega_p t = A_{mod}(0, t) \cos \omega_p t , \quad (7.48)$$

con

$$\omega_p = \omega_1 + \frac{\Delta \omega}{2} , \quad (7.49)$$

la frecuencia promedio (la que está en el medio del peine), tal como ya habíamos anticipado en la eq. (7.41).

Límite $N \rightarrow \infty$, con $\Delta \omega$ constante

Si se deja fijo $\Delta \omega$ pero se hace más pequeño el valor $\delta \omega$, es decir, se ponen mas dientes en el peine ($N \rightarrow \infty$), $T_{mod} \rightarrow \infty$. Y este resultado equivale a decir que en el límite $N \rightarrow \infty$, la moduladora deja de ser periódica (la portadora lo sigue siendo). Es casi evidente que en este límite hemos sintetizado una función no periódica a partir de una suma de cosenos. Veremos que, además de no ser periódica, la moduladora empieza y termina, es decir que puede servir para representar señales.

Para que el proceso tenga sentido, al aumentar el número de dientes en el peine tenemos que disminuir la altura de los dientes (las amplitudes a de las componentes

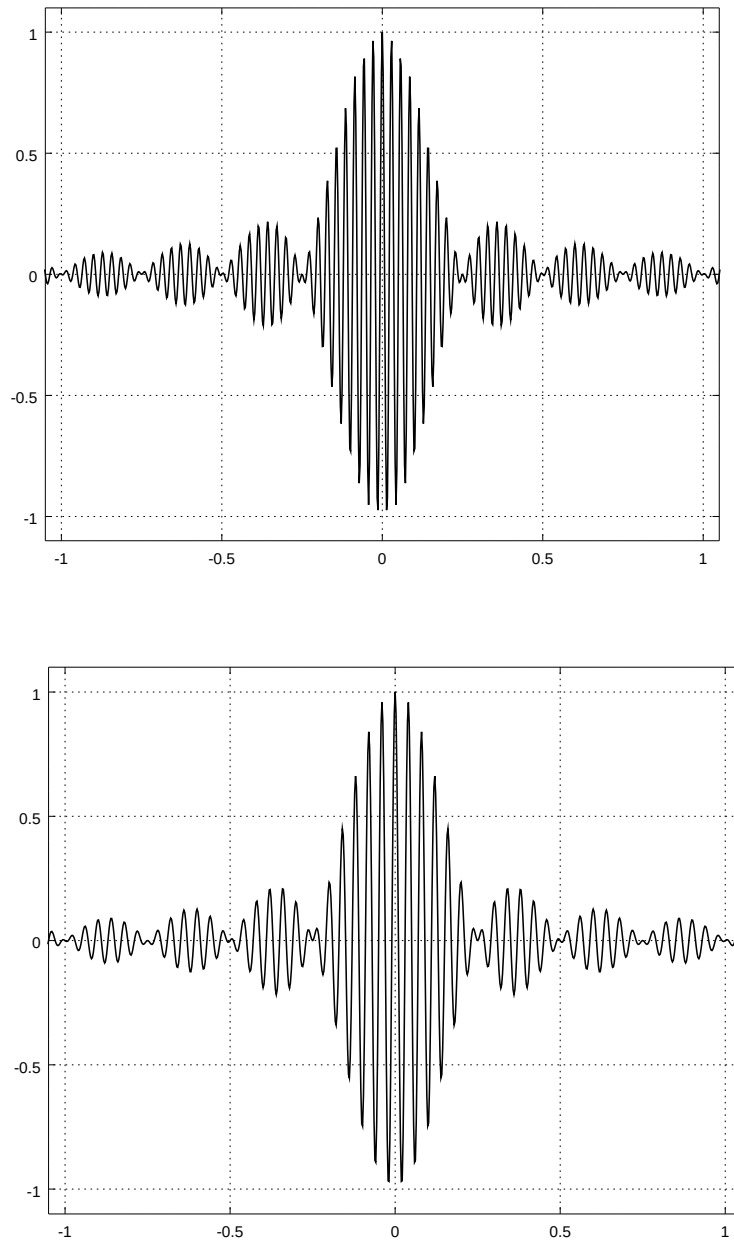


Figura 7.2: $\psi(0, t)$ dado por la ec. (7.54) en función del tiempo, $A_0 = 1$ y distintos valores de ω_1 y ω_2 . (a) $\nu_1 = 33\text{Hz}$ y $\nu_2 = 37\text{Hz}$; (b) $\nu_1 = 23\text{Hz}$ y $\nu_2 = 27\text{Hz}$.

armónicas), porque no queremos que la fuerza impulsora en el extremo del sistema $\psi(0, t)$ se haga infinita. Para que esto no pase, pedimos que la amplitud de modulación

$$A_{mod}(0, t) = a \frac{\sin(\frac{N}{2} \delta\omega t)}{\sin(\frac{1}{2} \delta\omega t)}, \quad (7.50)$$

siempre tenga el mismo valor en el instante inicial, independientemente del valor de N . Y como $A_{mod}(0, 0) = aN$, esto equivale a pedir que en este límite $aN = A_0$ cuando $N \rightarrow \infty$ (A_0 una constante). De esta manera, conviene escribir la amplitud de modulación

como

$$A_{mod}(0, t) = \frac{A_0 \sin(\frac{N}{2} \delta\omega t)}{N \sin(\frac{1}{2} \delta\omega t)} . \quad (7.51)$$

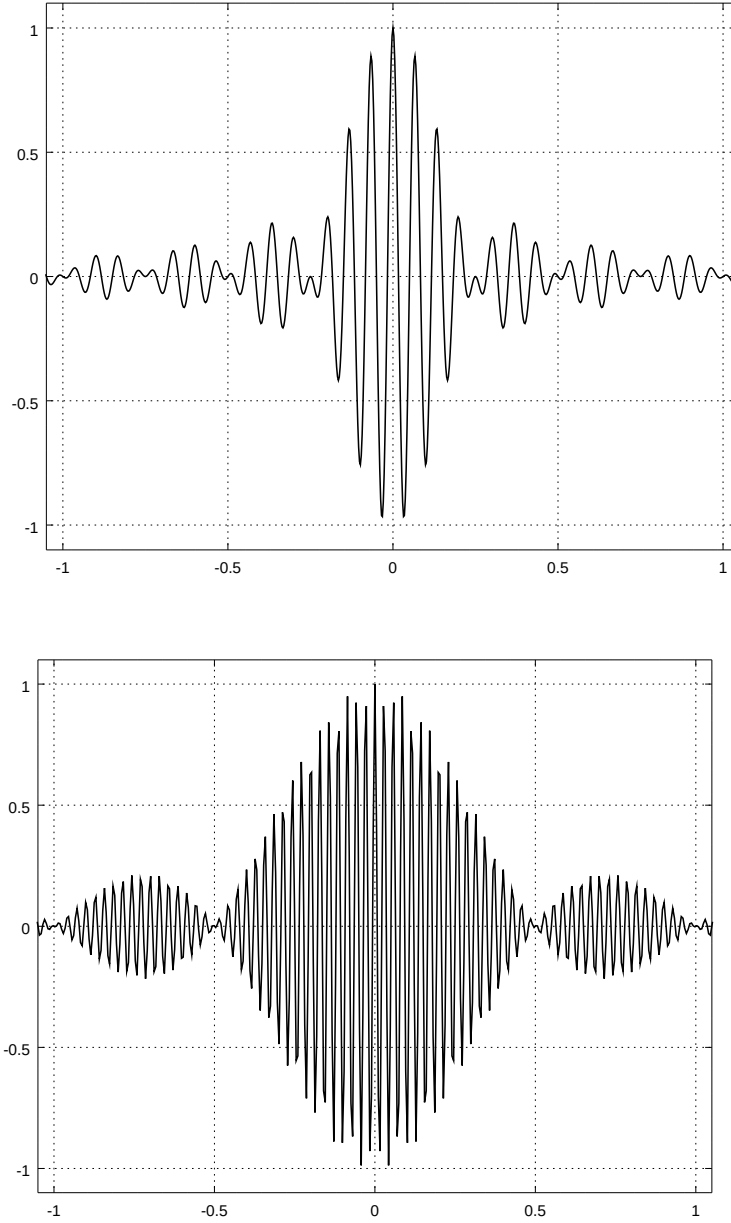


Figura 7.3: $\psi(0, t)$ como en la Fig. 7.2, pero (a) $\nu_1 = 13\text{Hz}$ y $\nu_2 = 17\text{Hz}$; (b) $\nu_1 = 34\text{Hz}$ y $\nu_2 = 36\text{Hz}$.

Ahora está todo preparado para hacer el límite $N \rightarrow \infty$, $\delta\omega \rightarrow 0$, con $\Delta\omega$ constante. En este límite,

$$N\delta\omega \approx (N - 1)\delta\omega = \Delta\omega = \omega_2 - \omega_1 , \quad (7.52)$$

y la amplitud de modulación se puede escribir

$$A_{mod}(0, t) = A_0 \frac{\sin(\frac{1}{2} \Delta\omega t)}{\frac{1}{2} \Delta\omega t}, \quad (7.53)$$

y entonces la dependencia temporal de la perturbación en el comienzo del sistema resulta

$$\psi(0, t) = A_0 \frac{\sin(\frac{1}{2} \Delta\omega t)}{\frac{1}{2} \Delta\omega t} \cos \omega_p t. \quad (7.54)$$

Por otra parte, este resultado se obtuvo calculando $\psi(0, t)$ como

$$\psi(0, t) = \lim_{N \rightarrow \infty} \sum_{n=0}^{N-1} a \cos[(\omega_1 + n \delta\omega)t] = \lim_{N \rightarrow \infty} \frac{A_0}{\Delta\omega} \sum_{n=0}^{N-1} \cos[(\omega_1 + n \delta\omega)t] \delta\omega, \quad (7.55)$$

que en el límite es igual a

$$\psi(0, t) = \frac{A_0}{\Delta\omega} \int_{\omega_1}^{\omega_2} \cos \omega t d\omega. \quad (7.56)$$

Las eqs. (7.54) y (7.56) muestran que

$$\frac{A_0}{\Delta\omega} \int_{\omega_1}^{\omega_2} \cos \omega t d\omega = A_0 \frac{\sin(\frac{1}{2} \Delta\omega t)}{\frac{1}{2} \Delta\omega t} \cos \omega_p t, \quad (7.57)$$

es decir, que sumar infinitos cosenos de la misma amplitud y la misma fase en el intervalo $\omega_1 < \omega < \omega_2$ (lo que equivale a hacer $\int_{\omega_1}^{\omega_2}$ de una función constante multiplicada por cosenos), da como resultado una señal portadora a la frecuencia promedio, modulada por una amplitud moduladora variable en el tiempo y que tiene la forma de la función $\sin u/u$, que se lee **seno cardinal de u** y que también se suele notar como $\text{sinc}(u)$. Es una función par, que tiende a cero cuando u crece (porque el seno está acotado) y que tiene los ceros del seno, excepto en el origen donde tiene su máximo. La variable u es

$$u = \frac{1}{2} \Delta\omega t = \pi \Delta\nu t, \quad (7.58)$$

y la (7.54) muestra que si se deja el ancho $\Delta\omega$ fijo pero se cambian los valores de ω_1 y ω_2 , la forma de la modulación es siempre la misma, lo que cambia es el valor de la frecuencia promedio, es decir cuántas oscilaciones de la portadora entran desde que la modulación toma su valor máximo, para $u = 0$, hasta que toma su primer cero, para $u = \pi$. Desde que la modulación toma su máximo valor hasta que se hace cero por primera vez transcurre un tiempo Δt que resulta

$$\Delta t = 1/\Delta\nu, \quad (7.59)$$

y que puede ser usado como medida de la localización temporal (cuánto dura) la modulación.

La función del tiempo resultante de sumar infinitos cosenos en un intervalo fijo se muestra en las figuras 7.2 y 7.3, donde se ha graficado $\psi(0, t)$, dado por la ec. (7.54),

en función del tiempo (también se puede graficar en función de u , cuál sería la ventaja) para $A_0 = 1$ y distintos valores de ω_1 y ω_2 . Notar que a pesar de que los gráficos de las figura 7.2a, 7.2b y 7.3a corresponden a valores diferentes de ν_1 y ν_2 , en todos se mantuvo el mismo valor $\Delta\nu = 4\text{Hz}$. Y por este motivo la la modulación en estas tres curvas tiene siempre el mismo ancho característico $\Delta t = 0.25\text{s}$, aunque lo que cambia es la frecuencia de la portadora. En cambio, en el ejemplo de la figura 7.3b, el intervalo de frecuencias donde se hace la suma de cosenos se reduce a la mitad, $\Delta\nu = 2\text{Hz}$. Por eso ahora la modulación tiene un ancho característico que es el doble, $\Delta t = 0.5\text{s}$. Este comportamiento, anunciado por la eq. (7.59), es común a todas las funciones sintetizadas a partir de funciones armónicas. Se conoce como **principio de incerteza** y dice que la localización de la función que se sintetiza a partir de funciones armónicas (en este caso localización temporal) es inversamente proporcional a la localización de las frecuencias de los senos y cosenos usados. No es una propiedad que valga solamente para síntesis temporales, por ejemplo la misma propiedad valdría para la localización espacial de funciones que se sintetizan a partir de funciones como $\cos kx$, con frecuencias espaciales k . El caso límite familiar en el contexto de estos ejemplos es un coseno, que no tiene localización temporal pero está perfectamente localizado frecuentemente. El caso límite opuesto es un impulso temporal, algo que se prende y se apaga instantáneamente, que está muy bien localizado temporalmente, pero que es de esperar que no esté bien localizado frecuentemente. Calcular analíticamente la disminución relativa de la altura de los máximos secundarios, mas notoria cuando se observa en la energía de la onda resultante, ya que la energía va con el cuadrado de la perturbación.

7.3.4. Transformada de Fourier

La superposición de cosenos que acabamos de hacer, representada por la eq. (7.56), es un caso particular de funciones del tiempo $f(t)$ obtenidas mediante superposiciones de la forma

$$f(t) = \int_0^\infty \tilde{A}(\omega) \cos \omega t d\omega + \int_0^\infty \tilde{B}(\omega) \sin \omega t d\omega, \quad (7.60)$$

que se conoce como **integral de Fourier**. Las funciones $\tilde{A}(\omega)$ y $\tilde{B}(\omega)$ dan la transformada de Fourier de la función $f(t)$. En nuestro ejemplo $\tilde{B}(\omega) = 0$ para todo ω (porque no usamos funciones seno) y $\tilde{A}(\omega)$ es una función partida con forma rectangular, que vale $\tilde{A}(\omega) = \frac{A_0}{\Delta\omega}$ para $\omega_1 < \omega < \omega_2$ y $\tilde{A}(\omega) = 0$ fuera del intervalo $\omega_1 < \omega < \omega_2$. Lo que hicimos al construir la función $f(t)$ se llama *síntesis* de Fourier. El proceso inverso, que consiste en descomponer una determinada $f(t)$ en sus componentes armónicas, se llama *análisis* de Fourier. Ambos procesos, síntesis y análisis, son completamente análogos a los vistos para funciones periódicas y pueden verse como el límite para período infinito. Si se hiciera este límite a partir de las ecuaciones obtenidas para series de Fourier (ver página 108), se obtendría que las funciones $\tilde{A}(\omega)$ y $\tilde{B}(\omega)$ resultan

$$\tilde{A}(\omega) = \frac{1}{\pi} \int_0^\infty f(t) \cos \omega t dt, \quad (7.61)$$

$$\tilde{B}(\omega) = \frac{1}{\pi} \int_0^\infty f(t) \sin \omega t \, dt. \quad (7.62)$$

Si se prefiere usar notación exponencial, hay que escribir los senos y los cosenos en términos de las exponenciales, separar los términos con exponenciales de exponente con distinto signo, introducir amplitudes complejas $\tilde{f}(\omega)$ definidas a partir de las amplitudes $\tilde{A}(\omega)$ y $\tilde{B}(\omega)$ y extendidas a frecuencias negativas. En notación compleja, la síntesis y el análisis se escriben

$$f(t) = \int_{-\infty}^\infty \tilde{f}(\omega) \exp(-i\omega t) \, d\omega. \quad (7.63)$$

$$\tilde{f}(\omega) = \frac{1}{2\pi} \int_{-\infty}^\infty f(t) \exp(i\omega t) \, dt. \quad (7.64)$$

Volviendo a la perturbación $\psi(0, t)$ al comienzo del sistema, la expresión (7.63) es muy útil para seguir la propagación de una señal. Con $\psi(0, t)$ como en nuestro ejemplo (eq. (7.57)), la respuesta progresiva $\psi(0, t)$ que resulta del forzado estacionario en $x > 0$ se obtiene reemplazando t por $t - x/v(\omega)$ en el lado izquierdo de la eq. (7.57). En medios no dispersivos, v es una constante y la integral debe dar el mismo resultado que antes, excepto que t se reemplaza por $t - x/v(\omega)$

$$\psi(x, t) = \frac{A_0}{\Delta\omega} \int_{\omega_1}^{\omega_2} \cos \omega (t - x/v) \, d\omega = A_0 \frac{\sin(\frac{1}{2} \Delta\omega (t - x/v))}{\frac{1}{2} \Delta\omega (t - x/v)} \cos \omega_p (t - x/v), \quad (7.65)$$

es decir que la forma inicial se propaga a la velocidad v sin deformarse. Pero si el medio es dispersivo, $v(\omega)$ no es una constante y entonces el pulso inicial por lo general cambia de forma a medida que avanza. Sólo en el caso en que la función $k(\omega)$ se puede aproximar en todo el intervalo de integración por una función lineal, la forma inicial se propaga a la velocidad v_g sin deformarse, porque en este caso la integral se puede aproximar por

$$\psi(x, t) \approx A_0 \frac{\sin(\frac{1}{2} \Delta\omega (t - x/v_g))}{\frac{1}{2} \Delta\omega (t - x/v_g)} \cos(\omega_p t - k_p x), \quad (7.66)$$

con $k_p = k(\omega_p)$, v_g la velocidad de grupo definida en la ec. (7.39).

Capítulo 8

Respuesta forzada estacionaria

En los capítulos anteriores hemos aprendido a encontrar los modos normales de sistemas con muchas partes móviles y seguramente nos dimos cuenta que, más allá de la descripción matemática, la simetría del sistema juega un papel fundamental para descubrir los modos normales. La relación entre simetría y modos normales se puede estudiar sistemáticamente mediante la teoría de grupos, una rama de las matemáticas que trata sobre simetrías. Usando teoría de grupos (fuera del alcance de este curso) es posible encontrar los modos normales de sistemas con un gran número de partes móviles. Por ejemplo, es posible encontrar los modos normales de un cristal perfecto de silicio, uno de los materiales más importantes en la tecnología electrónica, resolviendo determinantes de 6×6 , en vez de los determinantes monstruosos que habría que resolver solamente teniendo en cuenta el enorme número de átomos que componen un trozo macroscópico de cristal.

Otra manera de encontrar los modos normales de sistemas con muchas partes móviles es mediante el estudio de las resonancias que se observan en la respuesta forzada estacionaria. La respuesta forzada estacionaria es el movimiento que realiza el sistema cuando se lo somete a la acción de fuerzas exteriores (no solamente las fuerzas restitutivas que juegan en los modos) y cuando se han extinguido los movimientos transitorios que dependen de las condiciones iniciales. Desde el punto de vista matemático, el transitorio es la solución de ecuaciones de movimiento homogéneas, mientras que el estacionario es la solución particular cuando la fuerza impulsora es armónica.

La respuesta forzada estacionaria es notablemente intensa cuando la frecuencia de la fuerza impulsora coincide con la frecuencia de un modo normal, mientras que es poco intensa y casi despreciable en otros casos. Esto significa que las formas y las frecuencias de los modos de un sistema se revelan experimentalmente en los máximos de respuesta cuando el sistema es forzado mediante una fuerza impulsora de frecuencia variable. Y a partir de formas y frecuencias modales, es posible inferir propiedades y características físicas internas de sistemas que incluso pueden ser inaccesibles, como la atmósfera de un planeta lejano.

Notar que el movimiento colectivo de las partes móviles de un sistema en la respuesta forzada estacionaria tiene mucho en común con el movimiento colectivo de las partes

móviles cuando el sistema se encuentra en un modo normal: en ambas situaciones las partes móviles oscilan con la misma frecuencia y pasan simultáneamente por la posición de equilibrio (tienen la misma fase). Por este motivo, la descripción matemática de ambos movimientos es muy similar, aunque con varias diferencias, la más importante es que mientras que en la respuesta forzada estacionaria la frecuencia es un dato exterior conocido, impuesto por la fuerza impulsora, la frecuencia en un modo es una característica dinámica interna, impuesta por las fuerzas restitutivas. Para poner en evidencia la íntima relación que existe entre el problema de modos de cualquier sistema físico y la respuesta forzada estacionaria de dicho sistema, como así también las similitudes y diferencias entre ambos problemas, revisaremos situaciones familiares de cadenas líneas y sogas y también veremos un nuevo sistema con una nueva fuerza restitutiva.

8.1. Forzado estacionario, cadena lineal

Consideremos la respuesta forzada estacionaria de una cadena lineal con N cuentas idénticas equiespaciadas, como se esquematiza en la figura 6.1. Como ya estudiamos la dinámica de esta cadena para pequeños apartamientos transversales del equilibrio, sabemos que las ecuaciones linealizadas de movimiento cuando solamente actúan las fuerzas restitutivas internas vienen dadas por el sistema de ecuaciones diferenciales (6.4), con $n = 1, 2, \dots, N$. Si ahora forzáramos a la parte móvil 3 de este sistema mediante una fuerza impulsora armónica, entonces deberíamos retocar solamente la ecuación correspondiente a $n = 3$, el resto de ecuaciones quedaría sin cambios. Para no complicar demasiado, imaginemos que forzamos alguna de las partes móviles “virtuales”, por ejemplo el extremo izquierdo ($n = 0$) o el extremo derecho ($n = N$), así ni siquiera tenemos que retocar el sistema (6.4).

Y ahora nos damos cuenta que, si estamos interesados solamente en la respuesta forzada estacionaria, todo el trabajo ya está hecho en la sección 6.1.2. Esto es así porque en el estado forzado estacionario, al igual que en el modo normal, las funciones $\psi_n(t)$ ($n = 1, 2, \dots, N$) deben representar movimientos donde todas las partes móviles oscilan con la misma frecuencia. Mienstra que en la sección 6.1.2 esta frecuencia era la frecuencia del modo, que dependía del sistema, ahora la frecuencia es la de la fuerza impulsora, que resulta un parámetro externo. Pero más allá de esta diferencia, no queda otra posibilidad que estas funciones adopten la forma empleada en los tratamientos de la sección 6.1.2, es decir $\psi_n(t) = A_n \cos(\omega t + \varphi)$ y por este motivo, volvemos a obtener las relaciones de recurrencia (6.5) o (6.6). Que ya sabemos que se resuelven con dependencias dadas por la ec. (6.7) y que conducen a relaciones de dispersión como (6.9) o su equivalente (6.10). Las mismas relaciones entre el período temporal asociado con ω y la dependencia espacial asociada con Θ que teníamos para los modos. Estas relaciones están representadas gráficamente en la figura 6.3.

8.1.1. Rango dispersivo

En el problema de modos, las condiciones de extremo discretizaban los valores de Θ . Por ejemplo, en el caso de extremos fijos resultaban los valores reales $\Theta = p\pi/(N+1)$, con p entero y esto llevaba a valores $0 < \omega_p < \omega_{Max}$. En el caso de modos la existencia de una frecuencia máxima ω_{Max} no nos había preocupado para nada. Y ahora, si la frecuencia ω de la fuerza impulsora es menor que ω_{Max} , tampoco nos preocupa, porque para todo valor ω en el eje vertical de la figura 6.3, siempre será posible encontrar un valor de Θ en el eje horizontal de la misma figura y en virtud de la dependencia espacial dada por la ec. (6.7), este valor será interpretado físicamente como una periodicidad espacial de la respuesta forzada estacionaria de la cadena lineal, todo igual que antes, debo decir aunque parezca pesado.

Todo sistema físico donde la periodicidad temporal de la fuerza impulsora produce una periodicidad espacial de la respuesta forzada estacionaria se dice que se comporta dispersivamente. La conclusión es entonces que para fuerzas impulsoras con frecuencias en el intervalo $0 < \omega < \omega_{Max}$, la cadena lineal se comporta dispersivamente. Notar que la palabra *dispersivo* se usa con dos acepciones distintas: i) como en la sección 6.4, como sinónimo de que la velocidad de fase depende de la frecuencia, y ii) como ahora, como sinónimo de que una periodicidad temporal en un estímulo produce una periodicidad espacial en el comportamiento del sistema.

8.1.2. Rango reactivo

Pero si la frecuencia ω de la fuerza impulsora es mayor que ω_{Max} , entonces si que tenemos que salir de nuestra zona de confort. Que no está mal, porque la zona de confort en ciencia impide lograr grandes cosas. Así que ... bienvenidos a esta zona de incomodidad, en la que nos damos cuenta que cuando $\omega > \omega_{Max}$ nunca será posible hallar valores de Θ en el eje horizontal de la figura 6.3. Y como ya probamos que la dependencia espacial (6.7) funciona bien, porque según (6.8) el cociente $\frac{A_{n+1}+A_{n-1}}{2A_n}$ no depende de n , tal como lo pide la física en la ec. (6.6), entonces tenemos que reconocer que quizás lo que no funciona es el prejuicio de pensar que, para toda frecuencia de fuerza impulsora, necesariamente tenemos que encontrar valores reales de Θ . Que no es otra cosa que el prejuicio de pensar que los senos y cosenos están siempre acotados. Lo están para argumentos reales, y esto muestra que quizás tenemos que buscar en argumentos complejos. Por otro lado, hay un indicio muy importante que puede facilitarnos la búsqueda, y es que ya sabemos que en el límite continuo estas cosas raras no suceden, porque la relación de dispersión es lineal y ω_{Max} convenientemente tiende a infinito y esta observación sugiere que nuestra zona de incomodidad desaparece en el límite continuo. Y ¿qué perdimos para hacer la aproximación continua? Recordemos que al hacer desarrollos de Taylor perdimos formas que no tienen bien definidas las derivadas espaciales de las ecuaciones de onda, como movimientos zig-zag.

Juntando los indicios de que los senos y cosenos de argumentos imaginarios no están acotados y de que en la aproximación continua perdimos movimientos zig-zag y observando que los senos y cosenos de argumentos complejos se entienden mejor cuando se

escriben como combinación de funciones exponenciales, es entendible que para la zona $\omega > \omega_{Max}$ haya que usar sucesiones de la forma

$$A_n = (-1)^n (b e^{-\Xi n} + c e^{\Xi n}) , \quad (8.1)$$

donde $\pm i \Xi$ juega el papel de $\pm \Theta$ y el $(-1)^n$ representa el zig-zag (o vaivén, para oscilaciones longitudinales) de partes móviles sucesivas. Para completar la demostración, hay que comprobar que con esta dependencia espacial, el cociente entre amplitudes $\frac{A_{n+1}+A_{n-1}}{2A_n}$ de la ec. (6.6) ahora resulta igual a $-\cosh \Xi$. Y entonces, la relación de dispersión en esta zona se reescribe como

$$\omega = \omega_{Max} \cosh \frac{\Xi}{2} , \quad (8.2)$$

que siempre tiene soluciones Ξ reales para frecuencias $\omega \geq \omega_{Max}$. La ec. (8.1) muestra que en este rango de frecuencias, la periodicidad temporal de la fuerza impulsora NO produce una periodicidad espacial de la respuesta forzada estacionaria, si no que produce una respuesta que es una combinación lineal de exponenciales crecientes y decrecientes y la magnitud a/Ξ representa la distancia típica en que la exponencial decreciente cambia a $1/e$ su valor anterior (y la exponencial creciente aumenta a un valor igual a e veces su valor anterior). Veremos dentro de poco que las exponenciales crecientes no pueden aparecer en sistemas muy largos, de manera tal que si se fuerza la primera parte móvil a la izquierda, la perturbación no puede avanzar muy lejos y casi no se va a notar a varias distancias típicas.

En todo sistema físico donde la periodicidad temporal de la fuerza impulsora produce una respuesta forzada estacionaria con comportamientos espaciales exponenciales (por lo general decrecientes) se dice que el sistema se comporta reactivamente. La conclusión es entonces que para fuerzas impulsoras con frecuencias $\omega > \omega_{Max}$, la cadena lineal se comporta reactivamente.

Notar que un sistema no es dispersivo o reactivo, si no que se comporta de una u otra manera de acuerdo con el valor de la frecuencia de la fuerza impulsora. También notemos que en el caso de la cadena lineal, el comportamiento reactivo desaparece en la aproximación continua. En la siguiente sección veremos otros sistema de partes móviles que en la aproximación continua exhibe los dos comportamientos, dispersivo y reactivo.

8.2. Forzado estacionario, Klein-Gordon

En vez de buscar la solución más general de (6.38) es instructivo estudiar cómo responde el sistema cuando: i) se obliga a que el primer péndulo se mueva con un movimiento oscilatorio armónico de frecuencia ω (con una fuerza externa, por ejemplo con un motor) y ii) se espera el tiempo suficiente como para que se apague el transitorio. Con la hipótesis ii) estamos repitiendo el esquema realizado en 8.1, de suponer que hay una pequeña disipación que no afecta el fenómeno que queremos estudiar (y que por eso despreciamos) y que esa pequeña disipación es necesaria para apagar el transitorio (y por eso no la despreciamos). Luego de pensar en situaciones prácticas en las que esta

ambivalencia puede resultar una buena aproximación, retomamos la idea de la sección anterior, de aprovechar para la respuesta forzada estacionaria las cosas que sabemos de modos normales. Y así seguir los pasos que hicimos en la página 103 para estudiar los modos normales de una cuerda. De esta manera, proponemos una solución como en (6.42), reemplazamos esta expresión en (6.38) y así llegamos la siguiente ecuación diferencial ordinaria para encontrar la función $A(x)$

$$\frac{d^2 A}{dx^2} + \underbrace{\frac{M}{Ka^2} (\omega^2 - \omega_0^2)}_{k^2} A = 0 . \quad (8.3)$$

Como esta ecuación es formalmente idéntica a la ec. (6.43), su solución debe ser formalmente idéntica a la dada en la ec. (6.44)

$$A(x) = B \sin(kx) + C \cos(kx) , \quad (8.4)$$

donde B y C son constantes. La diferencia fundamental entre el modelo de la cuerda y el modelo de los péndulos acoplados aparece en la ecuación de dispersión que relaciona k con ω : en el primer caso la relación de dispersión es lineal y viene dada por la ec. (6.45), mientras que en el segundo caso la relación de dispersión es

$$k = \sqrt{\frac{M}{Ka^2}} \sqrt{\omega^2 - \omega_0^2} \quad (8.5)$$

que claramente no es lineal.

8.2.1. Dispersivo vs reactivo

Aparte de la no linealidad, la nueva relación de dispersión muestra la existencia de dos regímenes de frecuencia donde los comportamientos físicos son muy distintos:

- **zona dispersiva**, $\omega > \omega_0$, k es una cantidad real, $A(x) = A(x + 2\pi/k)$, la periodicidad temporal de la fuerza impulsora induce una periodicidad espacial $\lambda = 2\pi/k$ en las amplitudes de las partes móviles.
- **zona reactiva**, $\omega < \omega_0$, $k = i\kappa$ es una cantidad imaginaria pura, con

$$\kappa = \sqrt{\frac{M}{Ka^2}} \sqrt{\omega_0^2 - \omega^2} , \quad (8.6)$$

los senos y cosenos en (8.4) pasan a ser senos y cosenos hiperbólicos y es conveniente reescribir $A(x)$ como una combinación lineal de funciones exponenciales reales

$$e^{-\kappa x} \quad y \quad e^{\kappa x} . \quad (8.7)$$

En este caso, la periodicidad temporal de la fuerza impulsora NO induce una periodicidad espacial en las amplitudes de las partes móviles. La cantidad $\delta = 1/\kappa$ es la distancia en la cual la exponencial decreciente cambia a $1/e$ su valor anterior y se conoce como *distancia de penetración* (también es la distancia en la cual la exponencial creciente aumenta a un valor igual a e veces su valor anterior).

Conclusión: la respuesta forzada de este sistema depende dramáticamente de la frecuencia de la fuerza impulsora. Cuando la periodicidad temporal de una excitación externa (como en este caso la fuerza impulsora, pero también podrían ser los campos electromagnéticos de una antena emisora, o el voltaje alterno de 50 Hz que sale del enchufe eléctrico) induce una periodicidad espacial, se dice que el medio excitado se comporta dispersivamente. En cambio cuando la periodicidad temporal de la excitación externa induce comportamientos exponenciales, se dice que el medio excitado se comporta reactivamente. Un medio no es ni reactivo ni dispersivo, sino que su comportamiento es de un tipo u otro dependiendo de la frecuencia de la excitación externa. Si $\omega = \omega_0$, $k = 0$, $\lambda = \infty$: todos los péndulos se balancean en fase y la fuerza restauradora es solamente gravitatoria.

8.2.2. La parte con $e^{\kappa x}$ no está permitida en sistemas infinitos

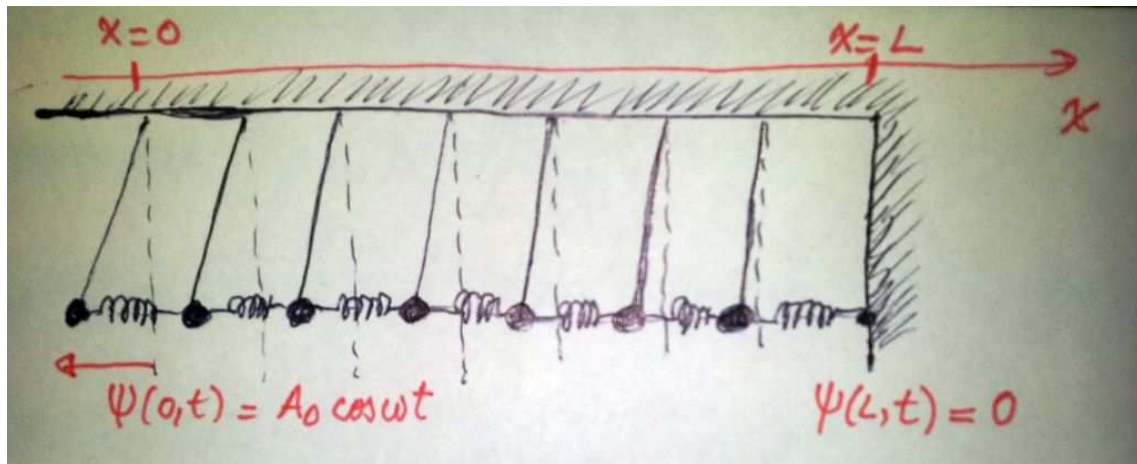


Figura 8.1: El sistema de péndulos idénticos se extiende desde $x = 0$ hasta $x = L$ y el último resorte está sujeto a una pared rígida.

El siguiente ejemplo, esquematizado en la figura 8.1, muestra que en el régimen reactivo, cuanto más grande es el sistema, menos influye la exponencial creciente en la combinación lineal de funciones exponenciales reales. El primer péndulo es forzado con un movimiento oscilatorio armónico de frecuencia ω y describe el movimiento $\psi(0,t) = A_0 \cos \omega t$. El último resorte está sujeto a una pared rígida en $x = L$, por lo tanto $\psi(L,t) = 0$ siempre. Si la frecuencia de la fuerza impulsora es menor que la frecuencia característica ω_0 , en el régimen estacionario las amplitudes $A(x)$ tienen la forma

$$A(x) = b e^{-\kappa x} + c e^{\kappa x}. \quad (8.8)$$

Las constantes b y c quedan determinadas cuando se ajusta (8.8) para que cumpla las condiciones en los dos extremos

$$\begin{aligned} b + c &= A_0, \\ b e^{-\kappa L} + c e^{\kappa L} &= 0 \implies c = -b e^{-2\kappa L} \xrightarrow{\kappa L \gg 1} 0. \end{aligned} \quad (8.9)$$

La segunda ecuación muestra que si el sistema mide varias distancias de penetración ($L \gg \delta$) $\implies |c| \ll |b|$. Para tener idea de órdenes de magnitud, si $L/\delta = 2$, $|c/b| \approx 0.14$ y si $L/\delta = 5$, $|c/b| \approx 0.007$. Resolviendo el sistema (8.9), resulta

$$\psi(x, t) = \frac{A_0}{1 - e^{-2\kappa L}} [e^{-\kappa x} - e^{-\kappa(2L-x)}] \cos \omega t. \quad (8.10)$$

Si $L \gg \delta$, $\psi(x, t) \approx A_0 e^{-\kappa x} \cos \omega t$ y se tiene una perturbación armónica con decaimiento espacial exponencial conocida como *onda evanescente*. En este caso resulta que en el régimen reactivo la fuerza impulsora es incapaz de mover partes móviles alejadas a distancias $x \gg \delta$. En este caso da lo mismo que el último resorte esté sujeto a una pared rígida o no, porque la perturbación de la fuerza impulsora no puede ser transmitida por el sistema.

8.2.3. Comentarios sobre el régimen reactivo

El régimen reactivo tiene aplicaciones en *filtros*: toda fuerza impulsora no armónica, Fourier mediante, se puede pensar como suma de funciones armónicas. Este sistema deja pasar solamente aquellas componentes de Fourier con frecuencias mayores que ω_0 , en cambio “rebotará” componentes de Fourier con frecuencias menores que ω_0 . Es lo que sucede en la ionósfera cuando se la intenta forzar electromagnéticamente. La ionósfera es la parte de la atmósfera terrestre ionizada permanentemente debido a la radiación solar. Y resulta que los campos electromagnéticos satisfacen en la ionósfera la ecuación de Klein-Gordon (6.38), ω_0 es la *frecuencia del plasma* y el factor que multiplica a la derivada segunda espacial es igual a la velocidad de la luz en el vacío al cuadrado. Para un plasma diurno la frecuencia del plasma es $\nu_p \approx 10 - 30$ MHz, por eso la ionósfera se comporta reactivamente para ondas de radio AM, con frecuencias $\nu \approx 1000$ KHz, mientras que se comporta dispersivamente para ondas de radio FM, con frecuencias $\nu \approx 100$ MHz. ¿Cómo se comporta la ionósfera terrestre para la luz?

Supongamos que a partir de la posición x_1 se produce un cambio en la longitud de los péndulos: a la izquierda de x_1 con longitud ℓ_1 y a la derecha con longitud ℓ_2 . La respuesta forzada estacionaria en cada tramo dependerá del valor de la frecuencia externa con respecto a $\omega_{01} = \sqrt{g/\ell_1}$ y $\omega_{02} = \sqrt{g/\ell_2}$. Si $\ell_2 < \ell_1$, $\omega_{01} < \omega_{02}$ (como en el sistema de péndulos del ejercicio 16 de la guía 1, que sería la versión discreta de lo que estamos haciendo acá). Si la frecuencia de la fuerza impulsora se elige para que $\omega_{01} < \omega < \omega_{02}$, la perturbación periódica causada por la fuente en el tramo izquierdo se transformará, a partir de x_1 , en una perturbación con decaimiento exponencial. Como no hay disipación y debido a que las perturbaciones con decaimiento exponencial no pueden entrar mucho en la zona $x > x_1$, la perturbación incidente “rebotará” y como

resultado se formará una onda estacionaria en la zona $x < x_1$, lo mismo que pasaba en un segmento finito de cuerda (independientemente de las condiciones en los extremos la energía se queda encerrada en el sistema). Comparar la solución del ejercicio 16 de la guía 1 con la solución que se obtendría en este caso en el continuo.

Si a partir de la posición $x_2 > x_1$ la longitud de los péndulos fuera de nuevo ℓ_1 , la región $x > x_2$ sería una zona con comportamiento dispersivo y en esta zona podría volver a aparecer una perturbación periódica, como la que había en el tramo izquierdo. Para que la perturbación en este tercer tramo sea apreciable se debería cumplir que $x_2 - x_1 \lesssim \delta$. Hay muchos fenómenos físicos en los que una perturbación se transmite a través de regiones “prohibidas” (como en el tramo entre x_1 y x_2 , que no puede ser traspasado si $x_2 - x_1 \gg \delta$). Cuando esto ocurre se dice que hay **tuneleo** de la perturbación, un efecto que aparece en óptica, en acústica y de manera espectacular en mecánica cuántica.

8.3. Forzado estacionario: péndulos acoplados

Para analizar el forzado estacionario del sistema discreto, partimos de (6.22) (la que nos llevó a Klein-Gordon). Ya dijimos que la solución procede formalmente igual que la de los modos. Aprovechemos lo que se hizo en la eq. (6.4) para el caso $\omega_0 = 0$, página 88. Es claro que si antes llegábamos a esta eq. en diferencias (ver (6.5))

$$\omega^2 = 2 \frac{K}{M} \left(1 - \frac{A_{n+1} + A_{n-1}}{2 A_n} \right), \quad (8.11)$$

ahora vamos a llegar a una ecuación en diferencias similar, pero con el nuevo término que corresponde a $\omega_0 \neq 0$

$$\omega^2 = \omega_0^2 + 2 \frac{K}{M} \left(1 - \frac{A_{n+1} + A_{n-1}}{2 A_n} \right), \quad (8.12)$$

que se puede interpretar recordando que el lado derecho de (8.12) da la fuerza restitutiva por unidad de desplazamiento y por unidad de masa cuando $\omega_0 = 0$ (no hay gravedad o péndulos muy largos). Cuando $\omega_0 \neq 0$, la fuerza restitutiva aumenta de la misma manera para todas las masas y se suma a la la fuerza restitutiva por unidad de desplazamiento y por unidad de masa que ya tenían las partes móviles sin gravedad.

Ya sabemos que el cociente entre amplitudes se puede hacer independiente de n con sucesiones de la forma

$$A_n = B \sin(n \Theta) + C \cos(n \Theta). \quad (8.13)$$

Así llegamos a la nueva relación de dispersión (ver eq. (6.10))

$$\omega^2 = \omega_0^2 + \underbrace{\frac{4 K}{M}}_{\omega_{MAX}^2} \sin^2 \frac{\Theta}{2}, \quad (8.14)$$

válida para la relación entre período temporal y período espacial, tanto en un modo como en el caso forzado estacionario.

Cuando $\omega = \omega_0$, $\Theta = 0$, todos los péndulos se balancean en fase (es decir, período espacial infinito, resultado ya obtenido en la aproximación continua, página 146) y la fuerza restitutiva por unidad de desplazamiento y por unidad de masa es solamente gravitatoria.

Cuando $\omega_0^2 \leq \omega^2 \leq \omega_0^2 + \omega_{MAX}^2$ siempre es posible encontrar un valor real de Θ que estará asociado con una periodicidad espacial. Vemos entonces que el rango dispersivo en el caso discreto tiene un límite superior

$$\sqrt{\omega_0^2 + \frac{4K}{M}} = \sqrt{\omega_0^2 + \frac{1}{a^2} \frac{4Ka^2}{M}} \xrightarrow{a \rightarrow 0} \infty \quad (8.15)$$

y que este límite superior tiende a infinito cuando nos aproximamos al caso continuo, resultado que ya habíamos obtenido con Klein-Gordon.

Rango reactivo bajo. Sería de esperar que la zona $\omega < \omega_0$ corresponda a un comportamiento reactivo. Y así es nomás. Una manera de comprobarlo es notar que la ec. (8.14) se satisface para valores imaginarios puros de $\Theta = i\Xi$, con Ξ real. En este caso los senos y cosenos de la ec. (8.13) se transforman en senos y cosenos hiperbólicos, el cociente entre amplitudes de la ec. (8.12) resulta igual a $\cosh \Xi$ y la relación de dispersión en esta zona se reescribe como

$$\omega^2 = \omega_0^2 - \frac{4K}{M} \sinh^2 \frac{\Xi}{2}. \quad (8.16)$$

Notar que en este rango tiene más sentido físico reescribir la ec. (8.13) en términos de combinaciones lineales de exponenciales crecientes y decrecientes del tipo $e^{\pm \Xi n}$

$$A_n = b e^{-\Xi n} + c e^{\Xi n}. \quad (8.17)$$

Cuando $\omega = \omega_0$ ya sabemos que todos los péndulos se balancean en fase, un resultado contenido en esta ecuación porque para esta frecuencia $\Xi = 0$, que corresponde a una distancia de penetración infinita. Cuando ω disminuye desde el valor ω_0 , Ξ aumenta, es decir que la distancia de penetración disminuye, hasta llegar a un valor máximo de Ξ cuando $\omega \rightarrow 0$. Nos damos cuenta que Ξ juega el papel de κa .

Rango reactivo alto. Ya nos dimos cuenta que la zona $\omega^2 \geq \omega_0^2 + \omega_{MAX}^2$ no corresponde a comportamientos dispersivos. Y también vimos que esta zona desaparece en el caso continuo. Nos damos cuenta que pasar algo muy parecido al caso $\omega_0 = 0$ que vimos para las oscilaciones transversales de la cadena lineal, aunque en vez de zig-zag deberíamos decir vaivén, porque los péndulos realizan movimientos longitudinales. Es claro que las sucesiones de la forma de la ec. (8.17) no sirven, porque dan frecuencias fuera del rango que nos interesa. Pero sí a través de un $(-1)^n$ intentamos con un vaivén

entre partes móviles sucesivas

$$A_n = (-1)^n \left(b e^{-\Xi n} + c e^{\Xi n} \right), \quad (8.18)$$

es fácil comprobar que ahora el cociente entre amplitudes de la ec. (8.12) resulta igual a $-\cosh \Xi$ y la relación de dispersión en esta zona se reescribe como

$$\omega^2 = \omega_0^2 + \frac{4K}{M} \cosh^2 \frac{\Xi}{2}. \quad (8.19)$$

que siempre tiene soluciones Ξ reales para frecuencias $\omega \geq \sqrt{\omega_0^2 + \omega_{MAX}^2}$.

Capítulo 9

Descripción geométrica

En los ejemplos mecánicos vistos en los capítulos anteriores, los comportamientos ondulatorios se pudieron mostrar a partir de estudiar la dinámica impuesta en cada parte móvil por las fuerzas restitutivas del sistema. Como la dinámica de sistemas mecánicos clásicos está gobernada por las ecuaciones de Newton, la dificultad quedaba acotada a resolver matemáticamente las ecuaciones diferenciales acopladas o la ecuación de ondas resultante. Pero para estudiar fenómenos físicos tan familiares para la humanidad como el sonido o la luz (tan familiares que nacemos con detectores para ellos) el camino recorrido no ha sido éste, sino un camino inverso, que arranca de la fenomenología y con análisis, sistematización, deducción y comprobación permite llegar a leyes y teorías, aún sin conocer los principios fundamentales de la dinámica de las partes móviles.

La descripción de comportamientos ondulatorios que veremos a continuación está formulada en base a conceptos puramente geométricos que no hacen ninguna referencia a la ecuación de ondas ni a la naturaleza ondulatoria del fenómeno que se estudia (como óptica o acústica) y recibe el nombre de descripción geométrica (DG). Es una aproximación válida cuando la onda interactúa con objetos grandes comparados con la longitud de onda, es decir

$$\lambda \ll D, \tag{9.1}$$

con λ la longitud de onda y D el menor tamaño característico del objeto con el que interactúa la onda.

¿Cuál es el interés de la descripción geométrica? Que en su rango de validez permite describir de manera muy sencilla la *cinemática* (cómo viaja la onda) en procesos ondulatorios como la luz o el sonido. Sin embargo, no dice nada sobre la *dinámica* de la interacción onda–objeto, que requiere conocer las leyes físicas que están detrás de cada onda. En otras palabras, la descripción geométrica permite predecir qué caminos siguen las ondas que se reflejan y transmiten en un dado obstáculo –y estos caminos serían iguales para todo tipo de onda, siempre que valga la condición (9.1)– pero no dice nada sobre qué fracción de la energía incidente se va por el camino reflejado y qué fracción por el camino transmitido.

La construcción teórica de la DG sigue históricamente los avances del hombre en el

entendimiento de la luz, por este motivo muchos libros la llaman *óptica geométrica*, si bien vale para cualquier tipo de onda. La luz es una onda electromagnética transversal con longitudes de onda en el rango $400 \text{ nm} < \lambda < 700 \text{ nm}$ ($1 \text{ nm} = 10^{-9} \text{ m}$). Luego, para la mayoría de los objetos de la vida cotidiana, se satisface ampliamente la condición (9.1). En cambio, para las ondas sonoras en aire ($\lambda \approx 1 \text{ m}$ para el λ_{440}), en la vida cotidiana por lo general no se satisface la condición (9.1) y en consecuencia no es aplicable la DG (o sea, que para que valga la DG hay que ir a otros rangos de frecuencia, es decir, de longitudes de onda).

9.1. Rayos

Un rayo se define como la línea que corresponde a la dirección del flujo de energía. Lo podemos imaginar como la línea recta que une la salida de un puntero laser con la manchita producida por el laser en un papel. Y se podría visualizar perfectamente si en la habitación donde se hace el experimento hay humo o polvo flotando en el aire. El concepto ideal de rayo se obtendría tratando de disminuir el tamaño de la manchita, por ejemplo limitando la apertura del puntero. Pero cuando la apertura resulta muy pequeña, la condición (9.1) deja de valer y a partir de este momento el procedimiento para visualizar el rayo ideal deja de funcionar: el tamaño de la manchita deja de disminuir y para describir lo que sucede hay que tener en cuenta el carácter ondulatorio, ya no se puede aplicar la DG y en estas circunstancias se dice que hay *difracción*.

Nos damos cuenta que los rayos, como caminos de las ondas, se corresponden con los vectores de onda $\vec{k}$ asociados con las ondas 1D (ver pag. 126 y siguientes). Y entonces son perpendiculares a los frentes de onda: rayos paralelos se corresponden con ondas planas, rayos radiales se corresponden con ondas esféricas, etc.

9.2. Leyes fenomenológicas

Los experimentos demuestran que el comportamiento de los rayos se puede condensar en las siguiente leyes:

- en un medio homogéneo, los rayos son líneas rectas (ley cero);
- como se muestra en el esquema de la Fig. 9.1, en una superficie de separación entre dos medios homogéneos el rayo incidente da lugar a un rayo reflejado en el primer medio y a un rayo transmitido en el segundo medio;
- el rayo reflejado y el rayo transmitido están contenidos en el plano determinado por el rayo incidente y la normal a la superficie de separación. Este plano se llama *plano de incidencia*. Notar que en el caso de incidencia normal ($i = 0$) el plano de incidencia no está bien definido, porque hay infinitos planos que contienen al rayo incidente y a la normal.

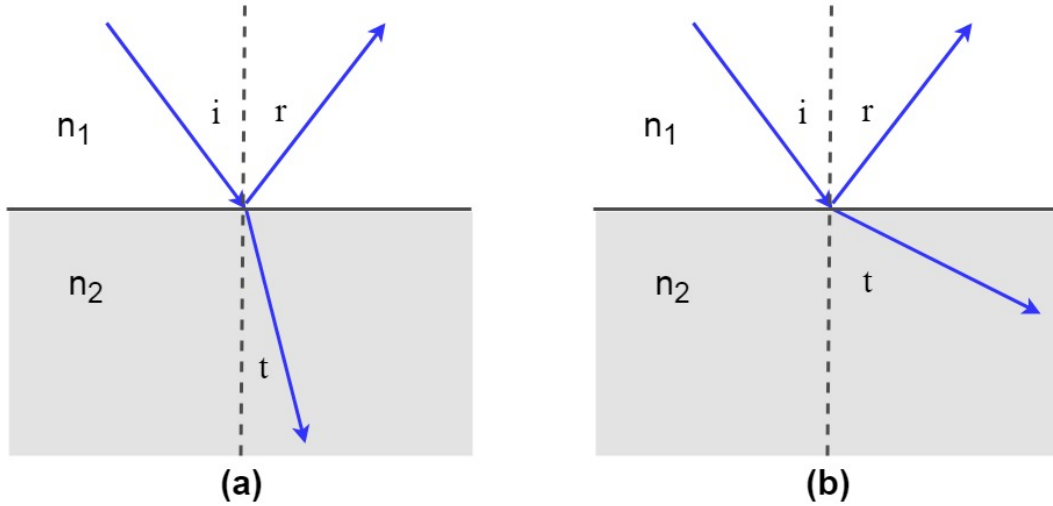


Figura 9.1: En la superficie de separación entre los medios 1 y 2, los experimentos demuestran que los rayos incidente, reflejado y refractado están contenidos en el plano determinado por el rayo incidente y la normal a la superficie (línea punteada) y las relaciones entre los ángulos i , r y t .

- $r = i$, es decir que el ángulo r que forma el rayo reflejado con la normal es igual al ángulo i que forma el rayo incidente con la normal;
- el cociente entre el seno del ángulo t que forma el rayo transmitido con la normal y el seno del ángulo i que forma el rayo incidente con la normal sólo depende del par de medios y no del ángulo de incidencia. Esta ley, descubierta por [Ibn Sahl](#), también se conoce como ley de Snell, de Descartes o ley de la refracción y se expresa

$$\frac{\sin i}{\sin t} = n_{21}. \quad (9.2)$$

La constante n_{21} se llama *índice de refracción relativo* entre el medio 2 y el medio 1. Si el medio 2 se reemplaza por un tercer medio, se verifica experimentalmente que valor del cociente entre senos independiente del ángulo de incidencia toma ahora otro valor n_{31} . Si se vuelve a hacer el experimento en la superficie de separación entre los medios 2 y 3, ahora el cociente entre senos resulta $n_{32} = n_{31}/n_{21}$ y esta propiedad sugiere tomar algún medio como referencia, entendiendo que las constantes involucradas en la ley de la refracción están medidas en relación a dicho medio: $n_{21} = n_2/n_1$. Así, la ley se escribe

$$n_1 \sin i = n_2 \sin t. \quad (9.3)$$

De la condición (9.1) se desprende que en esta aproximación todas las superficies son localmente planas en la zona donde incide el rayo, es decir que los radios de curvatura de la superficie son grandes comparados con la longitud de onda. Con experimentos caseros es posible poner en evidencia ciertos comportamientos luminosos que no responden a las leyes fenomenológicas de la DG. Por ejemplo basta con iluminar superficies, como la de un DVD o de ciertas tarjetas bancarias, que tienen detalles que el ojo humano no puede resolver. Por ejemplo, la distancia radial entre dos surcos consecutivos en la

superficie de un DVD es aproximadamente 740 nm. Además, a lo largo de cada surco hay montañitas de 320 nm de ancho y como mínimo 400 nm de largo y 120 nm de alto. Estas cifras muestran que para estas superficies no se cumple la condición de validez de la DG, porque la menor distancia característica D es del mismo orden de magnitud que λ .

9.3. Algunas consecuencias de la ley de Ibn Sahl

9.3.1. Reflexión externa e interna

En la figura 9.1a se muestra el esquema correspondiente al caso en que el índice de refracción del medio de incidencia es menor que el índice de refracción del medio de transmisión. En esta situación, conocida como *reflexión externa*, el rayo refractado *se acerca a la normal* y siempre hay un rayo refractado para todo valor del ángulo de incidencia.

En cambio, en la figura 9.1b se muestra el caso opuesto al anterior, cuando el índice de refracción del medio de incidencia es mayor que el índice de refracción del medio de transmisión. En esta situación, conocida como *reflexión interna*, el rayo refractado *se aleja de la normal*, tal como se pone en evidencia cuando despejamos $\sin t$ en función de $\sin i$. Si tenemos en cuenta que en el intervalo $[0, \pi/2]$, intervalo donde están definidos los ángulos i y t , el seno es una función creciente de su argumento, resulta

$$\sin t = \frac{n_1}{n_2} \sin i \implies \begin{cases} t < i & \text{si } n_1 < n_2, \\ t > i & \text{si } n_1 > n_2. \end{cases} \quad (9.4)$$

9.3.2. Reflexión total

Notar que en el caso $n_1 > n_2$ existe un ángulo crítico $i_c < \pi/2$ para el cual el rayo refractado es paralelo a la superficie $t = \pi/2$. Este ángulo se llama *ángulo crítico de reflexión total* y se puede calcular a partir de la ley de Ibn Sahl

$$n_1 \sin i_c = n_2 \sin \pi/2 \implies \sin i_c = \frac{n_2}{n_1}. \quad (9.5)$$

En este caso la ley de Ibn Sahl se puede reescribir como

$$\sin t = \frac{\sin i}{\sin i_c}. \quad (9.6)$$

Esta expresión muestra claramente que cuando

$$i > i_c \implies \sin t > 1$$

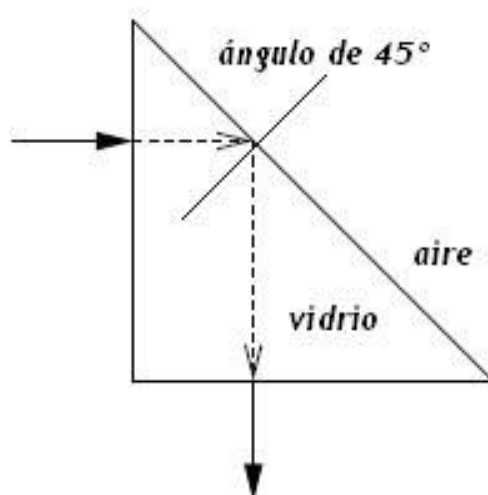


Figura 9.2: Esquema de funcionamiento del prisma de reflexión total. La sección del prisma tiene forma de triángulo rectángulo isósceles. Los rayos que inciden normalmente a uno de los catetos llegan a la hipotenusa con un ángulo mayor al que corresponde a i_c para la reflexión interna entre vidrio y aire.

una situación que no es nada clara en el lenguaje de rayos, porque no sabemos cómo dibujar ángulos con seno mayor que 1. Esto suena bastante parecido a lo que sucedía para comportamientos reactivos. Justamente, más adelante reexaminaremos el fenómeno con el lenguaje de las ondas y encontraremos que para $i > i_c$, el medio de transmisión se comporta reactivamente y entonces el rayo reflejado lleva toda la energía del rayo incidente. También encontraremos que para $i < i_c$ el medio de transmisión se comporta dispersivamente y la energía del rayo incidente en parte se refleja y en parte se transmite, tal como sucede para todo ángulo de incidencia en el caso de reflexión externa $n_1 < n_2$.

Si el rayo va de vidrio ($n_1 = 1.5$) hacia aire ($n_2 = 1$), entonces $\sin i_c = 1/1.5 = 2/3$ y resulta $i_c \approx 42^\circ$, un resultado que explica el diseño del prisma de reflexión total mostrado en la figura 9.2, donde se utiliza un prisma con forma de triángulo rectángulo isósceles para que el ángulo de incidencia en la hipotenusa resulte $i = 45^\circ > i_c$ y así entonces el rayo transmitido en el primer cateto se refleja totalmente.

El mecanismo de reflexión total permite el confinamiento y la propagación de luz en [fibras ópticas](#). Esencialmente la fibra consiste en un núcleo con un recubrimiento, con el índice de refracción del núcleo mayor que el índice del recubrimiento. De esta manera, la luz se propaga por el interior de la fibra con un ángulo de reflexión por encima del ángulo límite de reflexión total, rebotando en las superficies núcleo-recubrimiento.

9.4. La DG en otros marcos teóricos

Cuando estudiamos el problema de la cuerda mediante las descripciones de d’Alambert y de modos observamos que distintas descripciones de la misma situación pueden pro-

ducir ideas físicas y métodos matemáticos de análisis diferentes y complementarios. Las leyes de la descripción geométrica también admiten teorías duales que pueden aplicarse como alternativa a las leyes fenomenológicas enunciadas en la página 154. Las descripciones alternativas que veremos están basadas en

- Principio de Huygens;
- Principio de Fermat;
- Marco ondulatorio riguroso.

Cada descripción tiene su ventaja, en cada una hay ideas físicas muy importantes que toman protagonismo y si estamos en el rango de validez de la DG, da lo mismo cuál se usa.

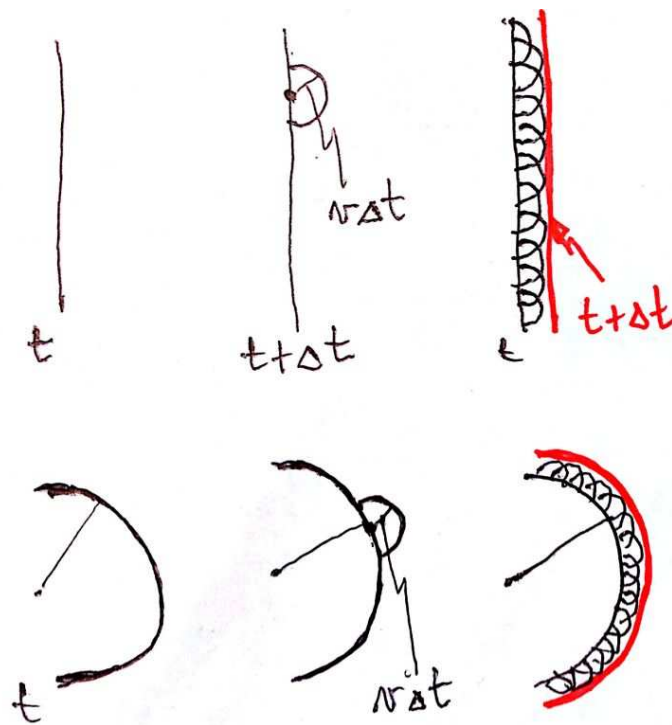


Figura 9.3: Esquema para obtener la propagación de frentes de onda planos y esféricos. Primero se muestra el frente de onda en el instante t . Luego una ondita secundaria en el instante $t + \Delta t$. Y finalmente la envolvente de todas las onditas secundarias en el instante $t + \Delta t$.

9.4.1. Principio de Huygens

En 1678 el físico holandés Christiaan Huygens propuso un mecanismo que permite explicar no solo la propagación de los rayos (normales a los frentes de onda) sino también las leyes fenomenológicas enunciadas en la página 154. El principio postula que todo punto de un frente de ondas actúa como fuente de onditas esféricas secundarias de igual velocidad y frecuencia que la onda original y que el frente de ondas en un instante

posterior es la envolvente de todas estas onditas esféricas (onditas de Huygens). Por envolvente se entiende la superficie tangente. Para que el mecanismo funcione, Huygens supuso, sin explicaciones adicionales, que las onditas secundarias sólo viajan “hacia adelante”. El principio fue reformulado en 1815 por Fresnel para: i) hacer desaparecer las onditas que viajan “hacia atrás”, introduciendo un factor de oblicuidad que depende de la dirección de propagación; y ii) interpretar la envolvente en términos de interferencia entre las onditas secundarias.

La figura 9.3 muestra como aplicar el principio de Huygens para obtener la propagación de un frente de onda plano (arriba) y de un frente de onda esférico (abajo). En la primera columna se muestra el frente en el instante t y en la segunda columna una sola ondita secundaria en el instante $t + \Delta t$, notar que su frente de onda tiene un radio $v\Delta t$. En la tercera columna se muestra la envolvente de todas las onditas secundarias en el instante $t + \Delta t$. Se ha supuesto que el medio es isótropo y homogéneo, por eso todas las onditas secundarias tienen el mismo radio, si el medio no fuera homogéneo la velocidad dependería del punto, las onditas secundarias tendrían distinto radio y los frentes de onda no serían como los originales. Como los rayos son perpendiculares a los frentes de onda, es evidente que estas construcciones equivalen a la ley cero de la página 154.

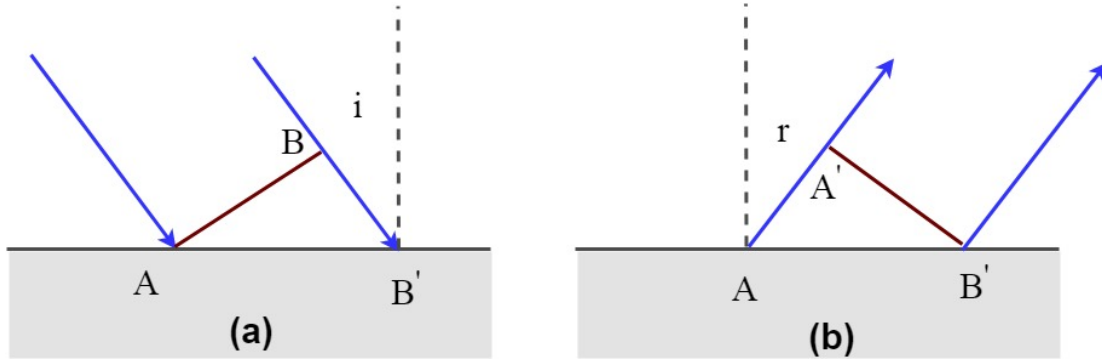


Figura 9.4: Esquema para deducir la ley de la reflexión a partir del principio de Huygens.

El Principio de Huygens también contiene la ley de la reflexión y la ley de Ibn Sahl. Para verlo, consideremos primero el esquema de la Fig. 9.4. La parte (a) de la figura representa el frente de una onda plana con rayos que forman un ángulo i con respecto a la normal a la superficie de separación. En el instante t , cuando el punto A del frente de ondas incidente toca a la superficie, genera una ondita esférica elemental reflejada que vuelve hacia el medio 1. Lo mismo hacen, en instantes posteriores, los otros puntos del frente de ondas $\overline{AB}$. Consideremos el instante $t + \Delta t$, cuando el punto B del frente de ondas incidente llega a la superficie. En ese instante, ver parte (b), la ondita esférica elemental reflejada en el punto A tiene radio

$$\overline{AA'} = v_1 \Delta t$$

mientras que la ondita esférica elemental reflejada en el punto B' recién está por empezar a emitir. Esto indica que el frente de ondas reflejado es el segmento $\overline{A'B'}$ y que por hipótesis también $\overline{BB'} = v_1 \Delta t$. Como

$$\text{sen } i = \overline{BB'}/\overline{AB'} \quad \text{y} \quad \text{sen } r = \overline{AA'}/\overline{AB'}$$

entonces debe ser $\sin i = \sin r$, luego $i = r$ y queda demostrada la ley de la reflexión.

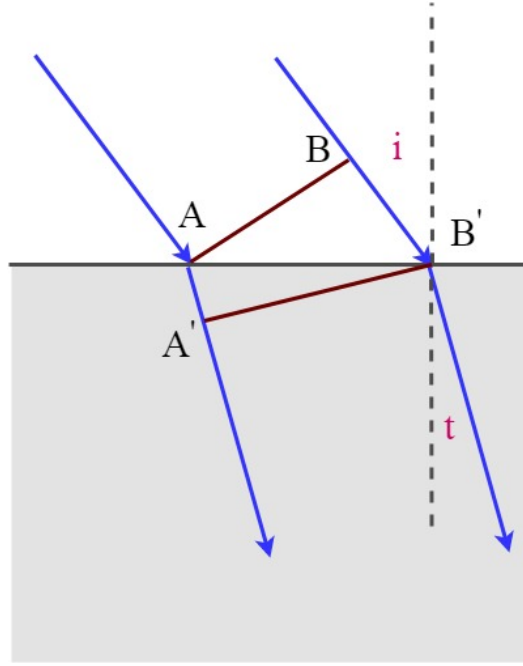


Figura 9.5: Esquema para deducir la ley de Ibn Sahl a partir del principio de Huygens.

Para deducir la ley de refracción a partir del principio de Huygens, consideremos ahora el esquema de la Fig. 9.5 y supongamos que v_1 y v_2 son las velocidades de fase correspondientes a los medios de incidencia y de transmisión respectivamente. El segmento $\overline{AB}$ representa, como antes, el frente de ondas de una onda plana con rayos que forman un ángulo i con respecto a la normal a la superficie de separación. En el instante t , cuando el punto A del frente de ondas incidente toca a la superficie, genera una ondita esférica elemental refractada que empieza a penetrar en el medio 2. Lo mismo hacen, en instantes posteriores, los otros puntos del frente de ondas $\overline{AB}$. Consideremos el instante $t + \Delta t$, cuando el punto B del frente de ondas incidente llega a la superficie, condición que impone que

$$\overline{BB'} = v_1 \Delta t.$$

En este instante, la ondita esférica elemental refractada a partir del punto A tiene radio

$$\overline{AA'} = v_2 \Delta t$$

mientras que la ondita esférica elemental refractada en el punto B' recién está por empezar a emitir (notar que en el esquema se supuso $v_1 > v_2$). Esto indica que el frente de ondas refractado es el segmento $\overline{A'B'}$ y como

$$\sin i = \overline{BB'} / \overline{AB'} \quad \text{y} \quad \sin t = \overline{AA'} / \overline{AB'}$$

entonces debe ser

$$\frac{\sin i}{\sin t} = \frac{\overline{BB'}}{\overline{AA'}} = \frac{v_1}{v_2}.$$

Esta expresión muestra que el cociente entre los senos solamente depende de los medios involucrados y no del ángulo de incidencia y equivale a la ley de Ibn Sahl escrita en la forma (9.2). El punto importante, además de entender cómo aplicar el mecanismo de Huygens cuando hay una discontinuidad, es notar que con este mecanismo Huygens permite interpretar la constante que aparece en la ley fenomenológica (9.2) como un cociente entre velocidades de fase de la onda en cada medio. Por eso la notación moderna escribe esta constante como un cociente entre índices de refracción, cantidades que para cada medio están definidas como cocientes entre la velocidad de fase en un cierto medio de referencia y la velocidad de fase en el medio en cuestión. De esta manera

$$\frac{1}{v_1} \sin i = \frac{1}{v_2} \sin t$$

que demuestra que, admitiendo las hipótesis de Huygens, es posible medir la velocidad relativa de la luz en dos medios distintos con sólo medir los ángulos de incidencia y refracción en una superficie de separación entre ambos medios.

Eligiendo convenientemente la forma de las superficies y los materiales, es posible diseñar elementos que introduzcan localmente desfases (retrasos o adelantos) en los frentes de ondas, ya sea mediante reflexiones o mediante refracciones. Por ejemplo, las antenas parabólicas transforman, mediante reflexión en la superficie de un paraboloide de revolución, un frente de ondas plano en un frente de ondas que converge en el foco de la parábola. El resultado se puede demostrar aplicando la ley de la reflexión en cada punto del paraboloide y para cada rayo incidente, o estudiando los desfases de las onditas elementales de Huygens y luego calculando la envolvente.

El esquema de la figura 9.6 muestra un elemento refractivo (lente convergente) que transforma el frente de onda plano $\overline{AB}$ de una onda incidente y lo convierte en el frente de ondas $\widehat{AB}$ con forma de superficie esférica con centro en C. En el lenguaje de rayos, la lente transforma un haz de rayos paralelos en un haz convergente.

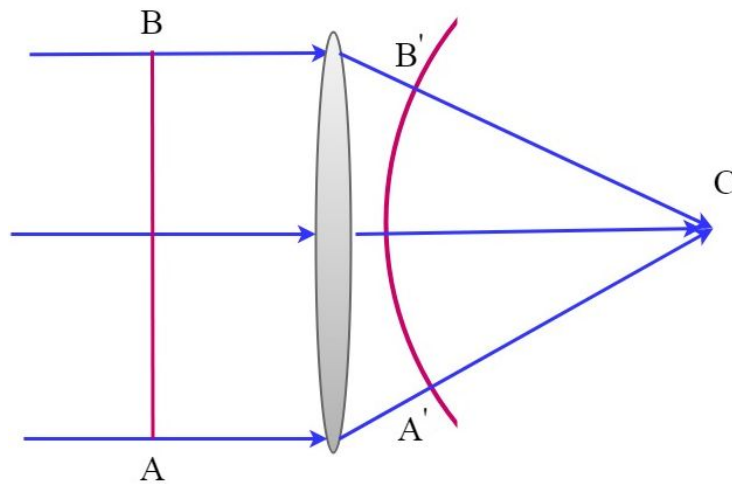


Figura 9.6: Un elemento refractivo (lente) transforma, mediante refracciones en dos superficies, un frente incidente plano en un frente de ondas esférico con centro en C (esto es, un haz de rayos paralelos en un haz convergente).

9.4.2. Principio de Fermat

Es un principio variacional para encontrar el camino que hace un rayo entre dos puntos (A y B). Es evidente que entre dos puntos hay infinitos caminos, algunos muy largos y otros más cortos. Matemáticamente, cada camino es una curva en el espacio 3D, que empieza en A y termina en B. ¿Cuál es el “verdadero” camino que “elige” un rayo que parte de A y llega a B y cumpliendo ciertas condiciones adicionales, como por ejemplo pasar por un espejo? Justamente el principio de Fermat da una receta para encontrar este camino.

El principio de Fermat se formula en términos de una magnitud llamada *longitud de camino óptico*, o directamente camino óptico a secas. La magnitud camino óptico está asociada con *dos puntos* (A y B) *y una trayectoria* entre ambos puntos. No tiene sentido hablar de camino óptico sin especificar dos puntos y una trayectoria.

En la figura ?? se esquematizan distintas trayectorias entre dos puntos A y B.

Definámosla. Dada la trayectoria j entre los puntos A y B, el camino óptico asociado a dicha trayectoria es la integral de camino del índice de refracción

$$[AB]_j = \int_j n(\ell) d\ell , \quad (9.7)$$

con $d\ell$ el elemento de arco en la trayectoria j .

Si el medio es homogéneo, el índice de refracción no depende del punto de la trayectoria. Entonces, la longitud de camino óptico asociada a una trayectoria entre dos puntos en un medio homogéneo es igual al producto del índice de refracción por la longitud de la trayectoria. Si además la trayectoria es rectilínea, $[AB]_{recta} = n \overline{AB}$, es decir el camino óptico es el producto del índice de refracción por la distancia entre A y B.

Usando la definición de índice de refracción, se ve que la longitud de camino óptico asociada con la trayectoria j resulta proporcional al tiempo que tardaría el rayo en recorrer dicha trayectoria j

$$[AB]_j = \int_j n(\ell) d\ell = c \int_j \frac{1}{v(\ell)} d\ell = c \int_j dt = c t_{AB} . \quad (9.8)$$

Y ahora que entendemos qué es el camino óptico asociado a una trayectoria entre dos puntos A y B, estamos en condiciones de formular el Principio de Fermat. Que dice así: *la trayectoria que siguen los rayos para ir de A a B es aquella para la cual el camino óptico es estacionario con respecto a variaciones de la trayectoria*. Y se debe recordar que dichas variaciones de la trayectoria deben ser compatibles con las condiciones del problema, por ejemplo en un problema de reflexión las variaciones de la trayectoria no pueden incluir caminos que dejen de pasar por el espejo.

Claramente este principio contiene la propagación rectilínea en un medio homogéneo, porque la única trayectoria con camino óptico estacionario, es decir que tenga camino óptico a primer orden igual que los caminos cercanos, es la trayectoria rectilínea, que

tiene camino óptico mínimo. Máximo no hay porque es claro que el camino óptico en este caso no está acotado.

El Principio de Fermat también contiene la ley de la reflexión y la ley de Ibn Sahl ...completar, por el momento ver <http://bit.ly/10GHKVY>.

Usando un espejo elíptico y sus deformaciones, vimos un caso donde el camino que predice el Principio de Fermat no es único, otro caso donde es único y es un máximo y por supuesto el conocido caso donde es único y es un mínimo. Un punto a tener en cuenta, hay textos que siguen diciendo que el Principio de Fermat es un principio de longitud de camino óptico mínimo.

El principio de Fermat no puede ser la causa, pues de lo contrario estaríamos atribuyendo conocimiento a la naturaleza ...

Claude Clerselier (1622)

9.5. Formación de imágenes

Los objetos luminosos (emisores o iluminados) se comportan como una colección de fuentes puntuales y cada fuente emite rayos divergentes. Ya vimos que si una fuente puntual se coloca en el foco de un espejo con forma de elipsoide, todos los rayos reflejados en el espejo convergen en el otro foco. En este caso se ha logrado transformar un haz de rayos divergentes con centro en un foco, en un haz de rayos convergentes con centro en el otro foco y se dice que el segundo foco es la imagen del primer foco.

Se llama sistema óptico a un conjunto de superficies reflectoras y refractoras. Si el sistema óptico transforma un haz de rayos divergentes desde el punto S en un haz de rayos convergentes en el punto S' , se dice que el sistema óptico forma en el punto S' una **imagen perfecta** del punto S . Excepto la energía perdida por efectos no deseados, como reflexiones inevitables cuando se usan sistemas refractores, **toda la energía** existente en el haz que incide en el sistema óptico llega al punto S' . S y S' también reciben el nombre de **puntos conjugados**, porque de acuerdo con la reversibilidad implícita en el principio de Fermat, una fuente puntual colocada en S' debe dar una imagen perfecta en S . El principio de Fermat también asegura que todas las trayectorias posibles que unen S y S' tienen el mismo camino óptico.

Por lo general la función de un sistema óptico es recoger una porción del frente de onda incidente (un haz de rayos) para formar una imagen, es decir que el sistema óptico se diseña para que el camino óptico de todas las trayectorias posibles entre un punto emisor S y otro punto S' tenga siempre el mismo valor. Desde este punto de vista, la formación de imágenes es la manipulación controlada de los rayos (o de los frentes de onda).

Cuando a la salida del sistema óptico todos los rayos se cortan en un punto, se dice que la imagen es **real**. En cambio, cuando las prolongaciones de los rayos que salen del sistema óptico se cortan en un punto, se dice que la imagen es **virtual**.

9.5.1. Dioptra plana

El caso más sencillo de sistema óptico, una superficie plana que separa dos medios de índices n_1 y n_2 . Para estudiar la formación de imágenes en sistemas ópticos en los libros hay varias convenciones de signo, acá usamos la convención más natural en física: tomar distancias positivas hacia la derecha y negativas hacia la izquierda tanto para los objetos como para las imágenes. Suponemos que la luz viaja de izquierda a derecha. La dioptra plana está en $x = 0$ (es el plano yz) y la fuente puntual S tiene coordenadas $(s, 0, 0)$ ($s < 0$ si la luz viene de izquierda a derecha). El rayo que sale de S e incide en la dioptra en el punto $(0, h, 0)$ forma un ángulo i con la normal y se refracta en el semiespacio $x > 0$ formando un ángulo t con la normal. Es claro que para cada i hay un h distinto. Y el rayo que sale de S e incide en la dioptra en el punto $(0, 0, 0)$ no se desvía cuando se refracta. Estos dos rayos no se cortan en el semiespacio $x > 0$, pero si en el semiespacio $x < 0$, en el punto $(s', 0, 0)$, $s' < 0$. En un sistema refractivo, $s' < 0$ equivale a que la imagen es virtual. Pero no sabemos si forma imagen. Para que esta dioptra forme imagen por transmisión, la distancia $s' < 0$ debe ser independiente del ángulo i . Y como $\operatorname{tg} i = h/s$ y $\operatorname{tg} t = h/s'$, resulta que

$$s' = \frac{\operatorname{tg} i}{\operatorname{tg} t} s = \frac{n_2 \cos t}{n_1 \cos i} s, \quad (9.9)$$

que claramente depende del ángulo i : el sistema no forma imagen perfecta. Pero si el sistema solamente recoge rayos que viajan cerca del eje, es decir, ángulos $i \ll 1$, entonces con muy buena aproximación se pueden poner los cosenos igual a 1, y resulta que con muy buena aproximación se puede decir que el sistema forma imagen, ubicada a la distancia

$$s' \approx \frac{n_2}{n_1} s. \quad (9.10)$$

Esta aproximación de suponer que el sistema solamente recoge rayos que viajan cerca del eje, es decir, ángulos $i \ll 1$, se llama **aproximación paraxial**, o también **aproximación de primer orden**, porque equivale a aproximar la ley de Ibn Sahl por $n_1 i \approx n_2 t$, es decir, aproximar los senos por los ángulos. Estrictamente hablando, la fuente puntual S no produce una imagen puntual S' , si no una mancha, que será más grande cuanto mayor sea la apertura de los rayos.

La formación de imágenes por reflexión se estudia de manera similar, en vez de seguir el rayo refractado en $(0, h, 0)$ hay que seguir el rayo reflejado. Que forma con la normal el mismo ángulo que el rayo incidente, pero con la orientación contraria. La ley de la reflexión es entonces $r = -i$, que es como la ley de Ibn Sahl aproximada $n_1 i \approx n_2 t$ pero con $n_1 = -n_2$ y con r en vez de t . Y además vale en forma exacta, es decir que el espejo plano forma imagen perfecta ubicada a una distancia

$$s' = -s, \quad (9.11)$$

que se obtiene de (9.10) con $n_1 = -n_2$. Notar que en un sistema reflector, $s' > 0$ indica que la imagen es virtual.

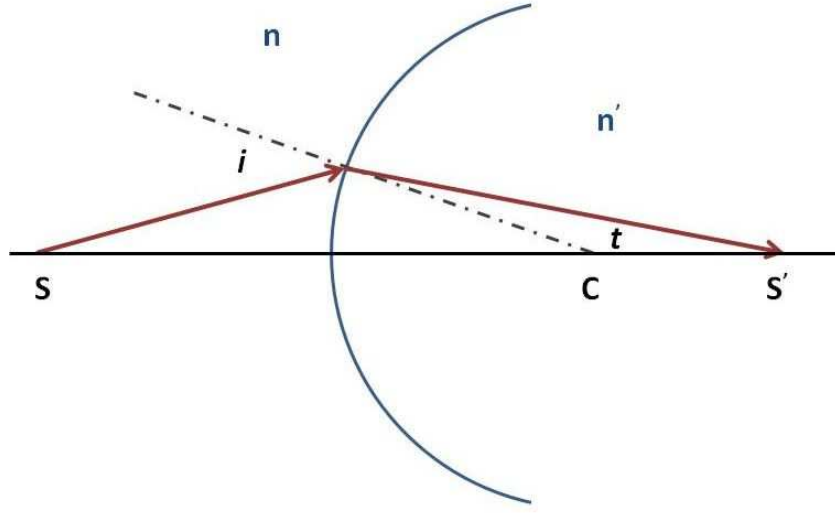


Figura 9.7: Refracción en dioptra esférica, radio R ($R > 0$ en el esquema), C es el centro de curvatura.

9.5.2. Dioptra esférica

Por tener curvatura constante, las dioptras esféricas son fáciles de fabricar. La mayoría de las lentes para corregir defectos de visión son esféricas y están diseñadas para trabajar en la zona paraxial. Ver esquema para la formación de imágenes por transmisión en la figura 9.7. La fuente puntual S tiene coordenadas $(s, 0, 0)$, $s < 0$ en el dibujo. El rayo que sale de S e incide en la dioptra con un ángulo i con la normal se refracta en el semiespacio $x > 0$ formando un ángulo t con la normal. Ahora la normal va del centro de curvatura C al lugar donde incide el rayo. El rayo que sale de S e incide en la dioptra en el punto $(0, 0, 0)$ no se desvía cuando se refracta. En el dibujo, estos dos rayos se cortan en el semiespacio $x > 0$ en el punto $(s', 0, 0)$, $s' > 0$. Esto es así porque se está suponiendo que $n' > n$ y que la dioptra es convexa (vista desde el lado que viene la luz). En cambio, con la misma curvatura, si $n' < n$, es claro que el rayo se habría alejado de la normal y que $s' < 0$. En la dioptra refractiva, $s' > 0$ indica que la imagen es real. Pero para que esta dioptra realmente forme imagen por transmisión, se debe cumplir que esta distancia $s' < 0$ sea independiente del ángulo i . Y resulta que esto se cumple en la aproximación paraxial. Queda como ejercicio demostrar que la siguiente ley

$$\frac{n'}{s'} - \frac{n}{s} = \frac{n' - n}{R}, \quad (9.12)$$

relaciona s y s' con todos los parámetros físicos de la dioptra. El lado derecho de esta igualdad se llama **potencia** Φ de la dioptra.

Si la imagen de un objeto está en infinito, el objeto está en el foco objeto y su distancia a la dioptra se llama distancia focal objeto, $s = f$.

Se define foco imagen a la posición donde se forma la imagen de un objeto que está en infinito, su distancia a la dioptra se llama distancia focal imagen, $s' = f'$. De acuerdo

con estas definiciones

$$\frac{n'}{s'} - \frac{n}{s} = \frac{n' - n}{R} = \frac{n'}{f'} = -\frac{n}{f} = \Phi, \quad (9.13)$$

Si $f' > 0$ la dioptra es convergente, si $f' < 0$ divergente. Notar que no sólo depende de que sea cóncava o convexa, también depende de la diferencia de índices de refracción.

El aumento lateral de una dioptra se define como el cociente entre la altura de imagen y objeto. Trazando el rayo que sale del extremo del objeto y pasa por el vértice de la dioptra, es fácil demostrar que el aumento lateral de la dioptra esférica vale

$$m = \frac{h'}{h} = \frac{n s'}{n' s}. \quad (9.14)$$

La formación de imágenes por espejos esféricos se estudia de manera similar a como se hizo para espejos planos: hay que seguir el rayo reflejado. Que forma con la normal un ángulo $r = t$, que el rayo incidente, pero con la orientación contraria. La ley de la reflexión es entonces $r = t$, que es como la ley de Ibn Sahl aproximada $n_1 i \approx n_2 t$ pero con $n_1 = -n_2$. Vemos entonces que la ecuación para espejos esféricos resulta

$$\frac{1}{s'} + \frac{1}{s} = \frac{2}{R} = \frac{1}{f'} = \frac{1}{f}, \quad (9.15)$$

que se obtiene de (9.13) con $n_1 = -n_2$.

9.5.3. Trazado de rayos

La imagen formada por una dioptra esférica se puede localizar gráficamente de una manera sencilla utilizando las propiedades de los focos objeto e imagen y del centro de curvatura. Teniendo en cuenta que dentro de los límites de validez de la aproximación paraxial el sistema forma imagen, cada punto de la imagen se obtendrá investigando cómo se refractan los rayos emitidos por cada punto del objeto. Es conveniente tomar los rayos emitidos por la parte superior del objeto y refractados en la dioptra. Donde estos rayos (o sus prolongaciones) se corten, allí estará la imagen, real (o virtual), de la parte superior de dicho objeto. Con la base del objeto se podría hacer lo mismo, pero no es necesario porque en la aproximación paraxial las dioptras y espejos simples transforman planos en planos ¹.

De los rayos emitidos por la parte superior del objeto, un rayo fácil de seguir es el paralelo a la línea central que pasa por el centro de curvatura y la base del objeto, porque este rayo se refracta pasando por el foco imagen en dioptras convergentes ($f' > 0$) o con su prolongación pasando por el foco imagen en dioptras divergentes ($f' < 0$). Otro rayo fácil de seguir es el que pasa, él o su prolongación, por el foco objeto. Este

¹A un plano objeto perpendicular al eje óptico, le corresponde un plano imagen también perpendicular al eje (planos conjugados), de tal forma que todo punto objeto contenido en el plano objeto, tiene su correspondiente punto imagen contenido en el plano imagen.

rayo se refracta en la dioptra saliendo paralelo a la línea central que pasa por el centro de curvatura y la base del objeto. Si bien con dos rayos es suficiente para determinar la imagen, también se podría usar un tercer rayo: el que pasa, él o su prolongación, por el centro de curvatura de la dioptra, que cuando se refracta no se desvía porque incide en la dirección de la normal local.

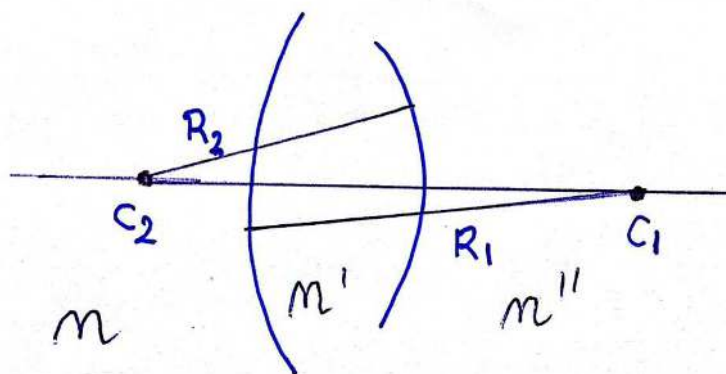


Figura 9.8: Lente gruesa: dos dioptras esféricas, radios de curvatura R_1 y R_2 , centros de curvatura C_1 y C_2 . La distancia entre dioptras, medida sobre el eje que une C_1 y C_2 es d .

9.5.4. Lentes

Una lente no es otra cosa que dos dioptras, el esquema en la Fig. 9.8, donde se muestran dos dioptras esféricas, de radios de curvatura R_1 y R_2 y con centros de curvatura C_1 y C_2 . La distancia entre dioptras, medida sobre el eje que une C_1 y C_2 es d . Si la distancia d es chica, se dice que se tiene una **lente delgada**, si no, una **lente gruesa**. El análisis para la formación de imágenes es la aplicación sucesiva del análisis hecho para una dioptra. El esquema se muestra en la Fig. 9.9: el objeto está situado en s_1 , la imagen de este objeto producida por la primera dioptra está ubicada en s_1' . Y según (9.12)

$$\frac{n'}{s_1'} - \frac{n}{s_1} = \frac{n' - n}{R_1} . \quad (9.16)$$

Análogamente, la imagen en s_1' actúa como objeto para la segunda dioptra, que ahora produce una imagen ubicada en s_2' . Según (9.12), las distancias s_2 y s_2' , medidas sobre el eje y desde la segunda dioptra, están relacionadas por

$$\frac{n''}{s_2'} - \frac{n'}{s_2} = \frac{n'' - n'}{R_2} . \quad (9.17)$$

En el esquema se ve que

$$|s_2| = d - s_1' . \quad (9.18)$$

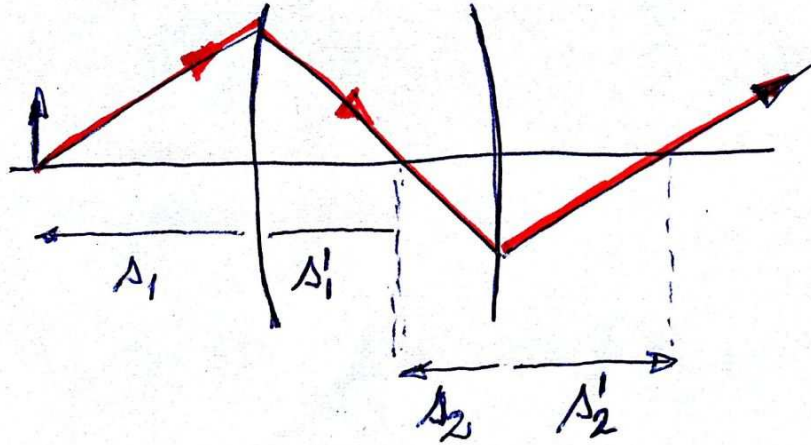


Figura 9.9: La primera dioptra produce una imagen en s_1' del objeto situado en s_1 y dicha imagen actúa como objeto para la segunda dioptra.

es decir que

$$s_2 = s_1' - d. \quad (9.19)$$

y esta relación da el esquema para aplicar (9.17) a partir del resultado de (9.16) (y para seguir iterando si hubiera más dioptras).

En el caso de lente delgadas $s_2 \approx s_1'$ y sumando miembro a miembro las ecuaciones (9.16) y (9.17) se obtiene

$$\frac{n''}{s_2'} - \frac{n}{s_1} = \frac{n' - n}{R_1} + \frac{n'' - n'}{R_2} = \Phi_1 + \Phi_2 = \frac{n''}{f'} = -\frac{n}{f}, \quad (9.20)$$

que muestra que la potencia de una lente delgada es la suma de las potencias de cada dioptra y que además define las distancias focales de la lente delgada. Si la lente es delgada no tiene sentido seguir distinguiendo los sistemas 1 y 2, todas las distancias se miden desde el mismo lugar y por este motivo la ecuación (9.20) se reescribe como

$$\frac{n''}{s'} - \frac{n}{s} = \frac{n' - n}{R_1} + \frac{n'' - n'}{R_2}. \quad (9.21)$$

El aumento lateral de una lente, el cociente entre la altura de la imagen y la altura del objeto, se obtiene siguiendo un procedimiento análogo al empleado en clase para el aumento lateral de una dioptra. El aumento lateral de una lente resulta

$$m = \frac{h'}{h} = \frac{n s'}{n'' s}. \quad (9.22)$$

Recordar que, de acuerdo con el principio de Fermat, todas las trayectorias posibles entre un punto objeto y su imagen dada por cualquier sistema óptico, tienen el mismo camino óptico.

9.6. La DG en el marco ondulatorio

A continuación reinterpretaremos las leyes de la descripción geométrica desde la visión de la teoría ondulatoria. Esto nos permitirá volver a las pobres ondas que quedaron un poco abandonadas. Y relacionarlas con las leyes fenomenológicas que fuimos discutiendo en términos de conceptos puramente geométricos como *rayos* o *caminos ópticos*. Veremos que así quedarán mas claros algunos aspectos que no se terminaban de entender, como por ejemplo *qué sucede con el rayo transmitido en reflexión total cuando el ángulo de incidencia es mayor que el ángulo crítico*

Recordemos lo discutido en la página 123, cuando pensábamos mayormente en sistemas 1D como sogas o gases en tubos. En particular, habíamos dicho que una onda progresiva en un medio continuo infinito continúa propagándose sin modificaciones, a menos que se encuentre con una discontinuidad, como un obstáculo fijo o impenetrable, un cambio en la densidad o en la tensión de la soga, o un cambio en la densidad o en la presión del gas. Hemos visto que para poder ajustar las condiciones de contorno en la discontinuidad, además de la onda transmitida que representa el avance de la perturbación en el segundo medio, también aparece una onda regresiva en el primer medio, llamada onda reflejada. Las dos ondas que aparecen en este ajuste de condiciones de contorno tienen la misma frecuencia que la onda incidente y sus periodicidades espaciales $2\pi/k$ surgen de las relaciones de dispersión $k(\omega)$ en cada medio, es decir que el k de la onda reflejada es igual al de la onda incidente, pero diferente en general del k de la onda transmitida.

En medios 3D los mecanismos físicos son completamente análogos a los que estamos acostumbrados para medios 1D, la necesidad de ajustar condiciones de contorno en una discontinuidad produce la aparición de ondas reflejadas en el medio de incidencia y de ondas transmitidas en el segundo medio. Estas ondas tienen la misma frecuencia que la onda incidente y sus periodicidades espaciales en la dirección de propagación $2\pi/k$ surgen de las relaciones de dispersión $k(\omega)$ en cada medio, el k de la onda reflejada es igual que el de la onda incidente y en general diferente al k de la onda transmitida. Pero mientras que antes la discontinuidad se podía describir sólo con un valor particular de la única coordenada involucrada (por ejemplo, $x = L$), en un medio 3D la discontinuidad entre dos medios homogéneos es una superficie y aún en el caso en que esta superficie fuera plana, la dirección de propagación de la onda incidente no es necesariamente perpendicular a dicha superficie. Justamente el formalismo de la descripción geométrica da la respuesta al problema cinemático de cómo orientar los vectores de onda de las ondas reflejadas y transmitidas, dada la dirección de la onda incidente y la forma de la discontinuidad.

Para encarar la interpretación de las leyes de la óptica geométrica en el marco de una descripción ondulatoria comenzaremos considerando el caso más sencillo posible de discontinuidad en un medio 3D: un plano infinito. Que tiene la ventaja de ser un obstáculo que no tiene “tamaños característicos” (esas cantidades que se comparaban con la longitud de onda para ver la validez de la descripción geométrica). En realidad un plano infinito **sí** tiene tamaños característicos: valen infinito, tanto su extensión como su curvatura. Y teniendo en cuenta que las ondas planas armónicas forman base

para cualquier solución de la ecuación de ondas clásica, (sección 10.3), a continuación analizaremos como satisfacer las condiciones de contorno cuando la discontinuidad está “iluminada” por una onda plana. Veremos que todos los aspectos cinemáticos de este problema surgen solamente de admitir la existencia de condiciones de contorno, *aún sin especificar de qué condiciones de contorno estamos hablando*. Justamente por este motivo, estos comportamientos cinemáticos (las leyes de la óptica geométrica) son comunes a cualquier fenomenología ondulatoria (mecánica, acústica, electromagnética, etc.).

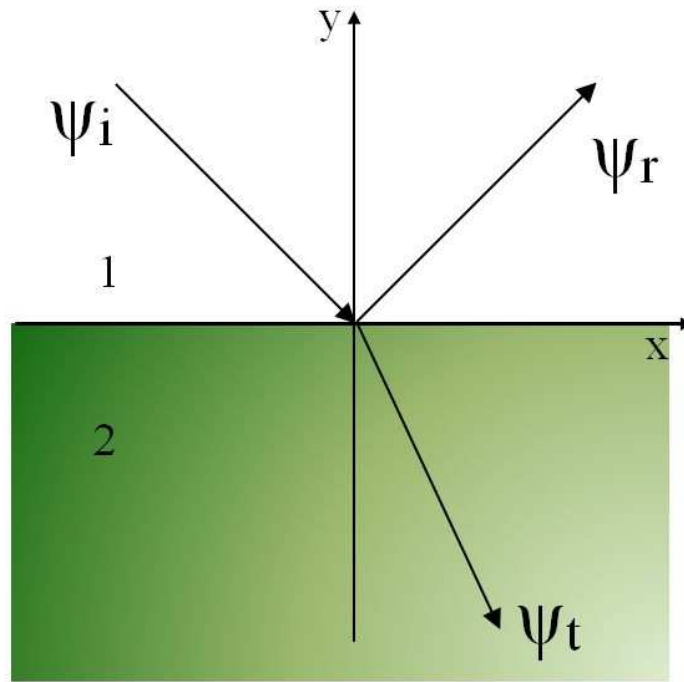


Figura 9.10: Onda plana incidente, reflejada y transmitida.

9.6.1. Reflexión y transmisión de una onda plana en una superficie plana

El plano $y = 0$ separa dos medios materiales isótropos, uno en la región 1 ($y > 0$) y otro en la región 2 ($y < 0$). La función de onda en cada región satisface la ecuación de ondas clásica con velocidades de fase $v_1(\omega) = c/n_1(\omega)$ y $v_2(\omega) = c/n_2(\omega)$, $n_1(\omega)$ y $n_2(\omega)$ los índices de refracción de los medios 1 y 2. Una onda incidente monocromática plana de frecuencia ω se propaga en la región $y > 0$ y se encuentra con la discontinuidad en $y = 0$. Podríamos pensar que esta onda incidente fue originada por una fuente situada en $y \rightarrow \infty$, pero hay otras posibilidades. Sin pérdida de generalidad, supongamos que

el plano $x - y$ es el plano de incidencia. Según (7.21)

$$\psi_i(\vec{r}, t) = A_{0i} e^{i(\vec{k}_i \cdot \vec{r} - \omega t)} , \quad (9.23)$$

con

$$\vec{k}_i = \underbrace{\frac{\omega}{c}}_{k_1} n_1 (\sin i \hat{x} - \cos i \hat{y}) = \alpha_i \hat{x} - \beta_i \hat{y} . \quad (9.24)$$

En analogía con lo discutido para el caso 1D, intentamos ajustar las condiciones de contorno en la discontinuidad $y = 0$ mediante una onda reflejada (en el medio 1) y una onda transmitida (en el medio 2)

$$\psi_r(\vec{r}, t) = A_{0r} e^{i(\vec{k}_r \cdot \vec{r} - \omega t)} , \quad (9.25)$$

$$\psi_t(\vec{r}, t) = A_{0t} e^{i(\vec{k}_t \cdot \vec{r} - \omega t)} . \quad (9.26)$$

Como en todo forzado estacionario de un sistema físico lineal, estas ondas tienen la misma frecuencia que la onda incidente. Y por ser soluciones de ecuaciones de ondas clásicas, las periodicidades espaciales en la dirección de propagación surgen de las relaciones de dispersión para cada medio. Como ya hemos discutido, estas relaciones fijan el módulo del vector de onda, pero no su dirección. Por eso escribimos los vectores de onda de las ondas reflejada y transmitida como

$$\vec{k}_r = \alpha_r \hat{x} + \beta_r \hat{y} + \gamma_r \hat{z} , \quad (9.27)$$

y

$$\vec{k}_t = \alpha_t \hat{x} - \beta_t \hat{y} + \gamma_t \hat{z} , \quad (9.28)$$

con

$$\alpha_r^2 + \beta_r^2 + \gamma_r^2 = \frac{\omega^2}{c^2} n_1^2 = k_1^2 , \quad (9.29)$$

$$\alpha_t^2 + \beta_t^2 + \gamma_t^2 = \frac{\omega^2}{c^2} n_2^2 = k_2^2 . \quad (9.30)$$

9.6.2. Las condiciones de contorno imponen la cinemática de los rayos

Aunque no sepamos en qué dirección apuntan $\vec{k}_r$ y $\vec{k}_t$, las ecuaciones (9.29) y (9.30) dicen que $\vec{k}_r$ y $\vec{k}_t$ están en esferas de radio k_1 y k_2 respectivamente. La dirección surge de aceptar que en el plano $y = 0$ hay que satisfacer condiciones de contorno lineales. Por condiciones de contorno lineales entendemos condiciones que se puedan escribir como combinación lineal de las funciones de onda y sus derivadas a un lado y otro de la discontinuidad. Es instructivo revisar los problemas que se hicieron en los casos 1D, porque justamente éste es el tipo de condiciones de contorno que han aparecido en estos problemas, donde la física de cada situación particular condujo siempre a condiciones del tipo “continuidad de la función”, o “continuidad de la derivada de la función”, o

“continuidad de la derivada de la función multiplicada por un coeficiente”. Supongamos que estamos en el caso “continuidad de la función”. Esto quiere decir que en todo instante de tiempo y para todo punto $\vec{r}_P$ en el plano $x-z$ ($y=0$), debe cumplirse que $\psi_i(\vec{r}_P, t) + \psi_r(\vec{r}_P, t) = \psi_t(\vec{r}_P, t)$. Esta condición nos lleva a pedir que

$$A_{0i} e^{i(\alpha_i x)} = -A_{0r} e^{i(\alpha_r x + \gamma_r z)} + A_{0t} e^{i(\alpha_t x + \gamma_t z)} . \quad (9.31)$$

Esto es raro ... ¿Una función de x (el lado izquierdo) igual a una función de x y de z (el lado derecho)? Sería posible (pero no demasiado interesante) si todas las amplitudes fueran cero. Si no, para que se cumpla la ecuación (9.31) debe ser

$$\gamma_t = \gamma_r = 0 . \quad (9.32)$$

Y esto equivale a decir que el vector de onda reflejado $\vec{k}_r$ y el vector transmitido $\vec{k}_t$ están en el plano que contiene al vector de onda incidente $\vec{k}_i$ y a la normal $\hat{y}$. [La ley del plano de incidencia](#). Es fácil darse cuenta que habríamos llegado a la misma conclusión si en vez de considerar condiciones de contorno del tipo “continuidad de la función” hubiéramos considerado “continuidad de la función multiplicada por un coeficiente” u otro tipo de condición de contorno lineal. Vemos entonces que la ley sobre el plano de incidencia es una consecuencia de la existencia de condiciones de contorno, toda onda plana reflejada o transmitida en una superficie plana va a estar contenida en el plano de incidencia, independientemente de la naturaleza de la onda.

Admitida la igualdad (9.32), como todos los vectores están en el plano $x-y$, en este plano se pueden definir ángulos de reflexión y transmisión entre los vectores de onda y la normal

$$\vec{k}_r = \underbrace{\frac{\omega}{c}}_{k_1} n_1 (\sin r \hat{x} + \cos r \hat{y}) = \alpha_r \hat{x} - \beta_r \hat{y} , \quad (9.33)$$

$$\vec{k}_t = \underbrace{\frac{\omega}{c}}_{k_2} n_2 (\sin t \hat{x} - \cos t \hat{y}) = \alpha_t \hat{x} - \beta_t \hat{y} . \quad (9.34)$$

Y para resolver el problema dinámico de encontrar las amplitudes de las ondas reflejada y transmitida, habría que encontrar antes las componentes α_r y α_t de los vectores de onda $\vec{k}_r$ y $\vec{k}_t$. Para eso, reescribimos (9.31) como

$$A_{0i} e^{i \alpha_i x} = -A_{0r} e^{i \alpha_r x} + A_{0t} e^{i \alpha_t x} . \quad (9.35)$$

El lado izquierdo de esta igualdad es una función periódica de la variable x , con período $2\pi/\alpha_i$. ¿Una función perfectamente periódica en todo el intervalo $-\infty < x < \infty$ igual a la suma de dos funciones con las mismas características pero con períodos distintos? Si las amplitudes no son cero, la única manera de que esto suceda es que los períodos sean todos iguales y para eso todos los α deben ser iguales ²

$$\alpha_i = \alpha_r = \alpha_t , \quad (9.36)$$

²Otra manera de decir lo mismo es notar que el conjunto de funciones $e^{i \alpha x}$ etiquetadas por α son funciones de x linealmente independientes. Y entonces, para que se cumpla (9.35) en todo punto de la superficie de separación, solamente existen dos posibilidades: i) todos los coeficientes de la combinación lineal son cero, o ii) todos los exponentes (todos los α) son iguales.

igualdad que equivale a la leyes de la reflexión, $i = r$, y de Ibn-Sahl, $n_1 \sin i = n_2 \sin t$. Como antes, la misma conclusión se obtendría considerando condiciones de contorno del tipo “continuidad de la función multiplicada por un coeficiente” u otro tipo de condición de contorno lineal, en vez de la “continuidad de la función” considerada en (9.35) y por este motivo concluimos que para el caso de una onda plana incidente, las leyes de la DG son todas consecuencias necesarias de la existencia de condiciones de contorno lineales en un plano infinito, independientemente de la naturaleza (acústica, electromagnética, mecánica, etc.) de la onda.

9.6.3. Conservación de componentes tangenciales de $\vec{k}$

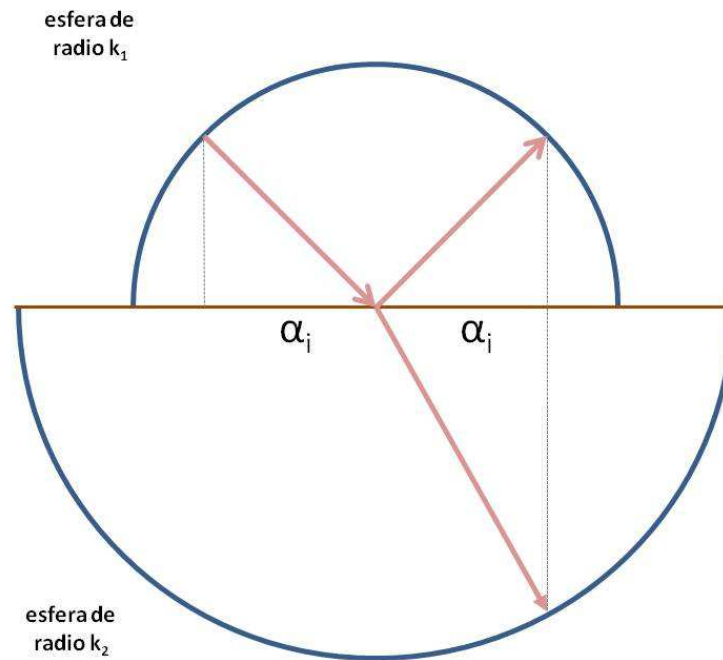


Figura 9.11: Construcción en el espacio de los vectores de onda, $n_1 < n_2$.

Vemos que todas las leyes de la descripción geométrica se pueden obtener como una consecuencia natural de la existencia de condiciones de contorno lineales en un plano infinito. Las ecuaciones (9.32) y (9.36) se pueden reformular como una ley de conservación de las componentes tangenciales de los vectores de onda. En cursos superiores se verá que esta ley de conservación es una consecuencia de la simetría que tiene un plano infinito, que resulta invariante frente a traslaciones infinitesimales. No como los techos de chapa acanalada, que tienen ondulaciones periódicas y que entonces son invariantes frente a traslaciones en la dirección perpendicular a las canaletas sólo si son traslaciones finitas e iguales a la distancia entre dos canaletas, mientras que en la dirección de las canaletas si son invariantes frente a traslaciones infinitesimales. Si una de estas

chapas actuara como espejo, se verían cosas raras, porque no se conservan todas las componentes tangenciales de los vectores de onda.

Trabajando en el espacio de los vectores de onda, la ley de conservación de las componentes tangenciales de los vectores de onda facilita la visualización del movimiento de los rayos reflejado y transmitido cuando cambia la inclinación del rayo incidente. Todos los vectores de onda involucrados en el ajuste de condiciones de contorno, tienen módulo impuesto por la ecuación de dispersión en cada medio. Los vectores de la onda incidente y la reflejada tienen que estar en la esfera superior (ver Fig. 9.11) y de la onda transmitida en la esfera inferior. Y como todos tienen que tener componente tangencial α_i , quedan fijas las direcciones de propagación. La Fig. 9.11 muestra el caso $n_1 < n_2$, mientras que la Fig. 9.12 el caso $n_1 > n_2$ justo cuando el ángulo de incidencia es igual al crítico de reflexión total, cuando la componente tangencial del vector de onda incidente es igual al radio de la esfera inferior. ¿Qué pasa cuando el ángulo de incidencia es mayor al ángulo crítico de reflexión total? ¿Cómo se dibuja en la esfera inferior un vector de onda con una componente horizontal mayor que el radio de la esfera?

9.6.4. La reflexión total en el marco ondulatorio

Si $n_1 > n_2$, $\sin \theta_t > \sin \theta_i$ y entonces existe un ángulo de incidencia crítico θ_{ic} para el cual la onda refractada se propaga en dirección paralela a la superficie

$$\sin \theta_{ic} = \frac{n_2}{n_1}. \quad (9.37)$$

En función de θ_{ic} , la ley de Snell se escribe

$$\sin \theta_t = \frac{\sin \theta_i}{\sin \theta_{ic}}, \quad (9.38)$$

que muestra que

$$\begin{aligned} \theta_i \leq \theta_{ic} &\Rightarrow \sin \theta_t \leq 1 \Rightarrow \cos \theta_t = \sqrt{1 - \left(\frac{\sin \theta_i}{\sin \theta_{ic}}\right)^2} \quad (\theta_t \in \mathbf{R}), \\ \theta_i > \theta_{ic} &\Rightarrow \sin \theta_t > 1 \Rightarrow \cos \theta_t = i\sqrt{\left(\frac{\sin \theta_i}{\sin \theta_{ic}}\right)^2 - 1} \quad (\theta_t \notin \mathbf{R}). \end{aligned} \quad (9.39)$$

¿Un ángulo de transmisión complejo? Seguro que las descripciones anteriores no contemplaban esta novedad. Sólo el tratamiento ondulatorio muestra que para ángulos de incidencia mayores que el crítico tenemos que revisar la interpretación de θ_t como ángulo de refracción. El significado queda en evidencia cuando investiguemos la estructura de la onda refractada en el medio 2 ($y < 0$)

$$\begin{aligned} \psi_t(\vec{r}, t) &= A_{0t} e^{i\left[\frac{\omega}{c}n_2(x \sin \theta_t - y \cos \theta_t) - \omega t\right]} \\ &= A_{0t} e^{i\left[\frac{\omega}{c}n_2 \sqrt{\left(\frac{\sin \theta_i}{\sin \theta_{ic}}\right)^2 - 1} y + \left(\frac{\omega}{c}n_2 \sin \theta_t\right)x - \omega t\right]}. \end{aligned} \quad (9.40)$$

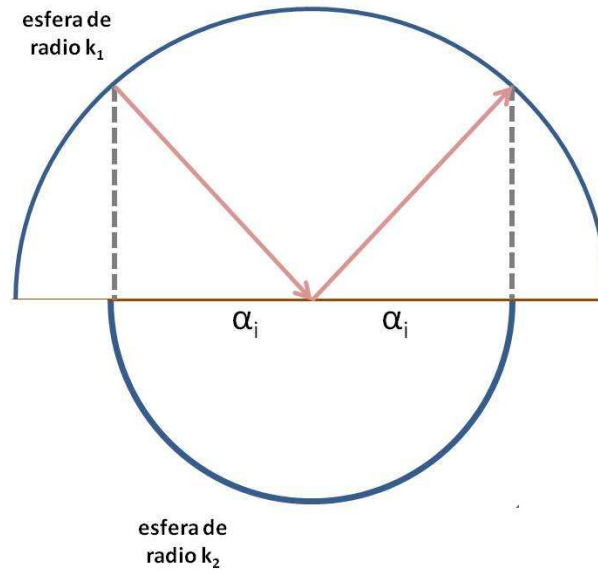


Figura 9.12: Construcción en el espacio de los vectores de onda, $n_1 > n_2$, ángulo crítico.

Esta expresión, válida para $y \leq 0$, representa una onda que se propaga a lo largo de la superficie sin atenuarse y cuya amplitud decae exponencialmente hacia el interior del medio 2 (notar que esto está relacionado con la elección de la raíz cuadrada en (9.39)). Esta onda se llama **onda evanescente** y es como las que aparecieron en el régimen reactivo de la ecuación de Klein-Gordon, página 145. La distancia de decaimiento δ de la onda transmitida resulta

$$\frac{1}{\delta} = \frac{\omega}{c} n_2 \sqrt{\left(\frac{\sin \theta_i}{\sin \theta_{ic}}\right)^2 - 1},$$

es decir

$$\delta = \frac{\lambda_1}{2\pi} \frac{1}{\sqrt{\sin^2 \theta_i - \sin^2 \theta_{ic}}}, \quad (9.41)$$

donde λ_1 es la longitud de onda en el medio 1. Como en el ejemplo de péndulos acoplados, este decaimiento no está asociado a ninguna pérdida, sino a un comportamiento reactivo del medio 2 para $\theta_i > \theta_{ic}$.

Capítulo 10

Ondas transversales

El objetivo de esta parte del curso es poner en evidencia fenómenos comunes a toda **onda transversal**. Fenómenos que son consecuencia del carácter **vectorial** de la perturbación. Y que hasta ahora hemos ignorado en aras de no complicar demasiado las cosas. Por este motivo, en los ejemplos previos hemos prestado atención a solamente a una de las dos componentes posibles de una onda transversal. Un ejemplo de esta afirmación se ve en la figura 6.2, allí se estudiaron las ecuaciones de movimiento de una cadena lineal para apartamientos del equilibrio solamente en el plano de la figura, pero no en el plano perpendicular. Veremos a continuación qué novedades trae considerar la presencia simultánea de las **dos** componentes posibles de una onda transversal. Ya sabemos que para ecuaciones de onda lineales basta con estudiar ondas con dependencia temporal armónica. Y dijimos, sin demostrar, que una consecuencia del análisis de Fourier espacial es que casi cualquier perturbación se pueda escribir como combinación lineal de ondas planas armónicas. Así que ...ondas planas armónicas **transversales** pasamos a estudiar.

10.1. Ondas planas armónicas transversales

Si el medio que soporta la onda es isótropo y no hay direcciones “especiales”, entonces ambas componentes de la onda transversal satisfacen la misma ecuación (7.18), con el mismo valor de v para las dos componentes. Esto equivale a decir que la perturbación satisface la ecuación (7.18), pero reemplazando ψ por $\vec{\psi}$. En un medio anisótropo, en cambio, hay direcciones “especiales” y a la componente de $\vec{\psi}$ en una dirección “especial” le corresponde en general una velocidad de fase distinta que a la componente de $\vec{\psi}$ en otra dirección “especial”.

Si elegimos que la dirección de propagación de la onda coincida con el eje z , entonces la perturbación $\vec{\psi}$ debe estar en el plano perpendicular a $\hat{z}$. Luego, la expresión más general para una onda plana transversal con dependencia temporal armónica se tiene que formar a partir de soluciones como la obtenida en la ecuación (7.21):

$$\vec{\psi}(z, t) = \psi_x(z, t) \hat{x} + \psi_y(z, t) \hat{y}, \quad (10.1)$$

una solución como en (7.21) para la componente $\hat{x}$ con amplitud A_x y otra solución de la misma forma para la componente $\hat{y}$ con amplitud A_y

$$\psi_x(z, t) = A_x \cos(kz - \omega t), \quad (10.2)$$

$$\psi_y(z, t) = A_y \cos(kz - \omega t + \phi), \quad (10.3)$$

donde ϕ es el desfase entre ambas componentes. Esta forma siempre es válida con una elección adecuada del origen de t o de z , porque si en vez de la expresión anterior hubiéramos tenido

$$\psi_x(z, t) = A_x \cos(kz - \omega t + \phi_x),$$

$$\psi_y(z, t) = A_y \cos(kz - \omega t + \phi_y),$$

entonces siempre se podría definir t' de manera tal que $-\omega t + \phi_x = -\omega t'$. Y con el nuevo origen de tiempo el argumento del primer coseno sería igual a $kz - \omega t'$ mientras que el argumento del segundo coseno sería igual a $kz - \omega t' + \phi$, con $\phi = \phi_y - \phi_x$. Una discusión similar vale para la elección del origen de coordenadas z .

Las eqs. (10.2) y (10.3) muestran que para definir una onda plana transversal con dependencia temporal armónica **no** son necesarios los cuatro parámetros A_x , A_y , ϕ_x y ϕ_y , si no sólo tres parámetros: A_x , A_y y la diferencia de fase ϕ . Notemos que estas ecuaciones definen dónde está la punta del vector $\vec{\psi}(z, t)$. Y la forma trazada por la punta del vector $\vec{\psi}(z, t)$ es lo que se conoce como **estado de polarización** de la onda transversal. O directamente, la polarización de la onda. Como las eqs. (10.2) y (10.3) no dependen de z y de t por separado, sino de la combinación $kz - \omega t$, es casi equivalente (a menos de un signo) estudiar la forma trazada por la punta del vector $\vec{\psi}$ en un lugar fijo a medida que pasa el tiempo, o la forma trazada por la punta del vector $\vec{\psi}$ en un tiempo fijo como función de z . Veamos primero los estados de polarización que son fácilmente visualizables (y más tarde veremos que todos los casos restantes se reducen a éstos cuando se rota convenientemente el sistema de ejes $x - y$ alrededor del eje z).

10.2. Polarización lineal

Cuando ϕ es un múltiplo par de π ($\phi = 2m\pi$, con m entero), las perturbaciones componentes ψ_x y ψ_y pasan por cero juntas, crecen juntas, disminuyen juntas y se hacen negativas juntas. Pensando en composición de vectores, esto quiere decir que la forma trazada por la punta del vector $\vec{\psi}(z, t)$ en un z fijo a medida que pasa el tiempo es una recta. Y que en un dado instante de tiempo, todos los vectores $\vec{\psi}(z, t)$ en distintos lugares están contenidos en un plano que contiene al eje z y que pasa por el primer y tercer cuadrante del plano $x - y$. Si necesitamos convencernos con algo más matemático, entonces dividimos miembro a miembro las ecuaciones (10.2) y (10.3) y queda que

$$\psi_y(z, t) = \frac{A_y}{A_x} \psi_x(z, t), \quad (10.4)$$

que dice que el ángulo que el vector $\vec{\psi}(z, t)$ forma con el eje x no depende ni de z ni de t . En la visualización a z fijo diríamos que la punta del vector $\vec{\psi}$ recorre una recta, por eso

esta polarización se llama **lineal** y decimos que la onda está **linealmente polarizada**. En la visualización a t fijo diríamos que si sacamos una foto, los vectores $\vec{\psi}$ están en un plano, por eso se dice que una onda con esta polarización está **plano polarizada**. Notar que según las (10.2) y (10.3), la onda plana transversal con dependencia temporal armónica más general es una combinación lineal de dos ondas planas linealmente polarizadas en direcciones ortogonales.

Cuando ϕ es un múltiplo impar de π ($\phi = (2m+1)\pi$ con m entero) las conclusiones son muy similares a las del párrafo anterior. Las dos perturbaciones, ψ_x y ψ_y , pasan por cero juntas, crecen en valor absoluto juntas, y en valor absoluto disminuyen juntas. Pero cuando una es positiva la otra es negativa. Pensando en composición de vectores, esto quiere decir que la forma trazada por la punta del vector $\vec{\psi}(z, t)$ en un z fijo a medida que pasa el tiempo también es una recta. Y que en un dado instante de tiempo, todos los vectores $\vec{\psi}(z, t)$ en distintos lugares también están contenidos en un plano que contiene al eje z . Pero ahora el plano no pasa por el primer y tercer cuadrante del plano $x - y$, sino por el segundo y el cuarto. Esto se ve al dividir miembro a miembro las dos ecuaciones (10.2) y (10.3), que ahora dan

$$\psi_y(z, t) = -\frac{A_y}{A_x} \psi_x(z, t) , \quad (10.5)$$

que de nuevo dice que el ángulo que el vector $\vec{\psi}(z, t)$ forma con el eje x no depende ni de z ni de t y por eso este caso también corresponde a una **polarización lineal**.

Otra manera de analizar el estado de polarización es a través del concepto de **vector amplitud**. En los dos casos considerados, el vector $\vec{\psi}(z, t)$ se puede escribir de la siguiente manera

$$\vec{\psi}(z, t) = (A_x \hat{x} \pm A_y \hat{y}) \cos(kz - \omega t) = \vec{A}_{\pm} \cos(kz - \omega t) , \quad (10.6)$$

que no es otra cosa que un vector cuyo módulo oscila armónicamente, tanto en el tiempo como en la dirección de propagación, pero que siempre apunta en la dirección del vector fijo $\vec{A}_{\pm}$. Como los vectores $\vec{A}_+$ y $\vec{A}_-$ son linealmente independientes, no es difícil justificar que con estos vectores se puede formar una base de perturbaciones para escribir cualquier perturbación transversal como combinación lineal de esta base.

10.3. Polarización elíptica

Si ϕ es múltiplo impar de $\pi/2$ y $A_y \neq A_x$, el coseno que aparece en $\psi_y(z, t)$ en (10.3) se transforma, a menos de un signo, en un seno. Entonces las dos perturbaciones ψ_x y ψ_y no pasan por cero juntas: cuando una es cero la otra es máxima en valor absoluto. Es decir que hay instantes donde la punta del vector $\vec{\psi}(z, t)$ está dirigida totalmente según un eje, mientras que en otros instantes la punta del vector $\vec{\psi}(z, t)$ está dirigida totalmente según el eje perpendicular al primero. Estos casos corresponden a **polarización elíptica**, llamada así porque la forma trazada por la punta del vector $\vec{\psi}(z, t)$ responde a la ecuación

$$1 = \frac{\psi_x^2(z, t)}{A_x^2} + \frac{\psi_y^2(z, t)}{A_y^2} , \quad (10.7)$$

que es una elipse con semiejes A_x y A_y . El caso particular $A_x = A_y$ corresponde a una circunferencia.

En la visualización a z fijo la punta del vector $\vec{\psi}$ describe una elipse y está rotando, en sentido horario o en sentido antihorario. El sentido de la rotación depende de que los desfases sean congruentes con $+\pi/2$ o con $-\pi/2$. En la visualización a t fijo diríamos que si sacamos una foto, los vectores $\vec{\psi}$ estarán en una especie de sacacorchos elíptico, algunos como los sacacorchos de la vida real, y otros como su imagen especular, justamente dependiendo del sentido de giro.

Para definir el **vector amplitud** en el caso de elipses centradas en los ejes, conviene pasar los cosenos a la notación exponencial, es decir,

$$\begin{aligned}\psi_x(z, t) &= \Re A_x \exp(kz - \omega t), \\ \psi_y(z, t) &= \Re A_y \exp(kz - \omega t \pm \pi/2),\end{aligned}$$

y factorizando las exponenciales, se puede expresar el vector $\vec{\psi}(z, t)$ como

$$\vec{\psi}(z, t) = \Re (A_x \hat{x} \pm i A_y \hat{y}) = \Re \vec{A}_{\pm} \exp(kz - \omega t). \quad (10.8)$$

Con la notación exponencial, donde es costumbre no escribir $\Re$, tenemos un **vector amplitud compleja** $\vec{A}_{\pm}$ para la onda elípticamente polarizada que cuando se observa a z fijo corresponde a una onda que gira en sentido antihorario, es decir para ϕ congruente a $-\pi/2$. Análogamente, el vector amplitud compleja $\vec{A}_{+}$ representa una onda elípticamente polarizada que en cada z gira en sentido horario, es decir para ϕ congruente a $\pi/2$. Como los vectores $\vec{A}_{+}$ y $\vec{A}_{-}$ en este caso también son linealmente independientes, no es difícil justificar que con estos vectores se puede formar una base de perturbaciones para escribir cualquier perturbación transversal como combinación lineal de ondas elípticamente polarizadas en sentidos opuestos.

10.4. Polarización circular

Este es un caso particular del anterior, ahora la elipse (10.7) degenera en una circunferencia y se dice que la onda transversal está **circularmente polarizada**. Notar que el valor absoluto de la perturbación $\vec{\psi}(z, t)$ ahora permanece constante en todo lugar y tiempo, aunque su dirección rota, en un sentido o en otro dependiendo de que los desfases sean congruentes con $+\pi/2$ o con $-\pi/2$. En este caso los vectores de amplitud compleja son $\vec{A}_{\pm} = A_{\pm}(\hat{x} \pm i \hat{y})$ que también se pueden usar como elementos de una base que permite expresar cualquier perturbación transversal como combinación lineal de dos ondas circularmente polarizadas en sentidos opuestos.

Los medios formados por estructuras desordenadas son isótropos y no distinguen entre estados de polarización ortogonales. Excepto los que tienen asociado un sentido de giro, que son sensibles al sentido de giro de la polarización de las ondas electromagnéticas. En estos medios la velocidad de fase para las ondas circularmente polarizadas en un sentido es distinta a la de las que están polarizadas en sentido contrario. Si una onda linealmente polarizada atraviesa este tipo de medio, al abandonarlo habrá rotado

su dirección de polarización. Para ver que esto es así, sólo hace falta escribir primero la onda linealmente polarizada en la base circularmente polarizada. Y notar que si ambas ondas circularmente polarizadas viajaran con la misma velocidad, la polarización a la salida seguiría siendo la misma que había a la entrada. Pero como cada elemento de la base circular viaja con distinta velocidad, las componentes en esta base se desfasan y esto se traduce en la rotación de la dirección de incidencia. Es el principio de funcionamiento de los polarímetros (ver video en <http://bit.ly/1PV5vvu>).

En la naturaleza abundan medios que tienen un sentido de giro asociado, son sustancias compuestas por grandes moléculas orgánicas, que tienen orientaciones espaciales en forma de hélice. El caso más conocido es el de la molécula de ácido desoxirribonucleico (ADN), ver Figura 10.1.

La mayoría de proyecciones 3D actuales usan luz circularmente polarizada en ambos sentidos para poder brindar imágenes levemente diferentes a cada ojo y así producir en el espectador la sensación estereoscópica. Si van a ver una película 3D, prueben el efecto de superponer el cristal de sus gafas correspondiente al ojo izquierdo con el correspondiente al ojo derecho de las gafas de su vecino.

10.5. Caso ϕ arbitrario

Si el valor de ϕ en la página 176 no es ninguno de los considerados hasta ahora, lo que hay que hacer es eliminar el parámetro $kz - \omega t$ de las ecuaciones paramétricas (10.2) y (10.3), para obtener una relación, ya sea explícita (como en (10.4) y en (10.5)) o implícita (como en (10.7)), entre las componentes ψ_x y ψ_y . De la primera ecuación despejamos el coseno del parámetro

$$\cos(kz - \omega t) = \psi_x(z, t) / A_x$$

Para sacar el parámetro de la segunda ecuación, desarrollamos el coseno de la segunda ecuación como

$$\cos(kz - \omega t + \phi) = \cos(kz - \omega t) \cos \phi - \sin(kz - \omega t) \sin \phi$$

Ahora reemplazamos $\sin(kz - \omega t)$ por

$$\sqrt{1 - [\psi_x(z, t) / A_x]^2}$$

y obtenemos una ecuación implícita que parece fea o desprolija en la primera impresión. Pero eso es porque recién se despierta y necesita un poco de ayuda. Para ayudarla a mejorar su aspecto, tratamos de sacarle esa raíz cuadrada que le queda tan mal, para eso pasamos la raíz cuadrada de una lado de la igualdad y dejamos todo el resto del otro. Y (todo sea por ayudar), elevamos al cuadrado miembro a miembro. Acá hay que aplicar eso del cuadrado del primero, mas el cuadrado del segundo, blablabla . . . y si nos tenemos fé, luego de emprolijar un poco las cosas, obtendremos la siguiente ecuación

$$\left(\frac{\psi_x}{A_x}\right)^2 + \left(\frac{\psi_y}{A_y}\right)^2 - 2 \frac{\psi_x \psi_y}{A_x A_y} \cos \phi = \sin^2 \phi \quad (10.9)$$

que hay que reconocer que mejoró muchísimo. No es otra cosa que la ecuación de una sección cónica. Y comparando el producto de los coeficientes de los términos cuadráticos (no cruzados) con el coeficiente del término cruzado, es claro que se trata de una elipse (ver <http://bit.ly/2epLUaC>). Es fácil demostrar que el eje mayor de esta elipse forma con el eje x un ángulo β dado por

$$\tan 2\beta = \frac{2 A_x A_y}{A_x^2 - A_y^2} \cos \phi \quad (10.10)$$

Corolario 1: en el sistema de ejes donde los ejes de la elipse están centrados, el desfase entre componentes ortogonales es $\pm \pi/2$ y sus congruentes módulo 2π .

Corolario 2: para describir una polarización elíptica, conviene (si se puede) girar los ejes para que coincidan con los ejes propios de la elipse, porque así los desfases serán siempre $+\pi/2$ o $-\pi/2$, dependiendo del sentido de giro.

10.5.1. Luz natural

La radiación electromagnética (luz incluida) proveniente de fuentes térmicas se puede describir con las eqs. (10.2) y (10.3), con la diferencia de que los cosenos, en vez de no empezar ni terminar nunca, están modulados por una función que empieza y termina, con una duración $\tau \approx 10^{-8}$ s. A estos cosenos modulados se los conoce como “trenes de onda”. Cada tren de onda difiere del anterior en el valor de la diferencia de fase ϕ y el valor de ϕ cambia aleatoriamente en cada intervalo τ . Todos los valores de ϕ tienen la misma probabilidad, lo que equivale a decir que todos los estados de polarización son equiprobables en cada intervalo τ y que cambian $\tau^{-1} \approx 10^8$ veces en un segundo. Como los detectores comunes son incapaces de resolver cien millones de cambios en un segundo, se suele decir que una onda con polarización natural está **no polarizada**. Aunque realmente lo esté, durante tiempos muy cortos. Lo más adecuado sería llamarla polarizada **al azar** (o también polarizada a la *a mí qué me importa*). Observar que una onda plana perfectamente monocromática, es decir con una frecuencia bien definida, siempre mantiene un estado de polarización bien definido. Porque si es tren, no es monocromática, no?

Si a una onda con polarización natural se le agrega o se le resta una onda polarizada, se dice que la onda tiene un estado **parcialmente polarizado** o que la onda está parcialmente polarizada.

10.6. Polarización por absorción selectiva

Una manera de polarizar un haz de luz natural es mediante el uso de materiales o dispositivos que absorben selectivamente una de las componentes transversales del campo eléctrico de una onda. La absorción puede estar ligada a la estructura molecular del material o a asimetrías espaciales. Un ejemplo son las rejillas de alambres conductores (Fig. 10.2) empleadas para microondas (las microondas son una forma de radiación

electromagnética, como la luz visible, pero con λ del orden de los centímetros). La energía de la componente del campo eléctrico paralela a los alambres se “gasta” en mover los electrones libres de los alambres, mientras que la energía de la componente del campo eléctrico perpendicular a los alambres permanece casi inalterada. De esta manera, a la salida de la rejilla se tiene una onda linealmente polarizada en dirección perpendicular a los alambres.

Otro ejemplo de polarización por absorción selectiva ocurre en ciertos cristales como la turmalina, o en las láminas Polaroid, donde las estructuras moleculares juegan un papel similar al que juegan los alambres conductores en las rejillas de microondas. En los polaroids reales hay una componente que se atenúa, aunque no completamente, la lámina es más cara cuanto mejor polariza (es decir, cuanto más atenúa). Salvo que se diga lo contrario, se supone que las láminas polarizadoras son ideales, es decir que a la salida se obtiene una onda perfectamente polarizada de manera lineal.

Si los ejes de transmisión de dos polarizadores ideales son perpendiculares, no se transmite luz. En cambio, si los ejes de transmisión de los polarizadores forman un ángulo θ , la intensidad de la luz transmitida sigue la **ley de Malus**

$$I(\theta) = I_{max} \cos^2 \theta , \quad (10.11)$$

que surge de proyectar la amplitud de la perturbación linealmente polarizada que sale de la primera lámina en la dirección del eje de transmisión de la segunda lámina y recordar que la intensidad es proporcional al cuadrado de la amplitud.

10.7. Polarización por reflexión

Cuando una onda plana incide oblicuamente sobre una superficie plana (ver esquema de la Fig. 10.3), existe un plano especial, el plano de incidencia, donde están contenidos los rayos reflejado y transmitido. Si bien el espacio es isótropo y un plano infinito es indistinguible frente a rotaciones alrededor de un eje perpendicular, este plano de incidencia introduce una asimetría para las ondas transversales, porque la componente de la polarización incidente que está contenida en el plano de incidencia no tiene por lo general el mismo coeficiente de reflexión que la componente de la polarización incidente que es perpendicular al plano de incidencia. Una manera fácil de ver que esto es así es pensando que si la superficie reflectante fuera metálica, la componente del campo eléctrico incidente perpendicular al plano de incidencia, paralela a la superficie reflectante, es muy eficiente en mover las cargas sueltas que existen en el metal. En cambio, a la otra componente le resulta más difícil, porque no es paralela a la superficie reflectante. En fin, de todo esto mejor no decir demasiado porque a nivel F2 no se sabe nada de electromagnetismo, que se empieza a ver en F3. Y el problema completo se resolverá recién en Física Teórica 1. Pero bueno, qué le vamos a hacer, la vida del científico es así, uno se pasa la vida completando huecos.

Lo importante es que la fracción de la potencia incidente que se lleva la onda reflejada es distinta en cada caso. Y entonces la polarización de la onda reflejada va a ser

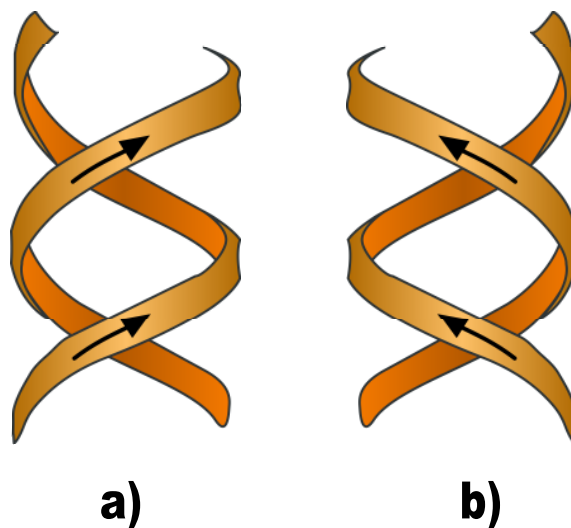


Figura 10.1: Estructuras helicoidales con sentidos de giro opuestos. Notar que la estructura de un tipo se obtiene como reflexión especular de la estructura del otro tipo (de Wikimedia Commons https://commons.wikimedia.org/wiki/File:Levo_dextro.svg)

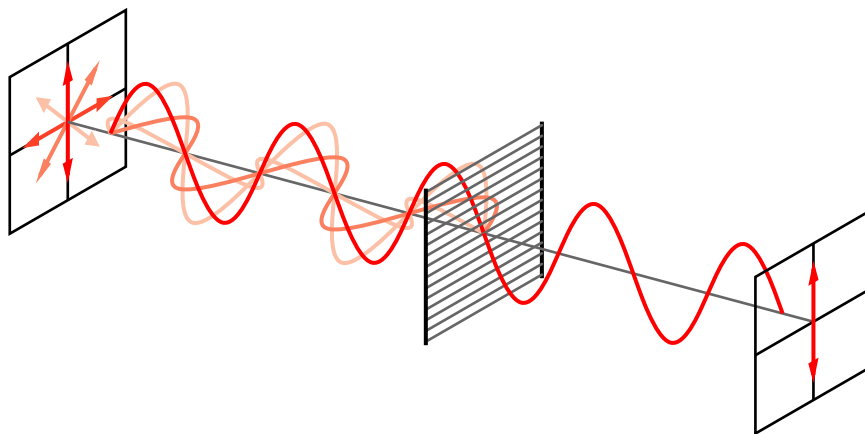


Figura 10.2: Polarizador de alambres usado para microondas (de Wikimedia Commons, archivo Wire-grid-polarizer.svg)

en general distinta que la polarización de la onda incidente. Los casos o “modos” de polarización, son (ver Fig. 10.3)

- polarización s , también llamada TE o transverso eléctrico, porque el campo eléctrico incidente es perpendicular al plano de incidencia. En el sistema de coordenadas de la figura es la componente $\hat{z}$;
- polarización p , también llamada TM, o transverso magnético, porque el campo magnético incidente es perpendicular al plano de incidencia (y el campo eléctrico incidente está contenido en el plano de incidencia). En el sistema de coordenadas de la figura es la componente en la dirección del producto $\hat{k}_i \times \hat{z}$.

En la Fig. 10.4 se muestran las curvas de los coeficientes de reflexión TE y TM en función del ángulo de incidencia para una dioptra plana aire-vidrio ($n_2 = 1.5$, $n_1 = 1$). Estos coeficientes dan la fracción de potencia incidente que se va en la onda reflejada. Notar que ambas curvas coinciden en incidencia normal, porque en el límite de $i \rightarrow 0$ no hay distinción entre modos TE y TM (porque no está definido el plano de incidencia). Usando las condiciones de contorno adecuadas (electromagnéticas), se puede demostrar que en incidencia normal el módulo del coeficiente de reflexión vale

$$R(i = 0) = \left| \frac{n_2 - n_1}{n_2 + n_1} \right|^2,$$

que sólo depende del índice relativo y que da lo mismo para incidencia desde el medio 1 hacia el medio 2 que para incidencia desde el medio 2 hacia el medio 1. Esta expresión muestra que la onda reflejada en la primera cara de una ventana de vidrio ($n_2 = 1.5$, $n_1 = 1$) se lleva un 4% de la energía incidente (y como la intensidad es proporcional al cuadrado de la amplitud, podemos decir que el módulo del campo eléctrico de la onda reflejada tiene una amplitud 0.2 veces menor que el del campo eléctrico de la onda incidente). Ambas curvas también coinciden en incidencia rasante donde la reflectividad vale 1. Este es el comportamiento usual en óptica, donde los medios son no magnéticos. La diferencia más importante entre modos TE y TM es que la reflectividad para el modo TM se anula para un ángulo de incidencia i_B conocido como **ángulo de Brewster**, mientras que la reflectividad para el modo TE nunca se anula. En el ángulo de Brewster el rayo reflejado (que en esta polarización no se refleja!) habría sido perpendicular al rayo refractado. Por eso, aplicando Ibn Sahl resulta

$$\tan i_B = \frac{n_2}{n_1}. \quad (10.12)$$

En el ángulo de Brewster, la onda con campo eléctrico en el plano de incidencia no se refleja en absoluto, sino que se refracta en su totalidad, una característica que se usa en las llamadas “ventanas Brewster”. Si la onda incidente tiene polarización natural, la luz reflejada en el ángulo de Brewster queda totalmente polarizada con campo eléctrico perpendicular al plano de incidencia. En cambio, la luz refractada queda parcialmente polarizada en dirección paralela al plano de incidencia (es una mezcla de la componente TM de la onda incidente, que se refracta en su totalidad, y de una fracción de la componente TE de la onda incidente, que no se refracta en su totalidad porque una parte sí se reflejó).

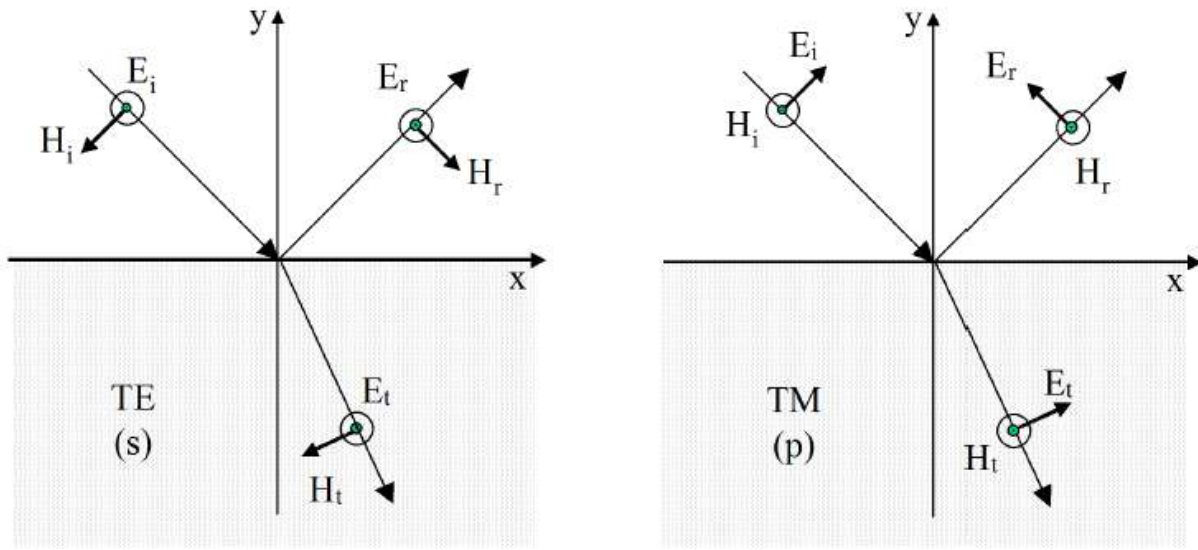


Figura 10.3: Modos de polarización en una superficie plana. E indica campo eléctrico, H indica campo magnético

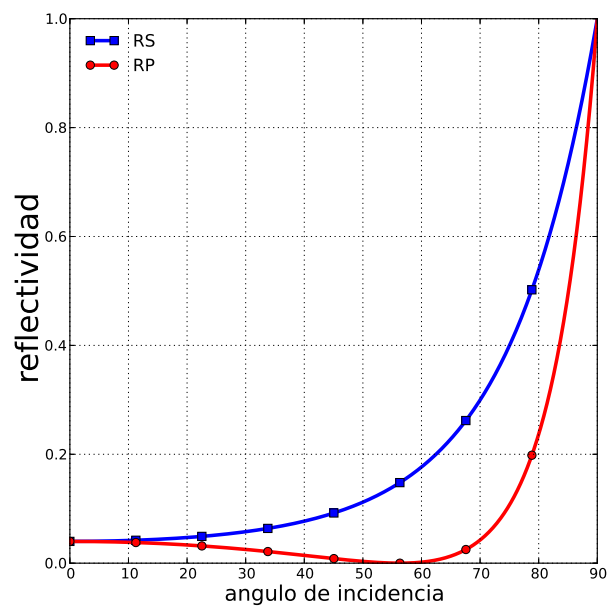


Figura 10.4: Coeficientes de reflexión en función del ángulo de incidencia para polarización s (TE) y p (TM). El medio de incidencia es aire $n_1 = 1$, el de transmisión es vidrio $n_2 = 1.5$

Es claro que la distinta respuesta para cada componente produce el efecto de polarización por reflexión. Justamente esta es la base del diseño de filtros polarizadores en los anteojos de sol. Cuando la luz solar se refleja en una superficie horizontal, el plano de incidencia es vertical y la luz reflejada contiene preponderantemente luz polarizada en dirección horizontal. Cuando la reflexión tiene lugar en la superficie lisa de una carretera de asfalto o en la superficie de un lago, produce una reverberación indeseable que deslumbra al observador interesado en las reflexiones que provienen de otros objetos (autos, barcos). La visión mejora si se elimina esta reverberación y por eso los fabricantes de lentes colocan en posición vertical el eje de polarización del material polarizador con que están hechas las lentes de los anteojos.

La Fig. 10.5 muestra las curvas para una dioptra vidrio-aire, en vez de aire-vidrio. Notar que el resultado es cualitativamente parecido al de la Fig. 10.4, pero la reflectividad toma el valor 1 en todo el intervalo entre el ángulo crítico de reflexión total e incidencia rasante. Claramente el ángulo crítico de reflexión total tiene que ser mayor que el de Brewster (hay argumentos físicos ... y además se ve en el resultado numérico ... y además **se puede demostrar analíticamente** usando las expresiones conocidas para cada ángulo: hacerlo).

Sabiendo electromagnetismo se demuestra que en medios no magnéticos los coeficientes de reflexión complejos para las amplitudes vienen dados por las siguientes expresiones

$$r^s = \frac{\beta_i - \beta_t}{\beta_i + \beta_t} = r_{12}^s \quad (10.13)$$

$$r^p = \frac{\beta_i/n_1^2 - \beta_t/n_2^2}{\beta_i/n_1^2 + \beta_t/n_2^2} = r_{12}^p \quad (10.14)$$

donde se usa la notación de las ecuaciones (9.24) y (9.34) y los subíndices 1 y 2 a la derecha de las igualdades recuerdan que la reflexión ocurre cuando la onda pasa desde el medio 1 al medio 2. La reflectividad $R(i)$ de la figura 10.4 es el módulo al cuadrado de los coeficientes r^s y r^p dados por las expresiones (10.13) y (10.14). Es más, usando los resultados de la ec. 9.39 es muy fácil comprobar que estos coeficientes son complejos cuando el ángulo de incidencia es mayor que el crítico, demostrando así lo que dije en las clases sobre los desfases introducidos en el régimen de reflexión total para cada polarización.

A continuación dejo de regalo un script de Python para calcular los coeficientes r^s y r^p y a partir de ellos la reflectividad que se muestra en las figuras 10.4 y 10.5.

```
# Coeficientes de reflexión vs ángulo de incidencia r_{12}
# Clase 21 del curso de F2
# Modos s y p pol, ver Figura 21.2 y 21.3 en apuntes
# Author R. Depine <http://users.df.uba.ar/rdep/>
#
import numpy as np
import math
import cmath
import matplotlib.pyplot as plt
```

```

pi=math.pi
Uno=1+0j

##### Entradas

ind1=1.
ind2=1.5
teta0=0.
tetaf=89.99
ihow=800

tetagrid=np.linspace(teta0,tetaf,ihow)
tetarad=tetagrid*math.pi/180

beta1=ind1*np.cos(tetarad)
beta2=np.sqrt(Uno*ind2**2-ind1**2*np.sin(tetarad)**2)
rs=(beta1-beta2)/(beta1+beta2)
rp=(beta1/ind1**2-beta2/ind2**2)/(beta1/ind1**2+beta2/ind2**2)
RS=np.absolute(rs)**2
RP=np.absolute(rp)**2

### resultados a un archivo con encabezamiento y tabla
f=open('clase21.ext','w')
print >>f, "===== "
print >>f, "Script para la Figura 2 de la Clase 20 - F2 2016 2C"
print >>f, "===== "
print >>f, ind1, ind2
print >>f, "ang, RS, RP"

for IJK in range(ihow):
    print >>f, "%g %g %g " %(tetagrid[IJK], RS[IJK], RP[IJK])

f.close()

### Mostramos los resultados

fig = plt.figure(figsize=(8, 8))
plt.grid()
plt.xlabel(r'ángulo de incidencia', fontsize=18)
plt.ylabel(r'reflectividad', fontsize=26)
plt.plot(tetagrid, RS, c='b', ls='-', lw=3, marker='s', markevery=100, ms=7)
plt.plot(tetagrid, RP, c='r', ls='-', lw=3, marker='o', markevery=100, ms=7)
plt.legend(['RS', 'RP'], loc='upper left', frameon=False)
plt.draw()
plt.savefig('Clase21Fig2.eps', format='eps', dpi=300)
plt.savefig('Clase21Fig2.jpg', format='jpg')

```

Para saber si es posible transformar en circularmente polarizada una onda que incide linealmente polarizada a 45 grados del plano de incidencia (en la clase anticipé que no), deberíamos poder evaluar los desfases para cada polarización en el régimen de reflexión total y a partir de ellos el desfase relativo entre las componentes s y p.

10.8. Láminas retardadoras

Para controlar la diferencia de fase ϕ entre las dos componentes transversales de las eqs. (10.2) y (10.3), y controlar así el estado de polarización, se usan láminas fabricadas con materiales anisótropos, que tienen distinto índice de refracción para componentes de polarización paralelas o perpendiculares a un eje preferencial del material, asociado con su estructura cristalina. Si la onda descrita por las eqs. (10.2) y (10.3) incide sobre una lámina de este tipo y orientada para que sus ejes preferenciales coincidan con los ejes x e y, para describir la propagación de la onda dentro de la lámina hay que tener en cuenta que la velocidad de propagación para la componente x no es igual a la velocidad de propagación para la componente y. En consecuencia, si el número de onda $k = \omega/c$ era el mismo para describir la propagación en el vacío de las dos componentes de la onda transversal, dentro de la lámina hay dos números de onda, uno para cada componente

$$k_x = \frac{\omega}{c} n_x, \quad k_y = \frac{\omega}{c} n_y. \quad (10.15)$$

De esta manera, mientras que en el vacío las dos componentes mantienen su diferencia de fase (en cada z mantienen su estado de polarización), en la lámina las dos componentes se van desfasando a medida que avanzan. En particular, a la salida de la lámina, luego de atravesar un espesor d , su diferencia de fase será la que tenían a la entrada (ϕ), más un desfase adicional, igual a

$$\Delta\phi = (k_y - k_x)d = \frac{\omega}{c}(n_y - n_x)d = \frac{2\pi}{\lambda_0}(n_y - n_x)d. \quad (10.16)$$

Si se elige el espesor de la lámina para que $\Delta\phi$ sea congruente a $\pm\pi/2$ (un cuarto de toda una vuelta), se dice que se tiene una **lámina de cuarto de onda**. Una lámina de este tipo transforma una onda linealmente polarizada a 45° de los ejes x - y en una onda circularmente polarizada y una onda elípticamente polarizada con ejes principales x - y en una onda linealmente polarizada.

Si se elige el espesor de la lámina para que $\Delta\phi$ sea congruente a $\pm\pi$ (media vuelta), se dice que se tiene una **lámina de media onda**. Esta lámina se puede usar para cambiar el sentido de giro de una onda elípticamente polarizada o para rotar el plano de polarización de una onda linealmente polarizada.

10.9. Birrefringencia en medios quirales

Así como la estructura interna de los medios anisótropos da como resultado la existencia de distintos índices de refracción para componentes de polarización en la dirección de ciertos ejes preferenciales, en la naturaleza existen medios cuya estructura interna tiene un sentido de giro privilegiado y esto produce que las ondas polarizadas circularmente con un sentido de giro tengan velocidades de propagación distintas que las que tienen sentido de giro opuesto. Este tipo de medios son de interés en la física de polímeros, en la industria farmacéutica o en química biológica, ya que las grandes moléculas orgánicas tienen generalmente asociado un sentido de giro.

Que las ondas polarizadas circularmente con sentidos de giro opuestos tengan distinta velocidad de propagación es una propiedad común a todo medio constituido microscópicamente por objetos que no tienen simetría de reflexión. Los medios formados por hélices izquierdas o derechas son un ejemplo, pero lo mismo ocurre con cualquier otro medio formado por colecciones de objetos tales que cada objeto difiere de su imagen especular. La palabra **quiral** tiene raíz griega, y proviene de **mano** (la mano izquierda no coincide con su imagen especular, que es la mano derecha). Este fenómeno es la base de lo que se conoce como actividad óptica: una onda linealmente polarizada entra en un medio quiral, se propaga, y al salir del mismo se comprueba que ha rotado su dirección de vibración (ver video en <http://bit.ly/1PV5vvu>).

Mediante un análisis similar al que hicimos en 9.6.3, ver figuras 9.11 y 9.12, es fácil ver que en la superficie plana entre un medio común (aquiral) y otro quiral, una onda incidente cuya dirección de propagación forma un ángulo no nulo con la normal a la superficie, da origen, por lo general, a dos ondas transmitidas que viajan en direcciones

distintas. La existencia de dos ondas transmitidas para una misma onda incidente se conoce como **birrefringencia**. Cuando el medio tiene un sentido de giro derecho, la onda circularmente polarizada derecha se propaga más rápido que la izquierda: $v_D > v_I$, en consecuencia $k^{(I)} > k^{(D)}$. Inversamente, cuando el medio tiene un sentido de giro izquierdo, la onda circularmente polarizada izquierda se propaga más rápido que la derecha: $v_I > v_D$, en consecuencia $k^{(D)} > k^{(I)}$. El punto importante es darse cuenta que en el medio quiral puede haber dos ondas circularmente polarizadas con sentidos de rotación opuestos. Esto se indica en las figuras mediante las esferas D e I. Las direcciones de propagación de estas dos ondas transmitidas quedan determinadas como en el caso de medios comunes, solamente pidiendo la conservación de las componentes tangenciales de los vectores de onda.

Dependiendo del tipo de medios, pueden darse situaciones en las que:

- la onda incidente da lugar a dos ondas transmitidas; una circular izquierda y otra circular derecha y las direcciones de ambas (respecto de la dirección incidente) se acercan a la normal. En este caso, que es el mostrado en la figura anterior, siempre hay dos ondas transmitidas para todo ángulo de incidencia.
- la onda incidente da lugar a dos ondas transmitidas; pero con respecto a la dirección incidente la onda circularmente polarizada izquierda se acerca a la normal, mientras que la onda circularmente polarizada derecha se aleja de la normal. En este caso, existe un ángulo de incidencia crítico a partir del cual la onda circularmente polarizada derecha se transforma en evanescente (pero la izquierda nunca se hace evanescente).
- la onda incidente da lugar a dos ondas transmitidas y tanto la circularmente polarizada izquierda como la derecha se alejan de la normal con respecto a la dirección incidente. En este caso, existe un ángulo crítico de incidencia a partir del cual la onda circularmente polarizada derecha se transforma en evanescente (y la izquierda no evanescente) y otro ángulo crítico mayor a partir del cual ambas ondas transmitidas se transforman en evanescentes (reflexión total).

10.10. Birrefringencia en cristales

En la clase vimos que la existencia de dos ondas transmitidas para una misma onda incidente también ocurre en la superficie plana entre un medio isótropo común y un medio anisótropo como el que se usa para construir láminas retardadoras. Las láminas retardadoras se usan en incidencia normal y por eso la birrefringencia no es evidente. Pero cuando la incidencia no es normal, se observan dos rayos transmitidos dentro del cristal y además las ondas asociadas a estos rayos están linealmente polarizadas en direcciones ortogonales, una propiedad que se usa para construir dispositivos polarizadores (ver, por ejemplo prismas de Nicol y de Wollaston, sección 8.4 del libro de Hecht y Zajac).

Para pensar

¿Cómo se debe proceder experimentalmente para determinar unívocamente el estado de polarización de una muestra incógnita?

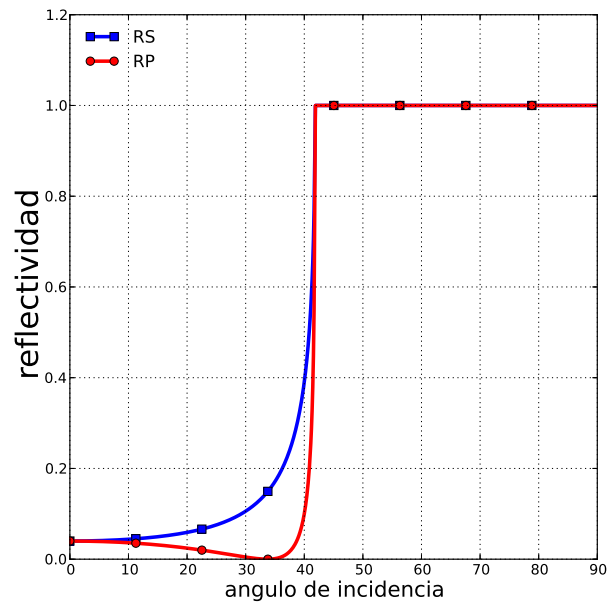


Figura 10.5: Coeficientes de reflexión en función del ángulo de incidencia para polarización s (TE) y p (TM). El medio de incidencia es vidrio $n_1 = 1.5$ y el de transmisión aire $n_2 = 1$.

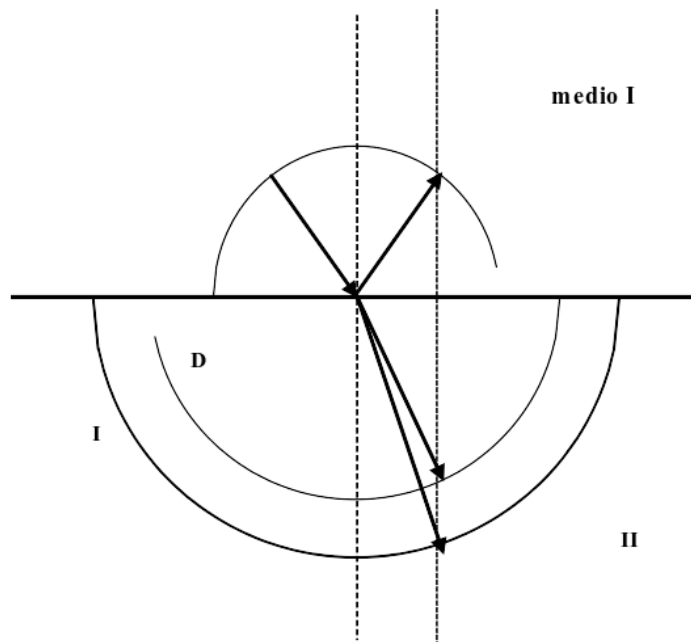


Figura 10.6: Birrefringencia en superficie aquiral-quiral, los dos índices de refracción de las ondas circularmente polarizadas izquierda y derecha son mayores que el índice de refracción del medio de incidencia.

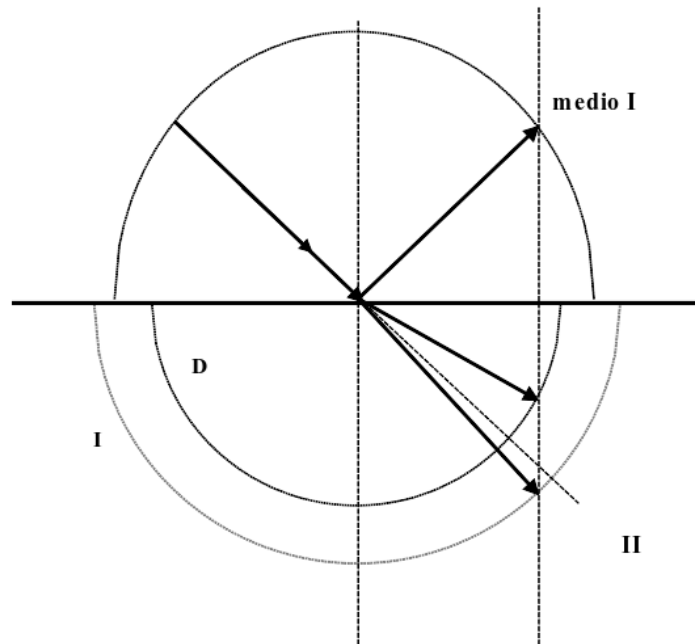


Figura 10.7: Birrefringencia en superficie aquiral-quiral, el valor del índice de refracción del medio de incidencia es intermedio entre los dos índices de refracción de las ondas circularmente polarizadas.

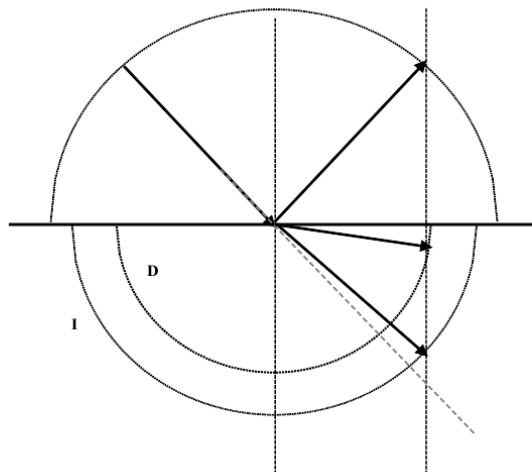


Figura 10.8: Birrefringencia en superficie aquiral-quiral, el valor del índice de refracción del medio de incidencia es mayor que los dos índices de refracción de las ondas circularmente polarizadas.

Capítulo 11

Interferencia

La ecuación de ondas clásica es lineal, entonces vale el principio de superposición. Esto quiere decir que si en el punto P estamos observando la perturbación producida por dos fuentes S_1 y S_2 (ver Fig. 11.1, la perturbación medida tiene que ser igual a la suma de la perturbación producida cuando la fuente S_1 estaba encendida y la fuente S_2 apagada, mas la perturbación producida cuando la fuente S_2 estaba encendida y la fuente S_1 apagada. La superposición de dos o más ondas en una región del espacio da lugar a una onda resultante que es la suma de las perturbaciones individuales. Cuando esto sucede se dice que hay **interferencia**. Desde este punto de vista, siempre se produce interferencia entre dos o más ondas regidas por ecuaciones de onda lineales. Pero con luz natural no es fácil que esta interferencia pueda ser detectada por los sentidos humanos, probablemente por eso se tardó tanto tiempo en evidenciar experimentalmente el carácter ondulatorio de la luz.

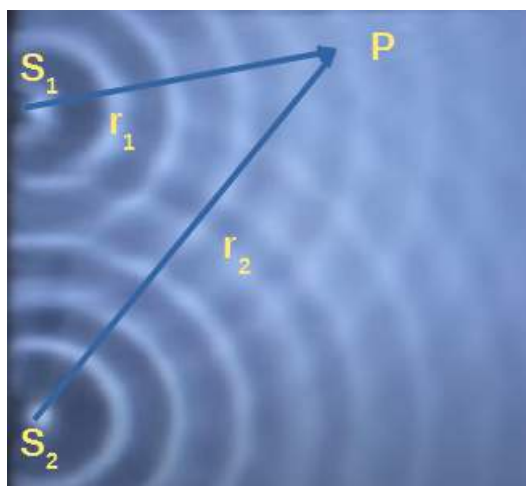


Figura 11.1: Esquema conceptual de la superposición con dos fuentes puntuales.

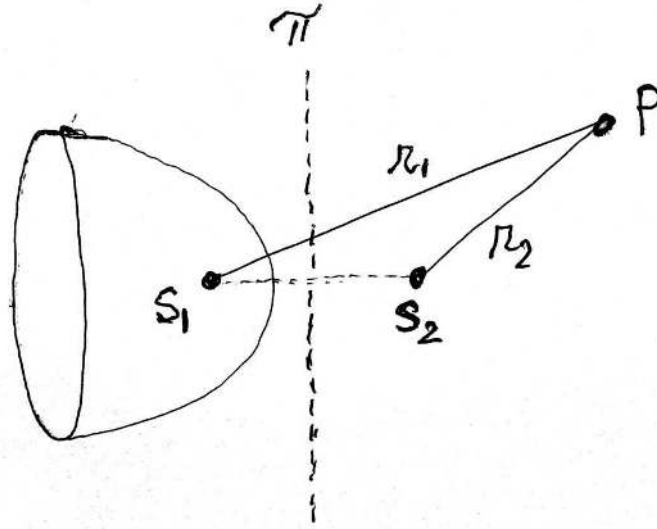


Figura 11.2: Las perturbaciones de dos fuentes puntuales interfieren en el punto P.

11.1. Fuentes puntuales

Empezamos con el caso más sencillo, analicemos la interferencia entre las perturbaciones producidas por dos fuentes puntuales monocromáticas de frecuencia ω . Cada fuente produce una onda esférica progresiva, como se vió en la eq. (7.28). Al punto P llegan dos perturbaciones de la forma

$$\begin{aligned}\psi_1(r_1, t) &= a_1 \cos(\omega t - kr_1 + \delta_1) = a_1 \cos(\omega t + \Delta_1), \\ \psi_2(r_2, t) &= a_2 \cos(\omega t - kr_2 + \delta_2) = a_2 \cos(\omega t + \Delta_2),\end{aligned}$$

donde $\Delta_1 = \delta_1 - kr_1$, $\Delta_2 = \delta_2 - kr_2$, $k = \omega/v$, r_1 y r_2 se miden desde cada fuente al punto P, a_1 y a_2 dependen de las intensidades de cada fuente y de r_1 y de r_2 y δ_1 y δ_2 son fases iniciales que se introducen para tener en cuenta que las fuentes se pueden encender en tiempos distintos y en consecuencia que pueden estar desfasadas.

La perturbación total en el punto P es

$$\psi_P(t) = \psi_1(r_1, t) + \psi_2(r_2, t)$$

y tiene que tener la forma

$$\psi_P(t) = a \cos(\omega t + \Delta). \quad (11.1)$$

Para ver que esto es así basta con encontrar la amplitud a y la fase inicial Δ de la onda resultante en función de amplitudes a_1 y a_2 y fases iniciales Δ_1 y Δ_2 de las ondas individuales. Notemos por un lado que $\psi_1(r_1, t) + \psi_2(r_2, t)$ se puede escribir como la suma de dos términos: uno que varía como $\cos \omega t$ y otro que varía como $\sin \omega t$. Por otro lado, $\psi(r, t)$ dado por la ec. (11.1) también se puede escribir de la misma manera. Entonces, para que la superposición valga en todo instante, el coeficiente del término

que varía como $\cos \omega t$ en $\psi_1(r_1, t) + \psi_2(r_2, t)$ tiene que ser igual al coeficiente del término que varía como $\cos \omega t$ en $\psi(r, t)$. Y lo mismo tiene que pasar para los coeficientes de los términos que varían como $\sin \omega t$. Al igualar estos coeficientes se obtienen las siguientes condiciones

$$\begin{aligned} a \cos \Delta &= a_1 \cos \Delta_1 + a_2 \cos \Delta_2, \\ a \sin \Delta &= a_1 \sin \Delta_1 + a_2 \sin \Delta_2, \end{aligned}$$

un sistema de ecuaciones con el que estamos familiarizados, hay que elevar al cuadrado miembro a miembro cada ecuación y luego sumar para obtener

$$a^2 = a_1^2 + a_2^2 + 2 a_1 a_2 \cos \phi, \quad (11.2)$$

donde la diferencia de fase ϕ entre las dos perturbaciones en el punto P

$$\phi = \Delta_2 - \Delta_1 = \delta_2 - \delta_1 + k(r_1 - r_2), \quad (11.3)$$

tiene dos aportes: uno $(\delta_2 - \delta_1)$ proviene de las fases iniciales de las fuentes, y el otro $(k(r_1 - r_2))$ proviene de la diferencia de camino óptico (porque k es proporcional al índice de refracción) entre los caminos recorridos por las dos perturbaciones. Análogamente, se podría hallar Δ en términos de a_1 , a_2 , Δ_1 y Δ_2 , pero por el momento no nos hace falta.

Como en todo fenómeno ondulatorio, la intensidad en el punto P, I_P , es proporcional a a^2 , el cuadrado de la amplitud de la perturbación resultante en dicho punto. En cambio, cuando solo está encendida la fuente 1, la intensidad I_1 que se mide en el punto P es proporcional a a_1^2 y cuando solo está encendida la fuente 2, la intensidad I_2 en el punto P es proporcional a a_2^2 . Entonces, en términos de las intensidades, la ec. (11.2) se puede reescribir como

$$I_P = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \phi \quad (11.4)$$

que dice que la intensidad en el punto P cuando están encendidas las dos fuentes simultáneamente **no es la suma de las intensidades** producidas por cada fuente por separado. Esto es así porque, además de la suma de intensidades individuales (proporcionales a a_1^2 y a a_2^2), hay un tercer término que puede ser positivo o negativo dependiendo del valor de ϕ . Este tercer término se llama **término de interferencia**. O sea, que la intensidad en el punto P tiene máximos y mínimos dependiendo del valor de la diferencia de fase ϕ entre las dos perturbaciones en dicho punto. Cuando ϕ es congruente con 2π el término de interferencia toma su máximo valor y entonces la intensidad en el punto P toma el máximo valor posible, proporcional a

$$a_{max}^2 = a_1^2 + a_2^2 + 2a_1 a_2 = (a_1 + a_2)^2. \quad (11.5)$$

En este caso se dice que la interferencia es **constructiva**. En cambio, cuando ϕ es congruente con π , el término de interferencia toma su mínimo valor, la intensidad en el punto P toma el mínimo valor posible, proporcional a

$$a_{min}^2 = a_1^2 + a_2^2 - 2a_1 a_2 = (a_1 - a_2)^2, \quad (11.6)$$

y se dice que la interferencia es **destructiva**. Notar que en el caso $a_1 = a_2$, los mínimos valen cero y podemos decir que “luz + luz = oscuridad”. Suena raro, porque no es usual observar que luz + luz da oscuridad. Y esto pasa porque estamos habituados a un mundo donde las fuentes de luz son fuentes térmicas, como las que mencionamos cuando hablamos de luz natural en la página 180. Y entonces, en vez de producir dos perturbaciones $\psi_1(z, t)$ y $\psi_2(z, t)$ como supusimos, en forma de coseno con fases Δ_1 y Δ_2 constantes en el tiempo, cada fuente térmica produce “trenes de onda”, es decir, cosenos modulados por una función con una duración τ , que para la luz suele ser $\approx 10^{-8}\text{s}$. Por este motivo, la diferencia de fase $\Delta_2 - \Delta_1$ cambia aleatoriamente en cada intervalo τ y entonces hay $\tau^{-1} \approx 10^8$ cambios de intensidad en un segundo, algo imposible de detectar con detectores comunes o a simple vista. Por este motivo se dice que las fuentes térmicas son **incoherentes**, en el sentido de que no pueden exhibir fenómenos de interferencia como los predichos por las ecs. (11.5) y (11.6). Es claro que esto es una convención impuesta por la resolución de los aparatos a los que estamos acostumbrados, porque si pudiéramos resolver tiempos muy pequeños, estaríamos mas familiarizados con las variaciones de intensidad producidas por interferencia.

11.2. Uso de fasores

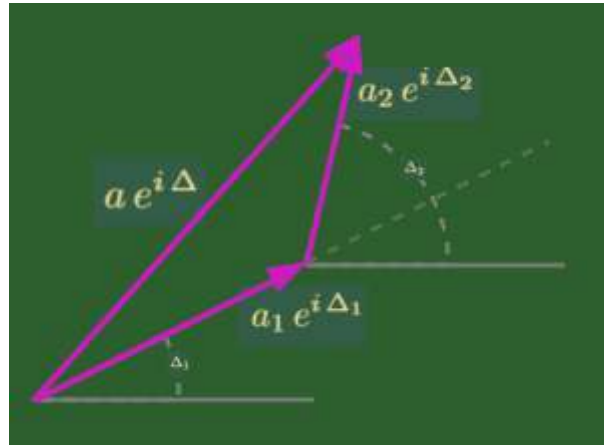


Figura 11.3: La suma de fasores es equivalente a la suma de dos perturbaciones armónicas.

Una vez entendido el procedimiento para encontrar la resultante de la suma de las dos perturbaciones armónicas usadas en (11.1), veamos cómo se puede simplificar el álgebra, pagando el precio de introducir notación compleja. Escribiendo los cosenos como la parte real de exponenciales imaginarias, obtenemos

$$\psi_1(r_1, t) = a_1 \cos(\omega t + \Delta_1) = \text{Re } a_1 e^{i(\omega t + \Delta_1)},$$

$$\psi_2(r_2, t) = a_2 \cos(\omega t + \Delta_2) = \text{Re } a_2 e^{i(\omega t + \Delta_2)}.$$

Con esta notación, y omitiendo Re (pero recordando que habrá que tomar la parte real de los resultados que obtengamos), la suma que hicimos en notación trigonométrica se

puede expresar como

$$\psi_P(t) = \sum_{n=1}^2 a_n e^{i(\omega t + \Delta_n)} = e^{i\omega t} \sum_{n=1}^2 a_n e^{i\Delta_n} = a e^{i\Delta} e^{i\omega t}. \quad (11.7)$$

Observando que el factor $e^{-i\omega t}$ está en todas partes, lo simplificamos y vemos que se puede interpretar la suma de oscilaciones armónicas como suma de dos números complejos, cada uno con su módulo y su fase, y que el resultado es otro número complejo, con módulo a y fase Δ . Y usando la representación de los números complejos como vectores del plano (el plano complejo), tenemos que el esquema de la sección anterior es completamente equivalente al esquema mostrado en la figura 11.3. Así vemos que la amplitud resultante dada por (11.2) no es otra cosa que la aplicación del teorema del coseno a la resultante que se obtiene de sumar los dos vectores del plano complejo. Y que este resultado no depende de la orientación de los vectores 1 y 2, solamente depende de su orientación relativa que viene dada por la diferencia $\Delta_2 - \Delta_1$ entre los ángulos de cada vector con el eje horizontal, que no es otra cosa que la diferencia de fase ϕ en (11.3).

11.3. Sistemas de franjas

La distribución espacial de máximos y mínimos de intensidad producida por la interferencia de distintas perturbaciones recibe el nombre de sistema de franjas. La visibilidad de los sistemas de franjas depende de la diferencia que haya entre las intensidades máxima y mínima. Se suele cuantificar la visibilidad V o contraste de un sistema de franjas mediante la cantidad

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \quad (11.8)$$

que varía entre 1 (máximo contraste, cuando la intensidad mínima es nula, es decir cuando $a_1 = a_2$) y cero (mínimo contraste, cuando la intensidad máxima y la mínima son iguales).

¿Cuál es el sistema de franjas asociado a dos fuentes puntuales coherentes? Para visualizar un sistema de franjas es conveniente determinar el lugar geométrico de los puntos con igual intensidad. Si las distancias de observación r_1 y r_2 son grandes comparadas con la longitud de onda, entonces la variación espacial más importante de a^2 es a través del coseno (ver (11.2)) y no a través de las amplitudes a_1 y a_2 . Entonces, el lugar geométrico de los puntos con igual intensidad es el de los puntos con igual valor de ϕ . Y según (11.3), esto corresponde a superficies con el mismo valor de $r_2 - r_1$. Si las fuentes emiten en fase, los puntos de intensidad máxima están en el plano $r_2 = r_1$, que es perpendicular al segmento que une S_1 con S_2 (el plano II en la figura 11.2). En este plano las dos perturbaciones llegan completamente en fase luego de haber recorrido el mismo camino óptico. Otros puntos de intensidad máxima están en los hiperboloides (de dos hojas) de revolución $r_2 - r_1 = \pm\lambda$, con focos en S_1 y S_2 , donde las dos perturbaciones llegan desfasadas en 2π . En general, el lugar geométrico de los puntos con intensidad constante es una familia de hiperboloides confocales, con focos en S_1 y S_2

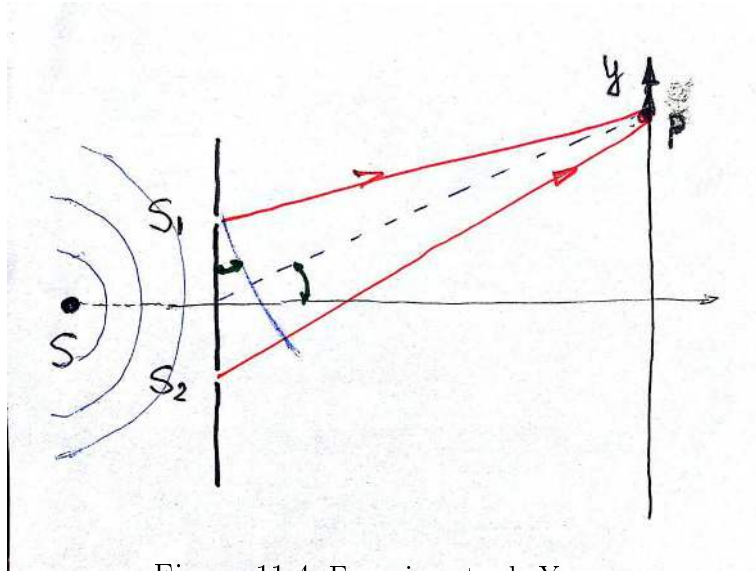


Figura 11.4: Experimento de Young.

(una de las hojas se indica en la figura 11.2). Si el sistema de franjas se observa en un plano perpendicular al segmento $\overline{S_1 S_2}$, se verán círcuncferencias concéntricas (anillos) con intensidad máxima y anillos con intensidad mínima. Y si el sistema de franjas se observa en un plano que contiene al segmento $\overline{S_1 S_2}$ (el eje del hiperboloide), se verán hipérbolas.

11.4. Experiencia de Young

Uno de los primeros experimentos cuantitativos encaminados a poner de manifiesto la interferencia de la luz proveniente de dos fuentes fue realizado en 1800 por el científico inglés Thomas Young. En la figura se esquematiza cómo hizo Young para obtener dos fuentes secundarias coherentes S_1 y S_2 a partir de una fuente de luz natural. Como las perturbaciones de diferentes partes de una fuente extensa ordinaria no están sincronizadas, se elige sólo una región de la fuente con un orificio estrecho S , de aproximadamente $1\mu\text{m}$ de ancho, en una pantalla opaca. Este orificio se comporta como una fuente puntual idealizada (en las versiones actuales de este experimento se usa un laser como fuente y este orificio no hace falta). La “fuente” S ilumina una pantalla con orificios S_1 y S_2 , separados entre sí una distancia d . A partir de S se propagan frentes de onda que recorren distancias iguales y que llegan a S_1 y S_2 en fase, por consiguiente se puede considerar que S_1 y S_2 son fuentes puntuales que están en fase, es decir que son fuentes coherentes. A partir de acá el análisis corresponde al de cualquier par de fuentes puntuales coherentes. En todo punto iluminado simultáneamente por las perturbaciones que salen de S_1 y S_2 se observará la interferencia de la luz, según como se interactúe con el sistema de hiperboloides asociado con S_1 y S_2 . Si esta interacción se hace con una placa fotográfica o con una pantalla ubicada a una distancia D y paralela a la pantalla con orificios, se visualizará la intersección con el sistema de hiperboloides, es decir trozos de hipérbolas.

Para simplificar el análisis del experimento de Young, se suele suponer que la distancia D es grande comparada con la distancia d entre S_1 y S_2 (en un experimento podría ser d aproximadamente unos pocos milímetros y D un metro o más). En este caso, la diferencia de camino $\overline{S_2P} - \overline{S_1P}$ satisface aproximadamente la relación

$$\frac{y}{D} \approx \frac{\overline{S_2P} - \overline{S_1P}}{d}, \quad (11.9)$$

Si suponemos que la “fuente” S equidista de los orificios S_1 y S_2 , entonces $\delta_2 = \delta_1$ y la diferencia de fase de la ec. (11.3) tiene solo el aporte de la diferencia de camino $\overline{S_2P} - \overline{S_1P}$. Cuando esta cantidad es igual a un número entero de longitudes de onda, la intensidad en el punto P alcanza su valor máximo posible, mientras que si esta cantidad es igual a un número impar de semilongitudes de onda, la intensidad en el punto P alcanza su mínimo valor posible. De estas consideraciones surge que

$$y_{max} \approx m \frac{\lambda D}{d}, \quad (11.10)$$

con m un número entero. Las perturbaciones que llegan a $y = 0$ interfieren constructivamente para todo valor de λ , entonces el máximo central no está coloreado si se hace el experimento con luz blanca. En cambio, las posiciones y_{max} donde hay interferencia constructiva para $m \neq 0$ dependen de λ , es decir que la posición del máximo para $m = 1$ del color azul no es la misma que la posición del máximo para $m = 1$ del color rojo y en consecuencia el sistema de franjas se ve coloreado cuando se usa luz blanca.

Si los orificios S_1 y S_2 tienen el mismo tamaño, podemos suponer que las amplitudes a_1 y a_2 son aproximadamente iguales en la zona central de la pantalla de observación (son exactamente iguales en $y = 0$) y la expresión (11.2) para a^2 , proporcional a la intensidad resultante, queda

$$a^2 = 2 a_1^2 (1 + \cos \phi) = 4 a_1^2 \cos^2 \frac{\phi}{2}, \quad (11.11)$$

y como sobre la pantalla

$$\phi = k (\overline{S_2P} - \overline{S_1P}) = \frac{k d}{D} y, \quad (11.12)$$

concluimos que la intensidad en la pantalla varía con y como un $\cos^2$. Si sobre la pantalla se coloca una película fotográfica, la transparencia del negativo varía con y como un $\cos^2$. Y si se coloca un material como las fotoreinas usadas en holografía, cuya solubilidad en un revelador es proporcional a la intensidad luminosa recibida, se obtiene una superficie rugosa con forma de seno (este dispositivo se llama red holográfica).

11.5. Coherencia

La falta de coherencia de las fuentes luminosas que nos rodean habitualmente es la gran dificultad para observar franjas de interferencia en la vida cotidiana. Ya vimos

que esta dificultad fue resuelta por Young mediante el dispositivo esquematizado en la Fig. 11.4. El dispositivo de Young divide el frente de onda generado por la fuente térmica, con fase inicial que cambia muy rápidamente, en dos frentes de onda que pasan a ser coherentes, porque su diferencia de fase permanece constante en el tiempo. Esta solución resuelve el problema de la coherencia temporal, porque ahora las dos porciones en que se dividió el frente de onda original pueden interferir y dar máximos y mínimos de intensidad, como si provinieran de dos fuentes coherentes. Pero no tenemos que perder de vista que los trenes de onda siguen teniendo una longitud finita y que la fase inicial en las dos porciones en que se dividió cada tren siguen cambiando rápidamente. El truco usado por Young consigue que en ambas porciones la fase inicial cambie de la misma manera y así la cantidad $\delta_1 - \delta_2$ (página 194) se mantiene constante. Esta afirmación es claramente cierta para el punto de la pantalla de observación que equidista de S_1 y S_2 y también para cualquier otro punto P en el que la diferencia de camino $r_1 - r_2$ no sea comparable con la longitud de un tren de onda. Pero si el punto P se aleja mucho del centro de la pantalla, o si la fuente primaria S se coloca de manera asimétrica con respecto a S_1 y S_2 , podría ocurrir que en el punto P se superpongan perturbaciones provenientes de dos trenes de onda diferentes. Y si pasa esto, el truco de Young deja de funcionar, porque la fase ϕ de la ec. (11.3) pasa a cambiar muy rápidamente, porque lo hace la diferencia de fase $\delta_2 - \delta_1$. Por este motivo, para que el truco de Young funcione, las diferencias de camino acumuladas desde la fuente al punto de observación deben ser menores que la longitud de coherencia

$$\overline{SS_2} + \overline{S_2P} - \overline{SS_1} - \overline{S_1P} < v\tau. \quad (11.13)$$

De esta manera, la coherencia temporal de la fuente se traduce en una condición espacial, asociada a la duración limitada de los trenes de onda. En la práctica aparece otra dificultad, relacionada con otro tema: el tamaño finito de la fuente. Esto sucede porque las perturbaciones de diferentes partes de una fuente extensa ordinaria no están sincronizadas. En este caso se habla de coherencia espacial, que es una condición distinta a la expresada en la ec. (11.13). En el caso de fuentes extensas, los orificios S_1 y S_2 deben estar iluminados por frentes de onda sincronizados, no por frentes de onda provenientes de puntos desincronizados de la fuente extensa (por este motivo Young usó como fuente S un orificio muy pequeño, si hubiera usado el Sol como fuente, habría fracasado como fracasó Grimaldi que en 1665 había intentado algo similar).

11.6. Interferómetros

Los interferómetros son dispositivos para ver sistemas de franjas, o sea, para controlar problemas de coherencia. El interferómetro de Young es el padre de todos los interferómetros.

11.6.1. División del frente de onda

Las mismas consideraciones matemáticas y físicas que discutimos en el experimento de Young, se aplican directamente en otras configuraciones donde el frente de ondas proveniente de una fuente puntual primaria se divide en dos frentes de onda que interfieren entre sí y que parecen provenir de dos fuentes puntuales secundarias. Por eso este tipo de interferómetros se llaman *de división del frente de onda*. Hay muchos ejemplos, como el espejo doble de Fresnel, biprisma de Fresnel, o espejo de Lloyd, muy bien descritos en la última parte de la sección 9.3 del libro de Hecht, Óptica (páginas 397-399 de la traducción al español, misma numeración de páginas en la versión en inglés, cuarta edición).

Entre los interferómetros de división del frente de onda merece destacarse, por razones históricas, el espejo doble de Fresnel, que jugó un papel muy importante para invalidar las críticas que Young recibió de sus contemporáneos. Cuando Young publicó su trabajo en 1804, la comprensión más ampliamente aceptada de la naturaleza de la luz era la teoría corpuscular de Newton. En esta visión, la luz era una corriente de partículas que podrían agitar el éter e interactuar con objetos materiales a través de fuerzas atractivas y repulsivas. En ese momento, la mayoría de las personas interesadas en óptica seguía a Newton sin discusiones, excepto algunos visionarios como Young en Inglaterra, y Arago y Fresnel en Francia. Para ellos la luz era una onda elástica en el éter. Hoy podríamos pensar que el experimento de Young es una prueba irrefutable de la naturaleza ondulatoria de la luz. Pero ese no fue el caso, no es fácil cambiar los paradigmas, especialmente cuando son sostenidos por autoridades como Newton. Los seguidores de la teoría corpuscular argumentaban que al pasar por las dos rendijas estrechas, las partículas de luz habían interactuado de manera mecánica con el material de los bordes de las rendijas O sea, no se querían convencer. El experimento del doble espejo de Fresnel en 1816 tenía la gran virtud de acabar con las rendijas por completo, y así proporciona más evidencia de que las franjas se producen, no por la acción de los bordes sino por la superposición de dos haces de luz . Con este experimento y otros trabajos teóricos y experimentales de Fresnel, la teoría ondulatoria de la luz gradualmente ganó preeminencia, y hacia 1830 comenzó a aceptarse que la teoría ondulatoria era más poderosa, porque explicaba más fenómenos observados, que la corpuscular.

11.6.2. División de amplitud

En los *interferómetros de división de amplitud*, el frente de ondas proveniente de una fuente puntual primaria se divide, mediante reflexiones y transmisiones parciales (no totales como en el espejo de Lloyd o en los espejos dobles de Fresnel) en láminas u otros objetos semitransparentes. Así se logran dos frentes de onda que por lo general tienen distinta amplitud, pero que son coherente entre sí debido a que provienen del mismo frente de ondas y por eso mantienen su diferencia de fase constante en el tiempo, siempre que no recorran caminos que difieran entre sí en cantidades mayores que la longitud de coherencia de la fuente utilizada. Un ejemplo de interferómetro por división de amplitud es el Interferómetro de Michelson.

Interferómetro de Michelson

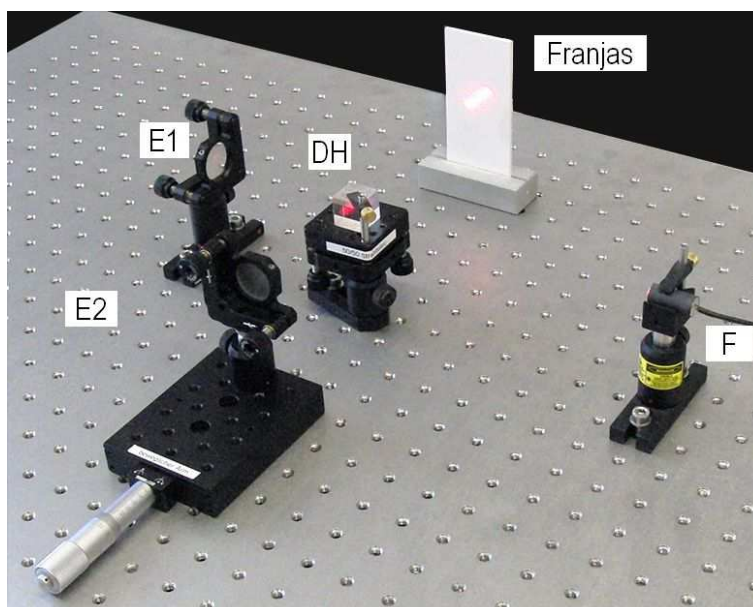


Figura 11.5: Interferómetro de Michelson (de Wikimedia Commons, archivo Aufbau-Michelson-Interferometer.jpg).

Además de sus múltiples aplicaciones actuales, este dispositivo experimental tiene importancia histórica ya que hace un siglo aportó uno de los puntales experimentales clave de la teoría de la relatividad. Hoy se continúa usando para mediciones muy precisas de la longitud de onda y en metrología óptica en general, cuando se necesita medir distancias muy pequeñas. Con el objetivo de detectar ondas gravitacionales, en el observatorio de LIGO se ha mejorado la sensibilidad hasta un nivel sin precedentes. LIGO, siglas de Laser Interferometer Gravitational-Wave Observatory, es decir, Observatorio de ondas gravitacionales por interferometría láser, utiliza largos interferómetros de tipo Michelson a fin de detectar ondas gravitacionales provenientes de agujeros negros, estrellas de neutrones, supernovas o remanentes del Big Bang.

Al igual que el interferómetro de Young, el interferómetro de Michelson se puede usar con una fuente de luz monocromática. La onda proveniente de esta fuente se divide en dos ondas que siguen caminos diferentes. Pero mientras que en el interferómetro de Young la división se hace enviando parte de la luz a través de una ranura y parte a través de otra; en el interferómetro de Michelson se emplea un dispositivo llamado divisor de haz, por eso pertenece a la familia de interferómetros por división de amplitud. En ambos casos hay interferencia cuando se combinan de nuevo las dos perturbaciones luminosas.

En la figura 11.5 se ve una foto del montaje en un banco óptico del interferómetro de Michelson. La luz proveniente de la fuente monocromática F incide en el divisor de haz DH, que es una placa de vidrio con un delgado recubrimiento de plata. Parte de la luz atraviesa DH y se refleja en el espejo E1 y la parte restante se refleja en la superficie plateada y se dirige hacia el espejo E2. La parte que se refleja en el espejo E1 regresa a través de DH y se refleja en la superficie plateada de DH hacia el espacio donde se coloca la pantalla. La parte que se refleja en el espejo E2 regresa, atraviesa

DH y se superpone con la otra parte en el espacio donde se coloca la pantalla. En clase vimos que las perturbaciones que llegan a esta zona parecen provenir de dos fuentes secundarias (virtuales) que están detrás de E2, en el eje que une E2 con la pantalla de observación. Estas fuentes secundarias definen el eje del sistema de hiperboloides, así que si está todo bien alineado (es decir, E1 y E2 a 90° entre sí y a 45° de la cara plateada de DH) en la pantalla se observan circunferencias concéntricas, tal como se muestra en el video <http://bit.ly/1k7XEho>.

El aparato de la figura está montado en su totalidad sobre una estructura antivibratoria (las mesas ópticas modernas lo son, Michelson llegó a usar mesas flotantes en mercurio porque los carruajes que pasaban por la calle hacían mover las franjas). La posición del espejo E2 se puede ajustar con un tornillo micrométrico muy fino y preciso (se observa en la foto). Cuando se desplaza el espejo E2 hacia atrás o hacia adelante una distancia $\lambda/2$, la diferencia de caminos entre las perturbaciones que pasaron por E1 y las que pasaron por E2 cambia en λ y cada franja se desplaza a la izquierda o a la derecha una distancia igual a la separación entre franjas, como puede verse en el video <http://bit.ly/1k7XEho>. Si se mueve el espejo una distancia d tal que la franja de orden $M \pm m$ esté en el lugar en que antes estaba la franja de orden M , entonces

$$d = m \frac{\lambda}{2}, \quad (11.14)$$

Si m es de varios miles, la distancia d es suficientemente grande como para medirla con buena exactitud y así se obtiene un valor muy exacto para la longitud de onda. Por otra parte, si λ ya se conoce, el método sirve para medir distancias del orden de una longitud de onda de la luz empleada con relativa facilidad.

11.7. Interferencia en láminas

También vimos en clase cómo analizar la interferencia producida con láminas y una fuente puntual, de nuevo todo se reducía al análisis anterior, o sea, a descubrir dónde están las fuentes secundarias S_1 y S_2 y tener en cuenta los desfases adicionales que se producen durante la reflexión. Si la onda incide desde un medio menos denso hacia otro más denso, la onda reflejada se desfasa en π , en cambio si la onda incide desde un medio más denso hacia otro menos denso, la onda reflejada no se desfasa. Las ondas transmitidas no se desfasan en ningún caso. En el caso de observar por reflexión, la lámina cumple la función de dividir la amplitud de la onda incidente de manera que se puede imaginar que las perturbaciones en el semiespacio superior provienen de dos fuentes coherentes virtuales colocadas detrás de la lámina, que son las dos imágenes de S formadas por: i) reflexión en la primera cara y ii) por transmisión en la primera cara, reflexión en la segunda cara y de nuevo transmisión en la primera cara.

11.8. Localización de franjas

Por todo lo visto anteriormente, es claro que cuando en la región donde coinciden las perturbaciones producidas por dos fuentes puntuales coherentes (por lo general, las fuentes secundarias creadas mediante un interferómetro), el sistema de franjas de interferencia producido siempre está asociado con los hiperboloides de revolución o con parte de estos hiperboloides. Estos hiperboloides llenan todo el espacio donde hay superposición, por eso se dice que el sistema de franjas es **no localizado**, o también “localizado en todo el espacio”. “Con fuente puntual, franjas no localizadas” sería la regla. Una fuente extensa es una colección de fuentes puntuales. Pero como por lo general estas fuentes puntuales no están sincronizadas, entre ellas *no interfieren*, es decir que cada fuente puntual de una fuente extensa es incoherente con respecto a las otras. Y haciendo las reservas del caso sobre todo lo que está detrás de la expresiones *no interfieren* y *son incoherentes*.

Si la fuente extensa se usa en un interferómetro, cada punto dará lugar a un par de fuentes secundarias con su sistema de hiperboloides. Pero a menos que los máximos y mínimos de todos los hiperboloides coincidan, las franjas asociadas con determinado punto de la fuente extensa van a borrar el sistema de franjas asociado con los puntos vecinos. Por este motivo, cuando hay fuente extensa, las franjas, si se ven, están **localizadas**: sólo se ven en aquellos lugares donde coinciden los máximos y mínimos de todos los sistemas de hiperboloides, independientemente del punto de la fuente extensa en el que se originaron. Todo problema de localización de franjas de interferencia se resuelve analizando si existe un lugar donde esto ocurre. Sólo se observarán franjas claras cuando las diferencias de fase entre pares de haces recibidos en un punto dado procedentes de partes diferentes del emisor sean iguales o casi iguales. En general, esta condición sólo se cumplirá cuando el punto de observación está en una cierta superficie, es decir, sólo se verán si el ojo del observador (o su instrumento de observación) se enfoca sobre dicha superficie.

Los ejemplos de localización más accesibles y fáciles de discutir en un curso introductorio de ondas son las franjas producidas en láminas o películas transparentes. En clase empezamos analizando las diferencias de camino y los desfases en reflexión necesarios para entender las **franjas de igual inclinación** que se observan con fuente puntual y una lente muy grande (texto asociado a la figura 9.17 del libro de Hecht). Una vez entendido este fenómeno, pasamos a ver lo que pasa cuando la lente (o el instrumento de observación) tiene pequeña abertura (texto asociado a figura 9.18 del libro de Hecht). Y así finalmente llegar al caso de fuente extensa (explicaciones de figuras 9.19, 9.20 y 9.21 del libro de Hecht), donde la localización se da en el infinito, porque es allí donde coincide la posición de los máximos y mínimos de todos los hiperboloides. También discutimos de manera muy general las llamadas franjas de igual espesor (explicaciones de figuras 9.22-9.26 del libro de Hecht).

Capítulo 12

Difracción

Si se acepta la validez del principio de Huygens, se debe reconocer que, cuando una onda interactúa con un obstáculo, la difracción siempre está presente. Sin embargo, y tal como vimos en el capítulo 9, la difracción aparece de manera significativa sólo cuando no se cumple la condición (9.1).

Recordemos que el principio de Huygens postula que todo punto de un frente de ondas primario actúa como fuente puntual que emite ondas esféricas secundarias de igual velocidad y frecuencia que la onda original. Y que el frente de ondas en todo instante es la envolvente de estas onditas esféricas (también llamadas onditas de Huygens) emitidas en un instante anterior. Por envolvente, Christiaan Huygens entendía *la superficie tangente*, un concepto fácil de visualizar geométricamente, pero quizás poco adecuado tanto para interpretar físicamente como para obtener resultados cuantitativos. A continuación, para poner en marcha el principio de Huygens y para cuantificar los apartamientos de la descripción geométrica, seguiremos la interpretación de Fresnel, el primero en imaginar que dicha envolvente no era otra cosa que la interferencia de todas las onditas secundarias (u onditas de Huygens). O sea: el mecanimso de Huygens según Fresnel consiste en sumar las onditas de muchas fuentes puntuales. Entonces, para poner en marcha el mecanimso de Huygens-Fresnel, y aprovechando lo aprendido en el capítulo anterior sobre la suma de ondas provenientes de dos fuentes puntuales (ver página 194), ahora generalizaremos el análisis para considerar la suma de muchas perturbaciones. Empecemos por el caso de N fuentes puntuales coherentes, es decir, de la misma frecuencia y que mantienen su fases iniciales constantes en el tiempo.

12.1. Interferencia de N fuentes

En la página 134 sintetizamos una función del tiempo a partir de N cosenos temporales con la misma amplitud (a), la misma fase y frecuencias equiespaciadas en el intervalo $\Delta\omega = \omega_2 - \omega_1$. El espaciado era $\delta\omega = \Delta\omega/(N-1)$. El resultado de esta síntesis se obtuvo con el procedimiento esquematizado en las ecuaciones (7.45), (7.47) y (7.48). Ahora queremos estudiar el efecto que producen las N perturbaciones armónicas que llegan desde las N fuentes a un punto P genérico. Si bien este problema parece com-

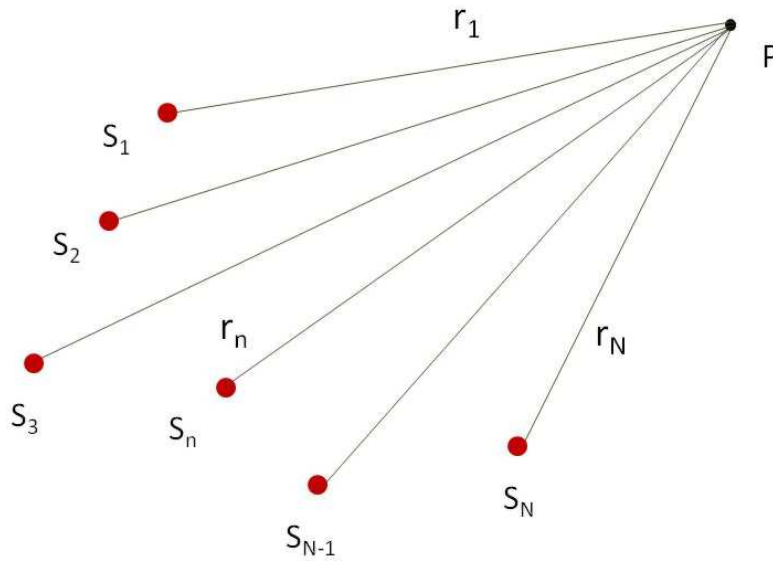


Figura 12.1: Las perturbaciones de N fuentes puntuales interfieren en el punto P .

pletamente diferente del problema de sintetizar una señal con N funciones armónicas de distintas frecuencias, veremos que ambos problemas tienen muchísimo en común.

12.1.1. Caso general

En la figura 12.1 se esquematiza una situación general. Notar que más allá de la alineación o el equiespaciado de las fuentes, la linealidad de la ecuación de ondas nos conduce a una suma como en la página 194, pero con N términos, en vez de solamente dos. La situación es completamente análoga: cada fuente produce una onda esférica progresiva y al punto P llegan N perturbaciones de la forma

$$\psi_n(r_n, t) = a_n \cos(\omega t - kr_n + \delta_n) = a_n \cos(\omega t + \Delta_n), \quad (12.1)$$

donde $\Delta_n = \delta_n - kr_n$, $k = \omega/v$, r_n se mide desde cada fuente al punto P , a_n depende de la intensidad de cada fuente y de r_n , mientras que δ_n es la fase inicial que se introduce para tener en cuenta que las fuentes se pueden encender en tiempos distintos y, en consecuencia, que pueden estar desfasadas. La perturbación total en el punto P es

$$\psi_P(t) = \sum_{n=1}^N a_n \cos(\omega t + \Delta_n), \quad (12.2)$$

y esta suma tiene que tener la forma

$$\psi_P(t) = A \cos(\omega t + \text{fase}). \quad (12.3)$$

Tenemos que encontrar una expresión para la amplitud A de la onda resultante en función de amplitudes a_n y fases iniciales Δ_n ($n = 1, 2, \dots, N$) de las ondas individuales

y la expresión buscada se tiene que reducir a (11.2) en el caso $N = 2$. Como en el caso $N = 2$, la fase de la resultante no es interesante en experimentos donde solamente se mida la intensidad de la perturbación.

Tal como hicimos en 7.3.3 pasamos a notación compleja y sobreentendiendo “parte real” (que tomaremos al finalizar los cálculos), podemos escribir

$$\psi_P(t) = \sum_{n=1}^N a_n e^{i(\omega t + \Delta_n)} = e^{i\omega t} \sum_{n=1}^N a_n e^{i\Delta_n}. \quad (12.4)$$

Comparando la expresión (12.4) con la ec. (7.47) obtenida en la sección 7.3.3, notamos un gran parecido entre ambas, excepto por: i) la presencia de las amplitudes a_n dentro de la sumatoria en (12.4) (que para el caso considerado en (7.47) son iguales y salen de factor común) y ii) el argumento de las exponenciales complejas, que en (12.4) es $\Delta_n = \delta_n - kr_n$, un valor bastante arbitrario porque aún no se ha dicho nada de las fases iniciales δ_n ni de cuál es la dependencia con n de las distancias r_n , (mientras que los δ_n para la (7.47) son proporcionales a n). Con esto queda bastante en evidencia lo que ya se había anunciado, que el problema de hallar la perturbación resultante de sumar N perturbaciones armónicas provenientes de N fuentes tiene un gran parecido formal con el problema de sintetizar una señal con N funciones armónicas de distintas frecuencias. La identidad formal de los dos problemas mencionados aparece justamente cuando se agregan hipótesis adicionales que permiten eliminar las diferencias mencionadas en i) y ii).

12.1.2. Fuentes alineadas y equiespaciadas

La dependencia de a_n con r_n se puede despreciar si las fuentes están localizadas en una región limitada del espacio y realizamos la medición en un punto P muy lejos de todas las fuentes. Para poner un caso concreto, consideremos que las fuentes están alineadas, como se muestra en la figura 12.2, y que están todas en fase (δ_n independiente de n , valor que podemos tomar cero con una elección conveniente del origen de tiempos y entonces $\Delta_n = -kr_n$). Con estas hipótesis, la suma de amplitudes complejas de la ec. (12.4) queda

$$\sum_{n=1}^N a_n e^{i\Delta_n} = \sum_{n=1}^N a_n e^{-ikr_n} = e^{-ikr_1} \sum_{n=1}^N a_n e^{-ik(r_n - r_1)} \quad (12.5)$$

donde el último paso equivale a tomar r_1 como referencia, sacar factor común fuera de la sumatoria y medir todos los caminos en las exponenciales con respecto a r_1 .

Para evaluar las diferencias $r_n - r_1$, consideremos que el plano del dibujo es el plano $x = 0$, que las fuentes están ubicadas a lo largo del eje y y en las posiciones y_n y que el plano de observación X-Y dista una distancia Z del plano x-y de las fuentes. Entonces

$$r_n = \sqrt{Z^2 + (Y - y_n)^2} \approx Z + \frac{(Y - y_n)^2}{2Z}$$

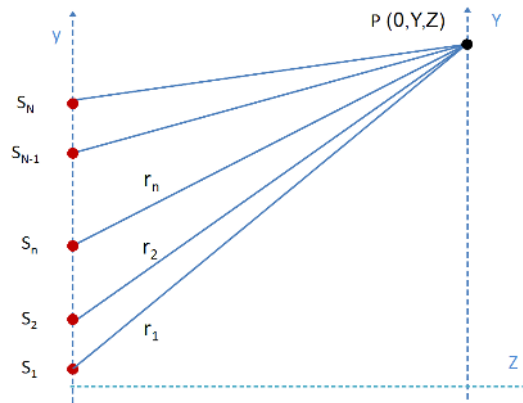


Figura 12.2: Las perturbaciones de N fuentes puntuales colineales (no necesariamente equiespaciadas) interfieren en un punto P alejado de la región donde están localizadas las fuentes.

$$r_1 = \sqrt{Z^2 + (Y - y_1)^2} \approx Z + \frac{(Y - y_1)^2}{2Z}$$

y restando y quedándonos con el primer orden en Y/Z , nos queda

$$r_n - r_1 \approx (y_1 - y_n) \frac{Y}{Z} \quad (12.6)$$

que coincide con la diferencia de camino obtenida en la aproximación paraxial para el experimento de Young (la distancia entre fuentes por el seno, o la tangente, de un ángulo). Esta expresión también muestra que una variación de camino óptico de media longitud de onda importa mucho en la fase, porque está pesada por k y entonces las perturbaciones individuales “tiran” juntas (si los caminos son iguales) o “tiran” en direcciones opuestas (si los caminos difieren en media longitud de onda). En cambio, esa misma variación de camino no modifica demasiado las amplitudes, porque hace que una amplitud pase de un valor inversamente proporcional a $1/r_1$ a otro valor inversamente proporcional a $1/(r_1 + \lambda/2)$, ambos muy parecidos entre sí cuando r_1 es mucho mayor que la longitud de onda λ . Como valores típicos que ilustran esta aproximación para luz visible, podríamos tomar $r_1 \approx 1m$ y $\lambda \approx 0.5 \cdot 10^{-6}m$.

La ec. (12.6) muestra que las fases de las exponenciales dentro de la sumatoria que representa la amplitud compleja resultante (ver ec. (12.5)) es siempre proporcional a la distancia entre la fuente 1 y la n . Entonces, si las fuentes están equiespaciadas, separadas todas la misma distancia d , las fases están equiespaciadas (como en el problema de suma de frecuencias), porque los caminos entre las perturbaciones provenientes de dos fuentes consecutivas cualesquiera, siempre son aproximadamente iguales a $\Delta = d \sin \theta$ (Figura 12.3), con θ la coordenada angular del punto lejano P medida desde la normal a la línea que pasa por todas las fuentes (eje Z).

El caso límite en que el punto P está en el infinito, equivale a observar en el plano focal de una lente. Y en este caso, a la interferencia en P solamente contribuyen aquellas perturbaciones asociadas a la dirección de propagación (a los rayos diríamos en la DG) con inclinación θ .

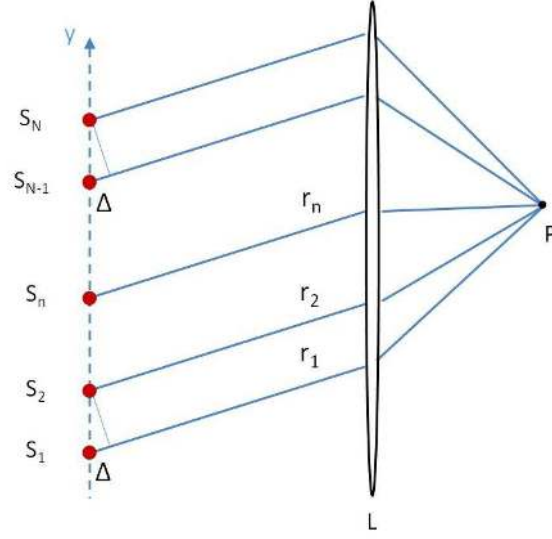


Figura 12.3: Las perturbaciones de N fuentes puntuales colineales equiespaciadas interfieren en el punto P en el plano focal de la lente L.

En resumen, en la configuración de la Figura 12.3, los caminos al punto de observación desde dos fuentes consecutivas difieren en el valor $\Delta = d \sin \theta$ y esta diferencia es independiente de la etiqueta que numera a las fuentes. De esta manera, $r_2 = r_1 - d \sin \theta$, $r_3 = r_1 - 2d \sin \theta$, $\dots$, $r_n = r_1 - (n-1)d \sin \theta$. Luego, la expresión (12.4) para $\psi_P(t)$ se escribe

$$\psi_P(t) = a e^{i\omega t} \sum_{n=1}^N e^{i\Delta_n} = e^{i\omega t} e^{-ikr_1} a \sum_{n=1}^N e^{ik(n-1)d \sin \theta}, \quad (12.7)$$

que en la representación compleja es una oscilación armónica con amplitud compleja igual al factor que multiplica a $e^{i\omega t}$. Este factor contiene una sumatoria entre 1 y N que mediante un cambio de índice se puede escribir como una sumatoria entre 0 y N-1 con la misma forma que la ec. (7.47) de la sección 7.3.3

$$\sum_{n=0}^{N-1} [e^{ikd \sin \theta}]^n, \quad (12.8)$$

donde $kd \sin \theta$ juega el mismo papel que $\delta\omega t$ en (7.47). De esta manera, es claro que la sumatoria se puede escribir como

$$\frac{e^{iNkd \sin \theta} - 1}{e^{ikd \sin \theta} - 1} = e^{i[(N-1)kd \sin \theta]/2} \frac{\sin(\frac{N}{2}kd \sin \theta)}{\sin(\frac{1}{2}kd \sin \theta)}. \quad (12.9)$$

Los términos e^{-ikr_1} en (12.7) y $e^{i[(N-1)kd \sin \theta]/2}$ en (12.9) tienen módulo unitario y por eso contribuyen solamente a la fase de la amplitud compleja de $\psi_P(t)$. Luego, la verdadera amplitud A_P (real) de la oscilación armónica resultante en P es

$$A_P = a \frac{\sin(\frac{N}{2}kd \sin \theta)}{\sin(\frac{1}{2}kd \sin \theta)} = a \frac{\sin N\alpha}{\sin \alpha}, \quad (12.10)$$

donde por brevedad definimos

$$\alpha = \frac{k}{2} d \sin \theta. \quad (12.11)$$

La intensidad I_P de la perturbación armónica resultante en P, proporcional al cuadrado de la amplitud A_P , queda

$$I_P = I \frac{\sin^2 N \alpha}{\sin^2 \alpha}, \quad (12.12)$$

donde I , proporcional al cuadrado de la amplitud a , representa la intensidad en P cuando está encendida una sola de las fuentes individuales.

Del resultado (11.5) obtenido en el caso $N = 2$ sabemos que cuando las fuentes interfieren constructivamente, la amplitud resultante es la suma de las amplitudes individuales ($2a$) y así la intensidad resultante es $4 = N^2$ veces la intensidad que se obtendría con una única fuente. Y antes de seguir con las cuentas, podemos convencernos que, bajo las hipótesis que hemos explicitado, tiene que pasar también algo similar en el caso $N > 2$. Porque si todas las perturbaciones están en fase (si todas “tiran” para el mismo lado), la amplitud de la perturbación armónica resultante en P debe ser N veces la amplitud de una perturbación individual (Na) y así la intensidad resultante debe ser N^2 veces la intensidad de una única fuente ($N^2 I$). Veamos a continuación que la (12.12) predice justamente este resultado.

12.1.3. Intensidad en la zona lejana

La ec. (12.12) describe la variación de la intensidad I_P en términos de la variable α , que contiene a las variables físicas $\lambda = 2\pi/k(\omega)$, d , N y θ .

Con numerador y denominador acotados, los máximos de I_P ocurren cuando se hace cero el denominador, es decir cuando

$$\alpha = m \pi, \quad \text{o sea} \quad \Delta = d \sin \theta = m \lambda, \quad (12.13)$$

con m un número entero. Esta condición confirma que la interferencia constructiva (máxima intensidad posible) aparece cuando todos los caminos desde las fuentes al punto de observación difieren en múltiplos de una longitud de onda. Cuando se cumple la condición (12.13) también se anula el numerador de (12.12) y al resolver la indeterminación se obtiene que la máxima intensidad es $N^2 I$, un resultado que también habíamos anticipado.

En cambio, la interferencia destructiva (mínima intensidad posible) ocurre cuando se anula el numerador de (12.12), pero no su denominador, es decir cuando

$$\sin N \alpha = 0 \quad \text{y} \quad \sin \alpha \neq 0 \quad (12.14)$$

o sea

$$N \alpha = j \pi, \quad \text{pero} \quad \alpha \neq l \pi \quad (12.15)$$

con m y l enteros, que en términos de θ se escribe

$$N \Delta = N d \sin \theta = j \lambda, \quad (12.16)$$

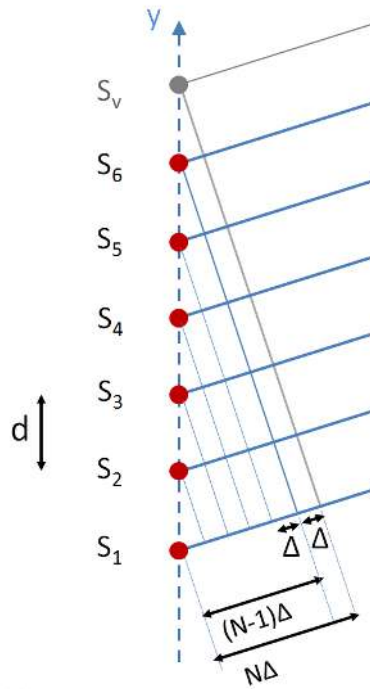


Figura 12.4: Cuando se cumple la condición (12.16) para $j = 1$, la primera fuente ($n = 1$) estaría en fase con una supuesta fuente virtual ($n = N + 1 = 7$) alineada y con el mismo espaciamiento, es decir que estaría en contrafase con la fuente en el centro ($n = 4$) del arreglo de $N + 1$ fuentes. Y lo mismo pasaría para las fuentes con $n = 2$ y $n = 5$ y así sucesivamente.

con j un entero distinto de cero y que no sea múltiplo de N , para que no se anule el denominador de la ec. (12.12). Como indica claramente la ec. (11.6) para en el caso $N = 2$, notemos al pasar que el hecho de que en este caso sea cero la mínima intensidad posible es una consecuencia de la suposición que se hizo al comienzo de la subsección 12.1.2 sobre la igualdad de las amplitudes de todas las perturbaciones en P.

Es instructivo interpretar físicamente la condición (12.16). Con este fin consideremos el esquema de la Figura 12.4 y observemos que según la ec. (12.7), $(N - 1) \Delta$ es la diferencia de camino entre las perturbaciones que llegan a P desde la primera y la última fuente. Es decir que el lado izquierdo de la condición (12.16) sería la diferencia de camino entre las perturbaciones que llegan a P desde la primera fuente y una eventual fuente virtual alineada con las fuentes reales y separada la misma distancia d de la última fuente. Y esto es equivalente a decir que la primera fuente estaría en contrafase con la fuente en el centro ($n = 4$) del arreglo de $N + 1$ fuentes y que toda fuente en la primera mitad del arreglo estaría en contrafase con otra fuente en la segunda mitad del arreglo. Un argumento similar vale para otros valores de j . En general, siempre que se cumpla la condición (12.16), es posible formar pares en contrafase y estos pares, que claramente interfieren destructivamente, producen ceros de intensidad.

También es instructivo interpretar las condiciones de máximos y mínimos de intensidad en el marco de la suma de números complejos $a_n e^{i\Delta_n}$ que aparece del lado derecho de la eq. (12.4). En el plano complejo estos números son vectores de módulo a_n e inclinación Δ_n con respecto al eje real. Si todas las inclinaciones Δ_n son iguales

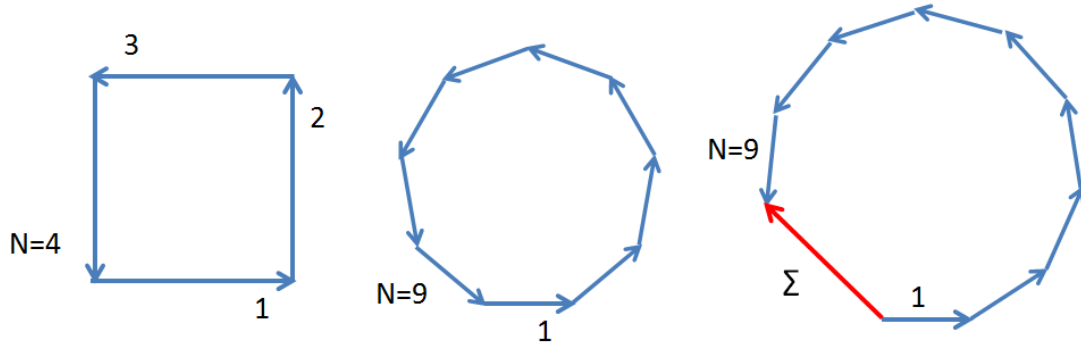


Figura 12.5: La suma de amplitudes complejas $a_n e^{i\Delta_n}$ de la eq. (12.4) como suma de vectores en el plano complejo cuando todos los vectores tienen el mismo módulo y los valores Δ_n difieren en el mismo valor 2α . Cuando el valor 2α está dado por la condición de mínimo (12.15), la resultante es cero, como se ilustra para el caso $j = 1$, con $N = 4$ (izquierda) y $N = 9$ (centro). A la derecha se muestra la resultante (vector rojo) para $N = 9$ y un valor 2α que no satisface ni la condición de mínimo (12.15) ni la de máximo (12.13) ($2\alpha = 33^\circ$).

o difieren en múltiplos de 2π , la resultante de la sumatoria es un vector de amplitud igual a la suma de las amplitudes, aunque sean distintas (interferencia constructiva, condición (12.13)). En cambio, si se tienen pares de vectores de igual módulo y orientaciones opuestas (con Δ_n 's que difieren en π), la resultante de la sumatoria es un vector de amplitud nula (caso de interferencia destructiva, del que la condición (12.13) es un caso particular cuando todas las amplitudes son iguales y N es par). Si todas las amplitudes son iguales y la inclinación relativa entre dos vectores consecutivos se mantiene constante, situación que es equivalente a la considerada en la ec. (12.7) (Δ independiente de n), es fácil ver que siempre que Δ satisfaga la condición (12.16), los vectores puestos uno a continuación del otro determinan en el plano complejo un polígono regular donde la punta del último vector coincide con el comienzo del primero, *cualquiera sea el valor de N (par o impar)*, o en otras palabras, que la resultante de la suma de vectores es nula. Esto es así porque según la ec. (12.11), 2α es la inclinación relativa entre vectores complejos asociados con fuentes consecutivas. Y según (12.15), en la condición de mínimo esta inclinación vale $j 2\pi/N$ (j distinto de cero y no múltiplo de N), es decir, justo la inclinación necesaria como para que la flecha del último vector toque el comienzo del primero, tal como se muestra en la figura 12.5 para $j = 1$ y en los casos $N = 4$ y $N = 8$.

En resumen, la ec. (12.12) muestra que la intensidad I_P en el plano focal de la lente tiene máximos cuando $\alpha = \dots - 2\pi, -\pi, 0, \pi, 2\pi \dots$ y la ec. (12.15) muestra que entre dos máximos consecutivos I_P se hace cero $N - 1$ veces, un resultado que está de acuerdo con lo que habíamos obtenido para el inteferómetro de Young ($N = 2$), donde entre dos máximos de interferencia constructiva encontramos solamente un lugar donde la intensidad se anula. Por ejemplo entre el máximo en $\alpha = 0$ y el máximo en $\alpha = \pi$, la intensidad vale cero cuando j en (12.15) vale $1, 2 \dots N - 1$, es decir, cuando $\alpha = \pi/N, 2\pi/N \dots (N - 1)\pi/N$, o sea cuando $d \sin \theta = \lambda/N, 2\lambda/N \dots (N - 1)\lambda/N$. Lo mismo ocurre entre el máximo en $\alpha = -\pi$ y el máximo en $\alpha = 0$, donde la intensidad vale cero $N - 1$ veces cuando j en (12.15) vale $-(N - 1), -(N - 2), \dots, -2, -1$, es

decir, cuando $\alpha = -(N-1)\pi/N, -(N-2)\pi/N, \dots, -2\pi/N, -\pi/N$ y mismo ocurre entre cualquier otro par de máximos con valores consecutivos de m en la condición (12.13).

La dependencia de la intensidad I_P con la variable α está ilustrada en la figura 12.6 para el caso $N = 5$, $I = 1$ y cuatro valores del cociente $d/\lambda = 1.2, 1.5, 2$ y 4 . Este cociente es el parámetro adimensional característico del fenómeno de interferencia que estamos estudiando. El resultado de la interferencia en este problema no depende de los valores de λ o d por separado ni tampoco de la naturaleza del fenómeno. Que los ejemplos de interferencia se hayan discutido mayormente en el contexto de las ondas luminosas no debería hacernos olvidar que, con las mismas hipótesis que hemos venido usando, se obtienen los mismos efectos de interferencia con otras ondas, como sonido audible, ultrasonido, ondas en la superficie libre de un fluido (olas) o con cualquier otro tipo de onda.

Como la intensidad producida en P por una única fuente se tomó arbitrariamente igual a uno, todos los máximos principales de la figura 12.6 alcanzan el mismo valor ($N^2 = 25$). Estos máximos son efecto de la interferencia constructiva, tienen un origen físico completamente análogo a los llamados órdenes del experimento de Young y cada uno tiene asociado un valor distinto del entero m de la ec. (12.13). Igual que en el experimento de Young, se los denomina *órdenes de interferencia*.

Notar que los órdenes de interferencia están equiespaciados en α (y en $\sin \theta$), pero no están equiespaciados en el plano focal de la lente. Es decir, no están equiespaciados en θ ni en las distancias medidas en el plano focal. Una vez fijado el valor de d , el equiespaciado en las variables α y $\sin \theta$ depende del valor de la longitud de onda: a menor longitud de onda (mayor valor de d/λ), menor es el equiespaciado entre órdenes. Y debido al hecho físico de que solamente tenemos acceso a ángulos θ reales entre $-\pi/2$ y $\pi/2$, la variación de $\sin \theta$ está acotada como así también los valores permitidos del entero m

$$-\frac{d}{\lambda} \leq m \leq \frac{d}{\lambda} \quad (12.17)$$

Vemos que el número total de órdenes resulta dos veces la parte entera del parámetro adimensional característico d/λ , y es menor cuanto mayor es la longitud de onda. Físicamente, la explicación es que si la longitud de onda es muy grande comparada con d , por más que variemos la coordenada angular que mide la inclinación θ del punto lejano P (ver figura 12.3), los caminos ópticos provenientes de cada fuente en el punto P siempre serán muy parecidos, sus variaciones serán siempre muy pequeñas comparadas con λ y así las variaciones de intensidad resultarán casi inapreciables. En cambio, si la longitud de onda es muy pequeña comparada con d , una pequeña variación de la coordenada angular θ puede producir variaciones de los caminos ópticos del orden de varias longitudes de onda, una condición necesaria para que al mover el detector en el plano focal de la lente, la intensidad pueda variar un número apreciable de veces entre sus valores máximo y mínimo y así tener un número apreciable de franjas. Tanto la discusión anterior como la expresión (12.17) muestran que N fuentes puntuales en fase y “muy juntas” ($d/\lambda < 1$) se comportan en la zona lejana casi como una única fuente puntual, en el sentido que no producen órdenes de interferencia ya que no existen variaciones apreciables de camino óptico en puntos de observación lejanos.

Estas consideraciones pueden usarse para establecer un criterio que involucre d , λ y la posición del punto de observación y que sirva para establecer cuándo una fuente extensa puede ser considerada puntual.

En la figura 12.7 exploramos cómo cambian las curvas de I_P predichas por la ec. (12.12) cuando se aumenta el número de fuentes. El valor $d/\lambda = 1.2$ se mantiene constante y se muestran los casos $N = 3, 10, 20$ y 50 . Como la intensidad $I = 1$ de las fuentes individuales se deja constante para todas las curvas, la intensidad máxima en P aumenta con el cuadrado del número de fuentes. Se observa que a medida que aumenta N :

- i) se hace mas pequeña la distancia angular entre cada orden de interferencia y su dos mínimos adyacentes a izquierda y derecha, tal como lo predice la ec. (12.15). Esto equivale a un afinamiento muy notable de los picos de intensidad correspondientes a los órdenes de interferencia, que pasan a ocupar regiones cada vez limitadas del espacio y a estar cada vez más definidas. En particular, notar que para $N = 2$ podríamos decir que las colitas de los órdenes ± 1 se van más allá del horizonte, que puede sonar casi esotérico pero no.
- ii) disminuye la intensidad relativa de los $N - 2$ máximos secundarios que hay entre dos órdenes consecutivos. Si bien nunca calculamos explícitamente la posición angular ni la intensidad de estos máximos secundarios, al hacer la cuenta se obtiene que no todos tienen la misma intensidad y que están aproximadamente a mitad de camino entre dos ceros consecutivos, como se puede comprobar muy fácilmente con un sencillo código Python como el de la Figura 12.8.
- iii) la función $I_P(\alpha)$ tiende a un peine de deltas, los dientes son los órdenes, cada vez mas delgados y el fondo de máximos secundarios se hace cada vez menos significativo. Excepto por el distinto significado de las variables físicas involucradas, este peine es completamente análogo al esquematizado en la segunda fila de la Figura 7.1.

12.1.4. Intensidad en la zona cercana

La figura 12.9 se corresponde con el esquema de la figura 12.3, pero con la diferencia de que el punto de observación, en vez de estar muy alejado de las N fuentes colineales, es un punto **“cercano”**. Entonces, ni las amplitudes ni las fases relativas de las perturbaciones individuales satisfacen las hipótesis que usamos para obtener la expresión (12.10). Por este motivo, no nos queda otro remedio que volver al punto de partida, que es el principio de superposición expresado en la suma de la ec. (12.2).

Este problema es muy importante, no solamente porque da lugar a muchísimas aplicaciones tecnológicas con fenómenos ondulatorios sino también porque la envolvente de las onditas secundarias postuladas por Huygens no es otra cosa que la *interferencia* de muchas onditas secundarias, tal como propuso Fresnel. El mensaje es el siguiente: si estamos familiarizados con expresiones del tipo (12.2), seremos capaces de entender

mejor el principio de Huygens y ponerlo en marcha para obtener una herramienta matemática que permita predecir resultados cuantitativos mucho más específicos que los que se obtienen solamente con el concepto geométrico de “envolvente”.

Es conveniente mencionar que el interés en la interferencia con muchas fuentes (como así también el interés en la difracción y en los apartamientos de la descripción geométrica) no está limitado a problemas donde el punto de observación está lejos de las fuentes. Si bien el estudio sistemático para puntos cercanos está fuera del alcance de un curso introductorio de ondas, es muy instructivo evaluar la suma involucrada en la expresión (12.4) empleando el método FB (de fuerza bruta), que bien usado y con la ayuda de una computadora o cualquier dispositivo con microprocesador programable (computadora de escritorio, notebook, tablet, celular) es hoy en día una herramienta tan válida e imprescindible como lo fue la regla de cálculo o las tablas de funciones especiales en otras épocas.

Consideremos que las N fuentes puntuales están ubicadas a lo largo del eje y con coordenadas cartesianas $x_n = 0$, $y_n = (n - 1)d$, $z_n = 0$ y que situamos el punto P “cercano” en el plano $x = 0$ que contiene a las fuentes, con coordenadas cartesianas $(0, Y, Z)$. Las fases Δ_n que intervienen en (12.4) siguen siendo $\Delta_n = \delta_n - kr_n$ y si mantenemos la hipótesis de que las fuentes están todas en fase, $\delta_n = 0$. Pero ahora dos valores de r_n consecutivos no difieren en un valor constante, como pasaba en el caso de puntos lejanos (recordar figura 12.3). Esto es así porque la nueva expresión para r_n es

$$r_n = \sqrt{Z^2 + (Y - y_n)^2}, \quad (12.18)$$

en vez de la dada unas líneas antes de la ec. 12.7. Luego, la sumatoria en la ec. (12.4) queda

$$\sum_{n=1}^N a_n e^{i\Delta_n} = \sum_{n=1}^N a_n e^{ikr_n} = e^{ikr_1} \sum_{n=1}^N a_n e^{-ik(r_1 - r_n)} \quad (12.19)$$

donde el argumento de la exponencial en la última sumatoria no depende linealmente de la etiqueta que identifica a las fuentes, como ocurría en el caso considerado en la sección 12.1.3. Si bien el cambio impide encontrar una expresión cerrada para la amplitud de la perturbación resultante en P, todo físico que se precie debe ser capaz de analizar el problema con herramientas computacionales.

12.1.5. Experimentos numéricos

La solución numérica de problemas físicos sin solución analítica explícita fue históricamente la primera aplicación científica de las primeras computadoras. Hoy en día constituye un área multidisciplinaria con identidad propia donde no solo se superponen física, matemáticas y computación (tres carreras de nuestra Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires) sino que también actúa de puente entre dichas disciplinas.

Por otro lado, desarrollar una herramienta numérica para resolver un problema físico y asegurarse de que la herramienta está funcionando correctamente es un excelente

ejercicio para poner a prueba nuestra comprensión de muchos detalles que sin este ejercicio hubiera pasado desapercibidos.

A continuación concentrémonos en la simulación numérica del problema planteado en la eq. (12.19). Un brevísimo código Python podría ser el siguiente

```
#####
# Campo Cercano
# N fuentes id'enticas equiespaciadas
#
# Curso de F2, 1er cuatrimestre de 2017
# ver http://users.df.uba.ar/rdep/>
#####

import numpy as np
import matplotlib.pyplot as plt
import cmath; import math

#
# Datos entrada

N=3 # Nro. de fuentes
dl=1.6 # d/lambda
#
# delimita zona
xini=-1+1.*10.**(-5)
xfin=2.5 # en unidades de d
nrox=150
yini=-3.6+1.*10.**(-5) # en unidades de d
yfin=6.8 # en unidades de d
nroy=500

# amplitud fuentes
af=1.
pi=np.pi
ceror=0.; ceroi=0.; cero=cero+i*j*ceroi

f=open('NfuentesCercano.dat','w')
print >>f, "x x psi2 logpsi2"
print >>f, "Nro. fuentes = %g" %N
print >>f, "d/lambda = %g" %dl
print >>f, "xini = %g, xfin= %g, nrox= %g" %(xini, xfin, nrox)
print >>f, "yini = %g, yfin= %g, nroy= %g" %(yini, yfin, nroy)

print >>f, " "
print >>f, "x y psi2 logpsi2"

x1 = dl*np.linspace(xini, xfin, nrox)
y1 = dl*np.linspace(yini, yfin, nroy)

yf = dl*np.arange(N) # yj de fuentes en x=0
# medidas en lambdas

for ii in range(nrox):
    for jj in range(nroy):
        aux1=cero
        r1=(x1[ii]**2+y1[jj]**2)**0.5
        for jf in range(N):
            rj=(x1[ii]**2+(y1[jj]-yf[jf])**2)**0.5
            deltarj=rj-r1
            aux1=aux1+af*cmath.exp(1j*2*pi*deltarj)/rj
        psi2=abs(aux1**2)
        print >>f, "%g %g %g" %(x1[ii],y1[jj],psi2)
```

donde los datos de entrada incluyen el número de fuentes N , el cociente d/λ y las coordenadas del punto de observación. Para explorar cómo varía espacialmente la intensidad de la perturbación resultante, el punto P se mueve en una grilla rectangular determinada por las variables de entrada $xini$, $xfin$ y $nrox$ (grilla $x1$ de puntos en la dirección Z) e $yini$, $yfin$ y $nroy$ (grilla $y1$ de puntos en la dirección Y). Se eligió dar las longitudes medidas en unidades de d , construir las grillas en estas unidades y luego multiplicar por dl . Luego, se usan dos ciclos para recorrer cada punto de la grilla y un tercer ciclo anidado para hacer la sumatoria indicada en (12.19). Al terminar la suma se calcula el valor absoluto al cuadrado, se escriben las coordenadas y la intensidad en un archivo y luego se pueden graficar los datos de este archivo con el método preferido. El gráfico se podría haber hecho en el mismo script con comandos Python, pero quizás es preferible tener los datos en un archivo y luego jugar con comandos Python o usar un graficador externo, como Gnuplot.

Consideremos un sistema de 3 fuentes equiespaciadas y usemos el archivo de comandos mostrado arriba para explorar la distribución de intensidad en las distintas zonas. Los resultados se visualizan mejor en un mapa de colores con escala logarítmica para las intensidades, tal como se muestra en la figura 12.10 para la región indicada en el script de arriba, comprendida entre $0.12 \leq Z/\lambda \leq 4$ (muy cercana). Lo primero que nos llama la atención es que en esta zona la perturbación resultante tiene una distribución espacial muy irregular, nada parecida a la que anuncia la ec. (12.12). Mientras la ec. (12.12) para campo lejano predice diagramas de interferencia que solamente dependen de las coordenadas angulares, no radiales, del punto de observación, el diagrama de intensidad obtenido en la figura 12.10 varía violentamente en la dirección radial, con máximos de gran intensidad (puntos calientes, rojos) muy cerca de las fuentes (en $y_n/\lambda = 0, 1.6$ y 3.2), junto a otras zonas menos intensas (amarillas, casi 1 decibel menos) y otras zonas donde la intensidad sigue bajando un decibel (verde, verde azulado) hasta llegar al azul oscuro que es una zona con intensidad casi cien mil veces menor que las zonas de intensidad máxima (zonas frías en un horno de microondas, o silenciosas en el lenguaje acústico).

Para sistemas de fuentes con tamaño típico menor que media longitud de onda, la zona cercana suele ser muy chica, menor que una longitud de onda. En cambio, para sistemas de fuentes de tamaño típico mayor que media longitud de onda, como en este caso, donde el tamaño típico es $D = (N - 1)d = 3.2\lambda$, el límite entre zona cercana y lejana viene dado aproximadamente por la *distancia de Fraunhofer* D_F

$$\frac{D_F}{\lambda} = 2\left(\frac{D}{\lambda}\right)^2, \quad (12.20)$$

donde D es la mayor distancia característica del sistema de fuentes. Que el valor D_F/λ en este caso sea ≈ 20 está de acuerdo con el hecho de que ni en la figura 12.10 ni en la figura 12.11 (para una región ampliada, comprendida entre $1.6 \leq Z/\lambda \leq 11.2$) podemos ver claramente la estructura ilustrada en el primer panel de la figura 12.7.

Sin embargo, en las figuras 12.10 y 12.11, especialmente a partir de distancias de algunas longitudes de onda, ya se empieza a vislumbrar una direccionalidad y las ganas de formar tres órdenes de interferencia, tal como se espera para la zona lejana. Esta tendencia se acentúa cuando el punto de observación se sigue alejando de las fuentes, tal como se muestra en los dos paneles de la figura 12.12.

No se puede negar que gracias a la ayuda del corto y sencillo script de la página 216 hemos sido capaces de revelar interesantes comportamientos físicos contenidos, pero bastante ocultos, en la expresión (12.4), comportamientos que habría sido muy difícil obtener sin ayuda computacional. Por este motivo, podríamos considerar que la física computacional es una rama intermedia entre la física teórica y la física experimental, es decir, un camino que complementa la teoría y el experimento. Es como tener un laboratorio particular, donde tenemos que entender muy bien los mecanismos y leyes físicas del fenómeno en estudio para poder instruir correctamente a nuestro ayudante computacional y que así nos entregue resultados confiables de cierto experimento. Los resultados del experimento ya están volcados en las figuras, ahora “midamos” nuevos observables. Por ejemplo, midamos en el panel superior de la figura 12.12 el ángulo que forma el “haz” cuyo centro parece pasar por el punto de coordenadas $Y/\lambda =$

16.5 y $Z/\lambda = 19.2$. Si medimos “a ojo” (se podría mejorar inspeccionando la tabla con la que se hizo el gráfico) este ángulo desde el centro del sistema de fuentes, su tangente resulta igual al cociente entre 14.9 ($= 16.5 - 1.6$) y 19.2, valor que corresponde aproximadamente a un ángulo de 37.81° . De igual manera midamos también en el panel superior el ángulo que forma el “haz” cuyo centro parece pasar por el punto de coordenadas $Y/\lambda = -13.5$ y $Z/\lambda = 19.2$. Medido desde el centro del sistema de fuentes, su tangente resulta igual al cociente entre 15.1 ($= 13.5 + 1.6$) y 19.2, valor que ahora corresponde aproximadamente a un ángulo de -38.18° . Como medimos $Y/\lambda =$ “a ojo”, con un error de digamos media rayita, salvo el signo ambos valores son iguales dentro del error, una simetría esperada en los resultados de campo lejano. Más aún, la ec. (12.13) para campo lejano predice que los ángulos con la horizontal $\theta_{\pm 1}$ formados por los órdenes de interferencia ± 1 satisfagan $\sin \theta_{\pm 1} = \pm \lambda/d = \pm 1/1.6$, esto es, $\theta_{\pm 1} \approx \pm 38.68^\circ$. Notar que hay otro “haz” mas intenso cuyo centro pasa muy cerca del punto con coordenadas $Y/\lambda = 1.6$ y $Z/\lambda = 19.2$, o sea, que es casi paralelo a la línea horizontal que pasa por el centro del sistema de fuentes y entonces corresponde al orden de interferencia de la ec. (12.13) con $m = 0$. Las conclusiones son similares cuando se “miden” ángulos de haces en el panel inferior de la figura 12.12, donde además se puede apreciar un único máximo secundario entre dos órdenes de interferencia consecutivos, en coincidencia con la intensidad mostrada para el caso $N = 3$ en la figura reffig:IPNequi2.

Realmente es notable la coincidencia entre los valores del “experimento numérico” y los valores teóricos correspondientes al caso límite en que el punto de observación está infinitamente alejado de las fuentes, especialmente si tenemos en cuenta todas las aproximaciones que hicimos para deducir la ec. (12.12), incluida la presencia de una lente.

12.1.6. Límite continuo, $N \rightarrow \infty$, $d \rightarrow 0$

Estudiemos un caso muy similar al de la figura 12.3, con la diferencia de que ahora aumentamos indefinidamente el número N de fuentes ($N \rightarrow \infty$) mientras mantenemos constante la distancia $b = (N - 1)d$ entre la primera y la última fuente. Claramente las fuentes están más juntas y entonces $d \rightarrow 0$. Y para que la intensidad resultante en situaciones de interferencia constructiva, que ya sabemos que crece con N^2 , no diverja (y para que no salga demasiado caro pagar la energía consumida por las fuentes), a medida que aumentamos N vamos a ir disminuyendo la intensidad de cada fuente, de manera tal que la intensidad para interferencia constructiva sea independiente del número de fuentes. Es claro que esto significa que en la ec. (12.12) el límite que interesa es mantener constante el producto $N^2 I$, llamémoslo I_0 . Y como escribimos en el renglón siguiente a la ec. (12.12), esto equivale a que las amplitudes individuales de cada fuente sean inversamente proporcionales a N .

Con estas consideraciones, procedemos a tomar el límite en la ec. (12.12). Según (12.11), $\alpha \rightarrow 0$, pero no $N\alpha$, ya que $Nd \approx (N - 1)d = b$. Entonces, el argumento del seno en el numerador de la ec. (12.12) se mantiene acotado e igual a

$$\lim_{N \rightarrow \infty} N\alpha \rightarrow \beta = \frac{k}{2} b \sin \theta = \pi \frac{b}{\lambda} \sin \theta, \quad (12.21)$$

mientras que el argumento del seno en el denominador de la ec. (12.12) tiende a cero y se puede aproximar por su argumento. Y al reemplazar I por I_0/N^2 , en el denominador aparece la variable β^2 , y así resulta que la intensidad producida por una distribución continua de fuentes puntuales que ocupan un segmento de longitud b cuando el punto de observación está infinitamente lejos, viene dada por la expresión

$$I_P = I_0 \left(\frac{\sin \beta}{\beta} \right)^2, \quad (12.22)$$

donde aparece la ya conocida función seno cardinal. Esta función ya apareció en las ecuaciones (7.54) y (7.57), en un problema que parece completamente distinto. Vemos que el problema de sumar infinitos cosenos de la misma amplitud y la misma fase con frecuencias distribuidas (uniformemente porque las amplitudes son iguales) en el intervalo $\omega_1 \leq \omega \leq \omega_2$ da una amplitud moduladora que tiene la forma de la función $\sin u/u$, la misma dependencia que según (12.22) tiene ahora la amplitud de la perturbación que resulta de hacer interferir infinitas fuentes monocromáticas idénticas uniformemente distribuidas en el intervalo $0 \leq y \leq b$. La función seno cardinal, máxima para $\beta = 0$ y ceros adyacentes para $\beta = \pm \pi$ (es decir, $\sin \theta = \pm \lambda/b$) está representada en las figuras 7.2 y 7.3 para distintos valores de la diferencia de frecuencias $\Delta\nu$, la intensidad es el cuadrado de esta función. Notar que de acuerdo a lo discutido para un número finito de fuentes, 2β es la diferencia de fase asociada a los caminos provenientes de ambos extremos del sistema de fuentes. Luego, en los ceros adyacentes al máximo principal, $2\beta = \pm 2\pi$ y el camino desde el extremo superior del sistema de fuentes estaría en contrafase con el camino desde el centro del sistema y todo camino desde un punto en la mitad superior del sistema estaría en contrafase con otro camino desde un punto en la mitad inferior del sistema.

Si comparamos el argumento u del seno cardinal en el problema del capítulo 7, dado por la ec. (7.58), con el argumento β del seno cardinal en la ec (12.22), ver lado derecho de (12.21), vemos que $\Delta\nu$ se corresponde con b/λ , mientras que el tiempo t se corresponde con $\sin \theta$. Y de acuerdo con lo discutido en la ec. (7.59), la posición angular del primer cero puede ser usada como medida de la localización. Antes era localización temporal, ahora es localización espacial, es decir localización en la variable $\sin \theta$: cuánto se extienden en el plano focal de la lente las manchas de interferencia producidas por el sistema de fuentes. La localización temporal de la moduladora era inversamente proporcional al contenido frecuencial $\Delta\nu$ (cuanto más monocromática, menos localizada), la localización angular del diagrama de interferencia es inversamente proporcional al ancho b del sistema de fuentes (cuanto mayor es b , más se parece el sistema de fuentes a un sistema de onditas de Huygens sin limitación en la dirección y , es decir a un sistema de rayos paralelos no obstruidos, que según la DG debería dar en el foco de la lente solamente una mancha (imagen) puntual.

Es claro que con estas herramientas de interferencia podemos pasar a estudiar difracción, simplemente como una aplicación de la interferencia de las infinitas ondas secundarias de Huygens.

12.2. Difracción de Fresnel y de Fraunhofer

La difracción, o desviación de una onda de la propagación rectilínea predicha por la DG, ocurre siempre que la onda interactúa con un obstáculo y así la envolvente de las onditas de Huygens, que mantenía la forma del frente de ondas original antes de interactuar con el obstáculo, cambia su forma. Los frentes de onda planos dejan de ser planos y aparecen curvaturas como resultado de efectos de borde. Así los rayos perpendiculares a los frentes dejan de ser todos paralelos y los rayos divergen, o los frentes de onda esféricos, que seguían siendo esféricos, empiezan a cambiar de forma también como resultado de efectos de borde, así aparecen rayos que dejan de tener todos el mismo centro. Cuando la onda interactúa con objetos grandes comparados con la longitud de onda, es decir, cuando se cumple la relación (9.1), los apartamientos de la propagación rectilínea son tan pequeños que se pueden ignorar. En cambio, cuando la onda interactúa con objetos de tamaños comparables con la longitud de onda, hay que investigar la envolvente de Huygens, o lo que es equivalente, hay que calcular la interferencia que producen en cada punto de observación las partes no obstruidas del frente de ondas que se propagaron más allá del obstáculo. Como se ve, no hay una diferencia física significativa entre interferencia y difracción. Y lo que sigue tiene que resultar muy familiar, ya que estamos estudiando interferencia desde el capítulo anterior.

Por razones fenomenológicas y metodológicas, el estudio de la difracción se divide en dos partes que se atacan por separado. En la **difracción de Fraunhofer**, tanto la fuente como el punto de observación están muy lejos del obstáculo difractor, idealmente ambos en infinito. La situación es muy parecida a la que vimos en la sección 12.1.3, si la fuente está en infinito, sobre el objeto incide una onda plana y todas las fuentes de Huygens están entonces en fase, resultando además que las diferencias de camino desde las distintas fuentes hasta el punto de observación P dependen linealmente de las etiquetas o coordenadas que identifican a las fuentes. En este caso, al alejarnos del objeto, y tal como vimos en los paneles de la figura 12.12, los diagramas de difracción cambian de tamaño, pero no de forma. En cambio, en la **difracción de Fresnel** la fuente o el punto de observación están “cerca” del obstáculo difractor y la situación es muy parecida a la que vimos en la sección 12.1.4, donde las diferencias de camino desde las distintas fuentes hasta el punto de observación *no dependen linealmente* de las etiquetas o coordenadas que identifican a las fuentes, tal como es evidente a partir de la ec. (12.18). En estos casos la distribución de intensidad de la perturbación resultante puede variar muy bruscamente cuando se cambia la posición del punto de observación, tal como vimos en las figuras 12.10 y 12.11.

Si bien estamos en condiciones de explorar los efectos de la difracción de Fresnel, partiendo de expresiones como (12.2) y usando procedimientos numéricos como los vistos en la sección 12.1.4, parece adecuado encarar primero el caso asintótico de Fraunhofer y dejar el estudio de la difracción de Fresnel para cursos avanzados.

12.3. Difracción de Fraunhofer

Por sus múltiples aplicaciones en ciencia y tecnología y por ser menos demandante matemáticamente, nos limitaremos a estudiar la difracción de Fraunhofer en el caso de *obstáculos planos*. La situación se esquematiza en la figura 12.13. La onda incidente proviene de una fuente puntual que, o está muy lejos o se coloca en el foco objeto de una lente (no dibujada). De esta manera, el frente de ondas incidente es plano y si el plano del frente de ondas es además paralelo al plano $x - y$, que es el plano que contiene al obstáculo, entonces, todos los elementos de superficie dS asociados con partes no obstruidas del obstáculo, identificados por sus coordenadas genéricas (x, y) , son iluminados al mismo tiempo. Y así resulta que están en fase *todas las fuentes secundarias*, que según el principio de Huygens iradian hacia adelante para dar lugar a la perturbación en los puntos de observación $P(X, Y, Z)$ más allá del obstáculo.

Notar que la situación es completamente análoga a la considerada en la figura 12.3 y que nos había conducido a expresiones como (12.4), (12.7) y (12.8), con dos diferencias importantes:

- i) son nulos todos los δ_n que aparecen en el argumento $\Delta_n = \delta_n - kr_n$ de las exponenciales complejas de la eq. (12.4), o sea $\Delta_n = -kr_n$;
- ii) la suma de perturbaciones se hace sobre una colección de fuentes que no están necesariamente alineadas y equiespaciadas, sino que están distribuidas de manera continua en las partes no obstruidas del plano $x - y$ (las coordenadas x e y juegan el papel de la etiqueta n).

La figura 12.14 da algunos ejemplos de la diversidad contenida en la familia de obstáculos planos: una chapa con una o varias perforaciones de la misma o distinta forma (figura 12.14a), un interferómetro de Young con dos rendijas lineales del mismo ancho b separadas una distancia d (figura 12.14b) y un disco opaco o una abertura circular en una pantalla opaca (figura 12.14d, según la parte sombreada se considere opaca o transparente).

Volviendo al esquema de la figura 12.13, observamos que la perturbación en cada punto de observación $P(X, Y, Z)$ tendrá aportes de todas las fuentes elementales situadas en $(x, y, 0)$. Cuanto más lejos del obstáculo se ubique el punto de observación P , más paralelos serán los rayos asociados a las perturbaciones provenientes de las fuentes elementales. En el caso límite de llevar el punto de observación al infinito, o lo que es lo mismo, de poner la lente y observar en su plano focal, solamente llegan a P las perturbaciones asociadas a rayos estrictamente paralelos y existe una relación unívoca entre las coordenadas $X - Y$ del punto P y la inclinación de estos rayos paralelos provenientes de fuentes elementales no obstruidas, es decir, que a cada $P(X, Y)$ solo llegan aquellas perturbaciones provenientes de distintos elementos de área dS que tengan la inclinación adecuada para converger en P .

12.4. Rendija de ancho b

El caso más sencillo de obstáculo plano es el de una rendija lineal infinitamente larga orientada en la dirección x y de ancho b en la dirección y , que sería la idealización de una abertura rectangular de largo L , con $L \gg b$, para la que no esperamos difracción asociada con las limitaciones a lo largo de x , pero sí esperamos difracción asociada con el ancho b , si $b \approx \lambda$. En el marco del principio de Huygens, estamos en presencia de esencialmente el mismo problema que ya consideramos en 12.1.6, aunque dada la simetría de este problema, en vez de sumar perturbaciones provenientes de fuentes puntuales podríamos simplificar las cuentas sumando directamente perturbaciones provenientes de fuentes lineales, es decir dividir la rendija en tiras de largo L y ancho dy . Así, cada fuente elemental ubicada entre y e $y + dy$ emite una onda cilíndrica con amplitud proporcional al elemento de área $dS = Ldy$ e inversamente proporcional a la potencia $1/2$ de la distancia entre fuente y punto de observación (ver (7.31)), siempre y cuando $-b/2 \leq y \leq b/2$, es decir, las partes obstruidas no contribuyen a la perturbación en P (ver figura 12.14c). Salvo este detalle de la dependencia de las amplitudes con la distancia entre fuente y punto de observación, para puntos lejanos sigue siendo válido considerar que las amplitudes de las perturbaciones individuales son iguales y sacarlas el factor común a de la suma en (12.7). Con esta discusión, no queda más remedio que aceptar que una rendija de ancho b produce un diagrama de difracción de Fraunhofer con las mismas características que el diagrama de interferencia de un sistema continuo de fuentes uniformemente distribuido en un segmento de longitud b y que en consecuencia la intensidad en la pantalla de observación responde a la ec. (12.22).

Podemos calcular fácilmente la distribución de intensidad para valores distintos de b/λ adaptando levemente el script de la página 234. Los resultados obtenidos para I_P/I_0 se muestran en la figura 12.15 para $b/\lambda = 10, 5, 2$ y 1 . Además del máximo principal que corresponde a $\beta = 0$, hay máximos secundarios, aproximadamente a mitad de camino entre dos ceros consecutivos. Entonces, la intensidad relativa de los máximos secundarios cerca de $\beta \approx \pm 3\pi/2$ es $I/I_0 \approx 5\%$, la de los máximos secundarios cerca de $\beta \approx \pm 5\pi/2$ es $I/I_0 \approx 2\%$, etc. La intensidad de los máximos secundarios sigue cayendo rápidamente a medida que nos alejamos del centro de la pantalla y por este motivo normalmente no es fácil apreciarlos.

Notar que la posición angular del primer cero, $\sin \theta = \pm \lambda/b$, mide la localización en la variable espacial $\sin \theta$ y cuantifica la difracción en el plano focal de la lente. Una onda plana no obstruida debería focalizarse como una imagen puntual en el foco de la lente, pero en vez de una imagen puntual se tiene una mancha de semiancho λ/b en la variable $\sin \theta$, que es muy parecido al semiancho angular si $\theta \ll 1$, como es razonable que pase en todo sistema óptico donde la difracción se considere un fenómeno indeseable (por ejemplo, en microscopios y telescopios).

12.5. Obstáculos planos con simetría de translación

Es instructivo volver a obtener los resultados de la sección anterior sin necesidad de pasar por el caso discreto, es decir, directamente a partir del principio de Huygens. Cuando el plano del frente de ondas incidente es paralelo al plano $x - y$, las fuentes secundarias asociadas con partes no obstruidas en dS y etiquetadas por (x, y) irradian en fase y se superponen en el punto de observación $P(X, Y, Z)$.

Si el obstáculo tiene simetría de translación en una dirección, digamos x , conviene considerar que las fuentes secundarias son tiras de largo L y ancho dy y que cada una emite una onda cilíndrica con amplitud proporcional al elemento de área $dS = Ldy$ e inversamente proporcional a la potencia $1/2$ de la distancia entre fuente y punto de observación. Y si en el campo lejano despreciamos la variación de la amplitud con la distancia, la amplitud de la perturbación en el punto $P(X, Y, Z)$ proveniente de la fuentecita ubicada entre y e $y + dy$ resulta

$$\frac{\mathcal{E}_0}{\sqrt{R}} L dy, \quad (12.23)$$

donde escribimos R , el camino al punto P desde el origen del sistema de coordenadas $x - y$ que describe al obstáculo, en lugar de r , el camino al punto P desde la fuente elemental que estamos considerando, ver figura 12.14c. A esta altura tiene que quedar claro que la combinación (12.23) juega en la superposición de Huygens el mismo papel que juega la amplitud a en la ecuación (12.7). De la figura 12.14c sale que $r = R - y \sin \theta$ y entonces la expresión para la contribución elemental $d\psi_P(t)$ resulta

$$d\psi_P(t) = \frac{\mathcal{E}_0 L}{\sqrt{R}} e^{i\omega t} e^{-ikR} e^{iky \sin \theta} dy, \quad (12.24)$$

que juega ahora el mismo papel que antes jugaban los términos de la sumatoria en (12.7).

En total paralelismo con (12.7), para obtener la perturbación resultante en el punto P solamente falta sumar sobre todas las fuentes, que equivale a integrar sobre todas las regiones del plano $x - y$ no obstruidas, que para la simetría considerada quiere decir sumar sobre todas las tiras paralelas a x , es decir, integrar en la variable y . Tanto lío para decir *integrar en la variable y* , pero decirlo tan resumido, sin saber decir todo lo anterior, es hacer matemática, no física. De esta manera, tenemos que

$$\psi_P(t) = \frac{\mathcal{E}_0 L}{\sqrt{R}} e^{i(\omega t - kR)} \int_{\text{PNO}} e^{iky \sin \theta} dy, \quad (12.25)$$

donde PNO quiere decir “partes no obstruidas”. Que no será una notación elegante, pero que se entiende muy bien. Notar que la integral es proporcional a la amplitud compleja de la oscilación armónica resultante en P , o sea que la intensidad en P será proporcional al valor absoluto al cuadrado de la integral en (12.25). Y también notar que la no elegancia de la notación está relacionada con nuestras pretensiones de generalidad, ya que (12.25) vale para cualquier distribución de rendijas paralelas en una pantalla opaca, por ejemplo también valdría para el sistema de dos rendijas en la figura 12.14b.

Cuando el plano del frente de ondas incidente no es paralelo al plano $x - y$ del obstáculo, las fuentes elementales ubicadas entre y e $y + dy$ asociadas con partes no obstruidas **no** se encienden al mismo tiempo, es decir, que no están en fase. Como se observa en la figura 12.16, si la dirección de propagación del frente de ondas incidente forma un ángulo $\theta_0 > 0$ con el eje x , la fuente con etiqueta $y > 0$ se enciende más tarde que la fuente con etiqueta $y = 0$. Es claro que el retardo es $y \sin \theta_0$ (proporcional a y con su signo: las fuentes con $y < 0$ están adelantadas, es decir tienen un retardo negativo). Y entonces, la cantidad $\delta(y)$, análoga a los δ_n del argumento Δ_n de las exponenciales complejas de la eq. (12.4), es $\delta(y) = -k y \sin \theta_0$ ya que para nosotros (ver eq. (12.4)), una fase inicial positiva es un adelanto. Luego, la expresión para la contribución elemental $d\psi_P(t)$ resulta

$$d\psi_P(t) = \frac{\mathcal{E}_0 L}{\sqrt{R}} e^{i\omega t} e^{-ikR} e^{ik y (\sin \theta - \sin \theta_0)} dy, \quad (12.26)$$

y la perturbación total

$$\psi_P(t) = \frac{\mathcal{E}_0 L}{\sqrt{R}} e^{i(\omega t - kR)} \int_{\text{PNO}} e^{ik y (\sin \theta - \sin \theta_0)} dy. \quad (12.27)$$

Vemos que la integral en la ec. (12.27) es formalmente idéntica a la de la integral en la ec. (12.25) y que se pasa de (12.25) a (12.27) simplemente reemplazando el parámetro $\sin \theta$ en (12.25) por $\sin \theta - \sin \theta_0$. Esto significa que, independientemente de la forma y distribución de las ranuras en el plano $x - y$, la intensidad como función de la variable $\sin \theta$ obtenida en el caso de incidencia inclinada es la misma función de $\sin \theta$ obtenida en el caso de incidencia paralela al eje x , pero desplazada del origen en el valor $\sin \theta_0$. Cuidado, esto no significa que la distribución de intensidad obtenida en la pantalla de observación en el caso de incidencia inclinada sea simplemente una translación de la distribución de intensidad obtenida en la pantalla de observación cuando la incidencia es paralela al eje x .

12.5.1. Otra vez una rendija

Una vez familiarizados con la idea de Fresnel, de que la envolvente del principio de Huygens no es otra cosa que la interferencia entre onditas elementales, resulta instructivo verificar que la distribución de intensidad obtenida a partir de la ec. (12.25) para la difracción de Fraunhofer producida por una rendija de ancho b iluminada por un frente de ondas incidente paralelo al plano $x - y$ es idéntica a la distribución de intensidad que resulta de la interferencia entre N fuentes idénticas, colineales, equiespaciadas y en fase, en el límite en que la distancia entre fuentes tiende a cero y el número de fuentes tiende a infinito manteniendo el sistema de fuentes en un segmento de longitud b .

Para una rendija de ancho b es muy fácil escribir los límites en la (12.25), porque las únicas partes no obstruidas son las que están en el intervalo $-b/2 \leq y \leq b/2$. Luego,

$$\psi_P(t) = \frac{\mathcal{E}_0 L}{\sqrt{R}} e^{i(\omega t - kR)} \int_{-b/2}^{b/2} e^{ik y \sin \theta} dy. \quad (12.28)$$

La integral de una exponencial que depende linealmente de la variable de integración es la misma exponencial multiplicada por una constante y la diferencia de esta exponencial evaluada en límites con signo cambiado es proporcional a la función $\sin \beta$ que aparece en la ec. (12.22). La cantidad $\beta = \frac{k}{2} b \sin \theta$ se forma en el denominador con las constantes que aparecieron al hacer la integral. Y así verificamos que el principio de Huygens predice que la amplitud compleja difractada por una rendija de ancho b está en total acuerdo con el resultado de la ec. (12.22), obtenido como un límite especial de la interferencia de N fuentes lineales equiespaciadas y uniformemente distribuidas en un segmento de longitud b . Cuando el frente plano incidente no es paralelo al plano de la rendija, las cuentas son iguales porque la exponencial sigue dependiendo linealmente de la variable de integración y todo lo dicho luego de la ec. (12.28) sigue valiendo, con la única diferencia de que ahora

$$\beta = \frac{k}{2} b (\sin \theta - \sin \theta_0), \quad (12.29)$$

un resultado que está de acuerdo con los desplazamientos que habíamos anticipado a partir de la ec. (12.27) para el caso general de cualquier obstáculo plano con simetría de translación.

Es interesante observar que el tamaño de la rendija se percibe en el campo lejano a través del tamaño angular del máximo central (figura 12.15). En cambio, muy cerca de la rendija predomina la influencia directa de las fuentes (secundarias), con una zona muy iluminada de ancho b y una zona de sombras, con intensidades mucho menores, tal como comprobamos en la simulación de campo cercano de la figura 12.10. Vemos entonces que en la zona de Fraunhofer, donde predomina la estructura de haces bien direccionales, la pantalla de observación tiene que estar iluminada esencialmente en una zona que corresponda al máximo central, mientras que en la zona cercana la pantalla de observación tiene que estar iluminada en una región del mismo orden que el ancho b . El máximo central se extiende en un rango angular del orden de λ/b (en realidad éste sería el rango del seno, pero nos estamos preparando para hacer una estimación para establecer un criterio, y en estas cosas todo vale, así que aproximamos nomás el seno y la tangente por el ángulo). Luego, en una pantalla colocada a una distancia Z del obstáculo, el máximo central ocupa una zona con dimensiones del orden de $Z \lambda/b$ (arco, distancia por ángulo). Y entonces diremos que la zona donde predomina la direccionalidad de la difracción (la zona de Fraunhofer) es donde se cumple

$$Z \lambda/b \gg b, \quad (12.30)$$

es decir

$$Z/\lambda \gg (b/\lambda)^2, \quad (12.31)$$

que coincide con la definición de la distancia de Fraunhofer dada en la ec. (12.20) (excepto por un factor 2 ... y no es momento ahora de hacernos problema por un factor 2 en un criterio tan aproximado).

12.5.2. N rendijas idénticas equiespaciadas

Un caso de sumo interés es el de una pantalla plana con N rendijas idénticas equiespaciadas. Si las rendijas tienen ancho b , están separadas una distancia d y están iluminadas por una onda plana incidente que viaja en la dirección x , entonces la integral sobre partes no obstruidas que aparece en la expresión (12.25) se escribe como una suma con N términos

$$\int_{\text{PNO}} = \int_{-b/2}^{b/2} + \int_{-b/2+d}^{b/2+d} + \dots + \int_{-b/2+(N-1)d}^{b/2+(N-1)d}. \quad (12.32)$$

Con el cambio de variable $y' = y - d$, la segunda integral resulta igual a la primera multiplicada por el factor $e^{ikd \sin \theta}$, con el cambio de variable $y' = y - 2d$, la tercera integral resulta igual a la primera multiplicada por el factor $e^{i2kd \sin \theta}$. Y repitiendo un procedimiento similar para las integrales restantes, la integral entre $-b/2$ y $b/2$ es un factor común en todos los términos y entonces

$$\begin{aligned} \int_{\text{PNO}} &= \left[1 + e^{ikd \sin \theta} + e^{i2kd \sin \theta} + \dots + e^{i(N-1)kd \sin \theta} \right] \int_{-b/2}^{b/2} \\ &= \sum_{n=0}^{N-1} \left[e^{ikd \sin \theta} \right]^n \int_{-b/2}^{b/2}. \end{aligned} \quad (12.33)$$

Hemos obtenido que el lado izquierdo de la ec. (12.33), proporcional a la amplitud compleja en el punto P, es igual al producto de dos términos:

- i) una sumatoria idéntica a la que se obtuvo en (12.8) para calcular la amplitud compleja resultante de la **interferencia** lejana de N fuentes equiespaciadas a lo largo del eje y . Según la ec. (12.10) esta sumatoria terminaba siendo proporcional a $\frac{\sin N\alpha}{\sin \alpha}$.
- ii) una integral idéntica a la que se obtuvo en la ec. (12.28) para calcular la **difracción** por una única rendija. Según la ec. (12.22) esta integral terminaba siendo proporcional a $\frac{\sin \beta}{\beta}$.

Así obtenemos la intensidad en la pantalla de observación

$$I_P = I_0 \left(\frac{\sin N\alpha}{\sin \alpha} \right)^2 \left(\frac{\sin \beta}{\beta} \right)^2, \quad (12.34)$$

con α y β funciones de $\sin \theta$ (a cada P le corresponde un θ) dadas por las ecs. (12.11) y (12.21) respectivamente. La expresión (12.34) también vale cuando las rendijas están iluminadas a lo largo de una dirección inclinada un ángulo θ_0 con el eje x (figura 12.16), pero en este caso

$$\alpha = \frac{k}{2} d (\sin \theta - \sin \theta_0), \quad (12.35)$$

y β viene dado por (12.29).

La validez de la **factorización** de I_P como producto de un término llamado *de interferencia*, asociado con la interferencia lejana de N fuentes puntuales idénticas colocadas en el centro de cada ranura, y otro término llamado *de difracción*, representativo de la difracción de una sola ranura, excede este ejemplo particular. Es fácil darse cuenta que la factorización también vale si las ranuras idénticas no están equiespaciadas. En este caso no cambia la forma del llamado término de difracción, que sigue siendo $(\sin \beta)^2/\beta^2$, pero si cambia la forma del llamado término de interferencia, que no viene más dado por $(\sin N\alpha)^2/(\sin \alpha)^2$, sino que se obtiene a partir de la primera sumatoria en la ec. (12.7). Usando transformadas de Fourier y el llamado teorema de convolución, es fácil demostrar que la amplitud de Fraunhofer en la pantalla de observación producida por un obstáculo plano con N aberturas idénticas (no necesariamente con simetría de translación) siempre se factoriza como producto de dos términos: uno asociado con **la interferencia** lejana de N fuentes puntuales idénticas colocadas en el centro de cada abertura y otro representativo de **la difracción por una sola abertura**.

La distribución de intensidad (12.34) se muestra en la figura 12.17 junto con la modulación asociada al término de difracción (línea punteada) para $N = 5$ rendijas idénticas equiespaciadas, $d/\lambda = 3.5$, $b/\lambda = 1.5$ (izquierda) y $b/\lambda = 0.5$ (derecha). Observar que el término de difracción puede modificar notablemente el diagrama de intensidad y eventualmente provocar la desaparición de máximos de interferencia.

Al aumentar el número de rendijas, los órdenes de interferencia se afinan angularmente, tal como se aprecia en la figura 12.18, correspondiente a los mismos parámetros que en la figura 12.17 pero para $N = 50$. Este afinamiento angular se aprovecha en instrumentos llamados *redes de difracción* para separar líneas espectrales de fuentes con longitudes de onda muy parecidas. Las redes de difracción tienen muchas ventajas respecto a otros elementos dispersivos basados en prismas de vidrio. Cuando una fuente que contiene dos frecuencias muy parecidas, por ejemplo una fuente estelar, ilumina la red de difracción, cada frecuencia da lugar a un diagrama como los de las figuras 12.17 y 12.18. La situación es equivalente a tener dos fuentes lejanas con distintas frecuencias (y por lo general, con distinta intensidad en cada frecuencia). La intensidad resultante es entonces la suma de las intensidades individuales correspondientes a cada fuente, una con longitud de onda λ y la otra con longitud de onda $\lambda + \Delta\lambda$. En la figura 12.19a los diagramas asociados con ambas longitudes de onda se superponen de manera tal que la curva de intensidad correspondiente a la suma (curva negra) de curvas de intensidad individuales ((curvas coloreadas) no exhibe ninguna característica que permita asociar posiciones de máximos con longitudes de onda: hay un máximo para el orden cero (este orden siempre es no dispersivo, porque tiene origen físico en la igualdad de caminos geométricos) y hay otros dos máximos para los órdenes ± 1 . Se dice que esta red no resuelve el valor de $\Delta\lambda$ en este rango espectral. Como muestra la figura 12.19b esto es así porque la curva negra tiene un máximo que enmascara los máximos de las curvas coloreadas. Si los máximos de las curvas coloreadas estuvieran más alejadas entre sí (es decir, para mayor $\Delta\lambda$), o si las curvas coloreadas fueran más angostas (es decir, para N más grande), entonces la curva envolvente negra debería revelar la existencia de los máximos individuales. En la figura 12.20 se muestra los diagramas obtenidos para parámetros físicos iguales a los de la figura 12.19, excepto que ahora el número de rendijas se decuplicó ($N = 50$). Cerca de los picos individuales correspondientes a

los órdenes ± 1 de λ y de $\lambda + \Delta\lambda$, no hay casi superposición y la curva negra tiene esencialmente la individualidad de las curvas coloreadas, salvo en la zona de máximos secundarios, que aquí es irrelevante desde el punto de vista espectroscópico por su baja intensidad. En este caso se dice que para esta λ la red resuelve la separación $\Delta\lambda$.

Para todo espectroscopista resulta relevante preguntarse cuál es el límite entre las situaciones mostradas en las figuras 12.19b (la red no resuelve $\Delta\lambda$) y 12.20b (la red resuelve ampliamente $\Delta\lambda$). Un criterio, con su cuota de arbitrariedad como todos los criterios, se conoce como *criterio de Rayleigh* porque fue propuesto por el físico inglés John William Strutt (Lord Rayleigh para los amigos). Rayleigh (Lord Rayleigh) hizo notar que si el máximo del orden m de la curva de $\lambda + \Delta\lambda$ (la curva roja en la figura 12.19b) coincide con el primer cero adyacente al máximo del orden m de la curva de λ (la curva azul en la figura 12.19b), entonces la envolvente (la curva negra) tiene un mínimo local, aproximadamente a mitad de camino entre los dos máximos y con el contraste suficiente como para que los picos individuales puedan dar el presente. Y propuso que se tome esta distancia como indicador de la capacidad del instrumento para separar longitudes de onda (o lo que es lo mismo, frecuencias) cercanas.

Se conoce como *poder resolvente* de una red de difracción al cociente $\lambda/\Delta\lambda$, donde $\Delta\lambda$ es la menor diferencia entre longitudes de onda parecidas que se pueden resolver. De acuerdo con el criterio de Rayleigh esta menor diferencia entre longitudes de onda parecidas se calcula usando que el máximo del orden m de la curva de λ está en un ángulo θ que satisface la condición (12.13)

$$d \sin \theta = m \lambda ,$$

mientras que al máximo de orden m de la curva para $\Delta\lambda$ le corresponde un valor mayor de $\sin \theta$ (es decir un valor mayor de θ), que también satisface la condición (12.13)

$$d [\sin \theta + \Delta(\sin \theta)] = m (\lambda + \Delta\lambda) .$$

Entonces

$$d \Delta(\sin \theta) = m \Delta\lambda . \quad (12.36)$$

Por otra parte, y de acuerdo a lo discutido en la página 212, la distancia entre la posición del máximo de orden m y la posición de su primer cero adyacente, medida en términos de la variable $\sin \theta$ es

$$d \Delta(\sin \theta) = \lambda / N . \quad (12.37)$$

Y según el criterio de Rayleigh, la situación límite de resolución ocurre cuando la distancia entre el máximo de orden m de las curvas de λ y de $\lambda + \Delta\lambda$ (lado izquierdo de la penúltima ecuación) resulta igual a la distancia entre el máximo del orden m y su primer cero adyacente en la curva correspondiente a λ (lado izquierdo de ecuación previa), es decir $m \Delta\lambda = \lambda / N$, de donde se obtiene la siguiente expresión

$$\frac{\lambda}{\Delta\lambda} = m N , \quad (12.38)$$

que dice que el poder resolvente de una red de difracción depende del número N de rendijas (“líneas” en la jerga espectroscópica) y del orden espectral m . Notar que

implícitamente también depende de d a través de la relación (12.17), porque $|m| \geq 1$. Cuando el lado derecho de la ec. (12.36) es justo igual al lado derecho de la ec. (12.37), se dice que $\Delta\lambda$ es el *límite justo de resolución* para esta λ . La definición del límite justo de resolución es un poco arbitraria, como se aprecia en las simulaciones mostradas en la figura (12.21).

En la práctica, hay una tensión entre la tendencia a usar órdenes de difracción altos, para aumentar el poder resolvente, y la tendencia a no tener demasiados órdenes, para no redistribuir la intensidad recibida. En ciertas aplicaciones la intensidad de la señal que llega al instrumento espectroscópico puede ser muy pequeña (en estrellas lejanas, por ejemplo) y repartirla en órdenes de difracción que no serán utilizados sería “malgastar” la intensidad del haz incidente. También hay un problema relacionado con la existencia de la campana de difracción, que por esas casualidades de la naturaleza tiene su máximo justamente en el orden $m = 0$, que no exhibe propiedades dispersivas. Para maximizar la fracción de la intensidad incidente que llega al orden que tiene interés espectroscópico, se han desarrollado diversas técnicas que permiten desplazar el máximo de la campana de difracción de la posición del máximo de orden $m = 0$. Lograr este efecto implica modificar el sistema de ranuras de manera tal que el efecto de interferencia en el punto P de las onditas secundarias que salen de una sola ranura cambie sustancialmente, pero sin que se modifique la interferencia en el punto P entre las N perturbaciones que provienen de cada ranura. Una manera de lograr el efecto, conocido como *efecto de blaze*, es cambiando la transmisión de las rendijas, por ejemplo colocando prismas.

12.6. Obstáculos planos, caso general

Consideremos ahora el caso general de obstáculos planos esquematizado en la figura 12.13 y generalicemos el procedimiento de la sección 12.5, que fue hecho con la hipótesis de obstáculos planos con simetría de translación. En vez de tiras, las fuentes elementales son elementos de área $dS = dx dy$. Y en vez de emitir una onda cilíndrica, cada elemento de área emite una onda esférica que, como antes, tiene amplitud proporcional al elemento de área. Sin embargo, en vez de ser inversamente proporcional a la potencia $1/2$ de la distancia r entre fuente y punto de observación, la amplitud es ahora inversamente proporcional a r .

Para campo lejano seguimos despreciando la variación de la amplitud con r ($r \approx R$) y entonces la contribución elemental $d\psi_P(t)$ a la perturbación en el punto $P(X, Y, Z)$, proveniente de la fuentecita ubicada entre x y $x + dx$ y entre y y $y + dy$, resulta

$$d\psi_P(t) = \frac{\mathcal{E}_0}{R} e^{i\omega t} e^{-ikr} dx dy, \quad (12.39)$$

donde R representa el camino al punto $P(X, Y, Z)$ desde el origen del sistema de coordenadas $x - y$ del obstáculo. En la figura 12.22 se muestran detalles de estos caminos cerca del obstáculo.

Si el punto P está en el foco de una lente, los caminos a lo largo de r y de R son paralelos, es decir tienen la misma inclinación y representan rayos perpendiculares a los

frentes de onda que llegan a la lente para converger en P. En otras palabras, estos rayos están en la dirección del versor unitario $\hat{k}$ paralelo al vector de onda $\vec{k}$ representativo de las ondas planas que llegan a P. Para referir la distancia r , entre cada fuente elemental $dS(x, y)$ y el punto de observación, al camino R desde el origen (donde puede que haya o no fuente elemental, todo depende de si el origen no está o está obstruido), notemos que los trayectos por estos caminos son iguales desde cualquier plano perpendicular a $\hat{k}$. Luego, r y R difieren en la proyección sobre $\hat{k}$ del vector $x\hat{x} + y\hat{y}$ que apunta a dS desde el origen y entonces el argumento de la exponencial espacial en 12.39 se escribe

$$kr = kR - k \left[\hat{k} \cdot (x\hat{x} + y\hat{y}) \right] = kR - k_x x - k_y y, \quad (12.40)$$

con $k_x = \vec{k} \cdot \hat{x}$ y $k_y = \vec{k} \cdot \hat{y}$. Así, la contribución elemental (12.39) se puede reescribir como

$$d\psi_P(t) = \frac{\mathcal{E}_0}{R} e^{i(\omega t - kR)} e^{i(k_x x + k_y y)} dx dy, \quad (12.41)$$

una cantidad que juega el mismo papel que jugaban los términos de la sumatoria en (12.7).

Observar que la fase de la contribución elemental (12.41) es otra vez una función lineal de las coordenadas que etiquetan a la fuente elemental en $dS(x, y)$. Sumando sobre todas las fuentes, es decir, integrando sobre todas las regiones del plano $x - y$ no obstruidas, se obtiene finalmente la siguiente expresión

$$\psi_P(t) = \frac{\mathcal{E}_0}{R} e^{i(\omega t - kR)} \int_{\text{PNO}} e^{i(k_x x + k_y y)} dx dy, \quad (12.42)$$

que representa la perturbación total en el punto P debida a la difracción por un obstáculo plano caracterizado por partes que son completamente transparentes o completamente opacas (es decir, no hay partes con zonas intermedias, más o menos transparentes).

Las cantidades $\hat{k} \cdot \hat{x} = k_x/k$ y $\vec{k} \cdot \hat{y} = k_y/k$ en (12.42) definen la inclinación de la dirección de observación a lo largo de los ejes x e y respectivamente. Representan la dirección en la que debería apuntar un telescopio para enfocar al punto P muy lejano y juegan el rol análogo al de la cantidad $\sin \theta$ en la ec. (12.25). Las componentes de $\hat{k}$ se pueden escribir en términos de los ángulos de coordenadas esféricas ϑ (el ángulo polar entre $\hat{k}$ y el eje z) y φ (el ángulo azimutal, entre la proyección de $\hat{k}$ en el plano $x - y$ y el eje x) y estos ángulos se expresan en función de las coordenadas X, Y y Z del punto P. En el caso de usar una lente de distancia focal f , $\tan \vartheta = \sqrt{X^2 + Y^2}/f$ y $\tan \varphi = Y/X$. Es instructivo seguir estos pasos para confirmar que a cada par $X - Y$ (a cada punto P de la pantalla de observación) le corresponde un par $\vartheta - \varphi$ (una inclinación del rayo que pasa por el centro de la lente y no se desvía) y a cada par $\vartheta - \varphi$ le corresponde un par $k_x - k_y$. Dicho de otra manera: de todas las perturbaciones emitidas por las onditas de Huygens, solamente se detectan en el punto P aquellas que tienen la inclinación correcta, descartando todas las demás. Y eligiendo la forma de las aberturas en el plano $x - y$ se controla la superposición de componentes con la misma inclinación.

Cuando el plano del frente de ondas incidente no es paralelo al plano $x - y$, las fuentes elementales no se encienden al mismo tiempo. Si a la dirección de propagación

del frente de ondas incidente le corresponde el vector de onda $\vec{k}_0$, es fácil repetir los pasos que dimos en el caso de rendijas y obtener que

$$\psi_P(t) = \frac{\mathcal{E}_0}{R} e^{i(\omega t - kR)} \int_{\text{PNO}} e^{i[(k_x - k_{0x})x + (k_y - k_{0y})y]} dx dy, \quad (12.43)$$

12.6.1. Abertura rectangular

Para una abertura rectangular de lados a (según x) y b (según y) centrada en el origen, los límites en las (12.42) y (12.43) son $-a/2 \leq x \leq a/2$ y $-b/2 \leq y \leq b/2$. Luego, como el integrando se factoriza en una función exclusivamente de x por una función exclusivamente de y , se obtiene un producto de dos integrales con la misma forma que en la ec. (12.28). De esta manera, la amplitud compleja es proporcional a un producto de senos cardinales, uno correspondiente a la integral en x y otro a la integral en y . La intensidad en el punto P queda

$$I_P = I_0 \left(\frac{\sin \beta_x}{\beta_x} \right)^2 \left(\frac{\sin \beta_y}{\beta_y} \right)^2, \quad (12.44)$$

con

$$\beta_x = \frac{1}{2}(k_x - k_{0x}) a = \pi \frac{a}{\lambda} [(\hat{k} \cdot \hat{x}) - (\hat{k}_0 \cdot \hat{x})], \quad (12.45)$$

$$\beta_y = \frac{1}{2}(k_y - k_{0y}) b = \pi \frac{b}{\lambda} [(\hat{k} \cdot \hat{y}) - (\hat{k}_0 \cdot \hat{y})]. \quad (12.46)$$

Suponiendo $\hat{k}_0 = \hat{k}_0 = 0$ (frente de ondas incidente paralelo al plano $x - y$), la figura de difracción de la abertura rectangular tiene un máximo principal con intensidad I_0 en $\beta_x = \beta_y = 0$, acompañado de un doble juego de máximos secundarios y que están limitados por ceros en inclinaciones definidas por $\beta_x = n\pi$ y $\beta_y = m\pi$, con n y m enteros distintos de cero. El rango de inclinaciones horizontales y verticales abarcadas por el máximo principal es $\Delta(\hat{k} \cdot \hat{x}) = 2\lambda/a$ y $\Delta(\hat{k} \cdot \hat{y}) = 2\lambda/b$ respectivamente.

12.6.2. Abertura circular

a ser actualizado próximamente

12.7. Óptica de Fourier

a ser actualizado próximamente

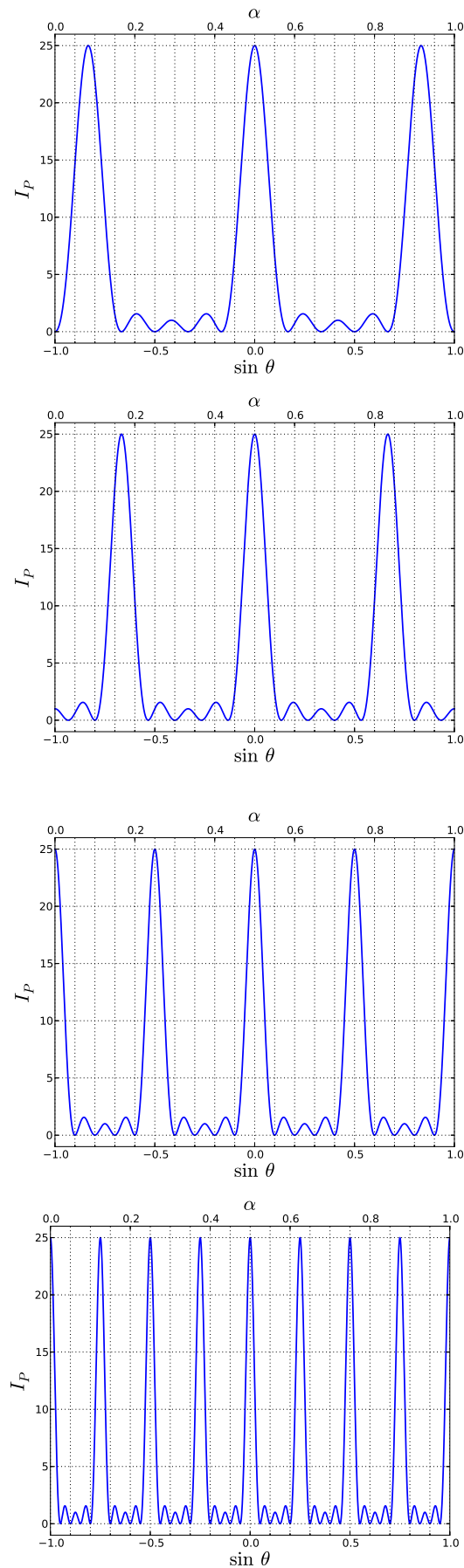


Figura 12.6: Intensidad I_P dada por la ec. (12.12) como función de α (eje superior) y de $\sin \theta$ (eje inferior) para $N = 5$ fuentes idénticas equiespaciadas y para distintos valores del cociente d/λ : 1.2 (arriba izquierda), 1.5 (arriba derecha), 2 (abajo izquierda) y 4 (abajo derecha).

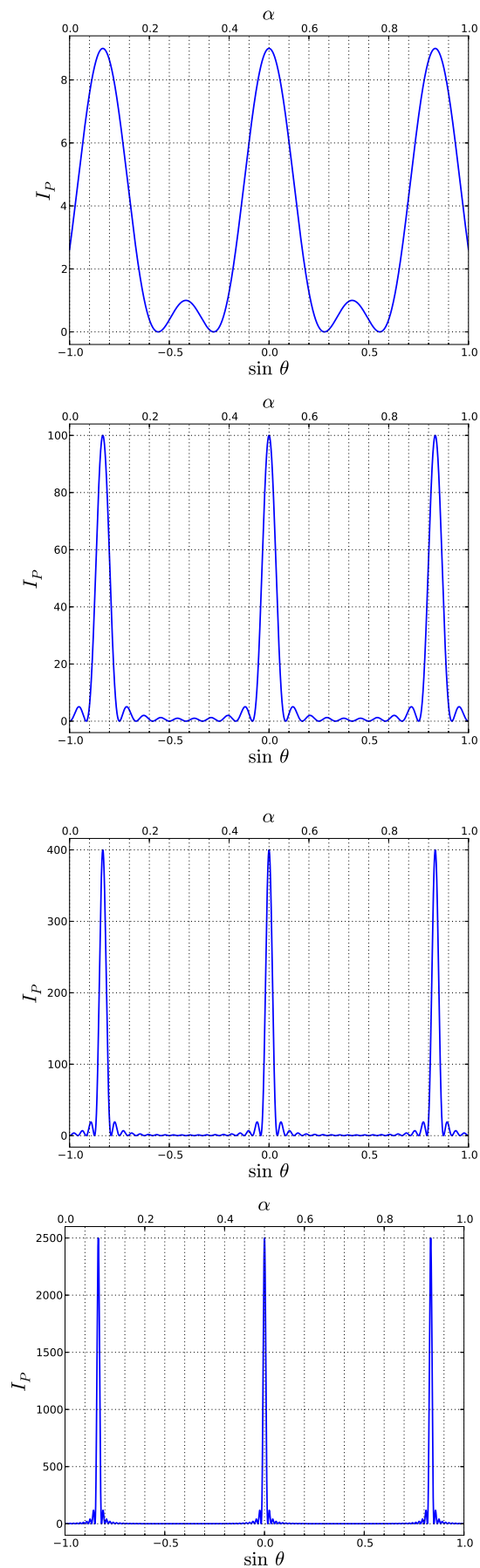
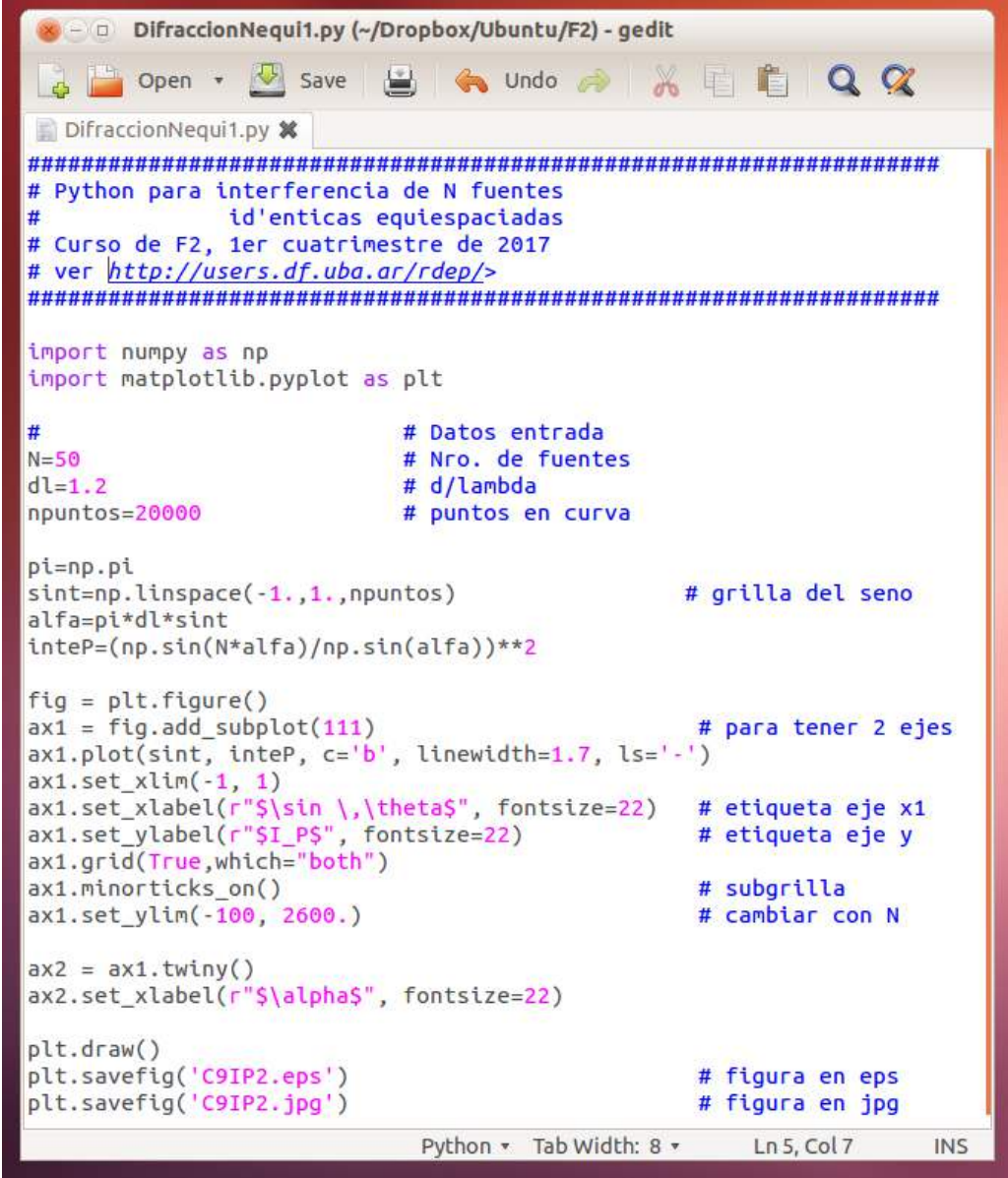


Figura 12.7: Intensidad I_P dada por la ec. (12.12) como función de α (eje superior) y de $\sin \theta$ (eje inferior) para $d/\lambda = 1.2$ y distintos valores de N : 3 (arriba izquierda), 10 (arriba derecha), 20 (abajo izquierda) y 50 (abajo derecha).



```

DifraccionNequi1.py (~/.Dropbox/Ubuntu/F2) - gedit
Open Save Undo
DifraccionNequi1.py
#####
# Python para interferencia de N fuentes
#           id'enticas equiespaciadas
# Curso de F2, 1er cuatrimestre de 2017
# ver http://users.df.uba.ar/rdep/
#####

import numpy as np
import matplotlib.pyplot as plt

#                               # Datos entrada
N=50                             # Nro. de fuentes
dl=1.2                           # d/lambda
npuntos=20000                   # puntos en curva

pi=np.pi
sint=np.linspace(-1.,1.,npuntos) # grilla del seno
alfa=pi*dl*sint
inteP=(np.sin(N*alfa)/np.sin(alfa))**2

fig = plt.figure()
ax1 = fig.add_subplot(111)       # para tener 2 ejes
ax1.plot(sint, inteP, c='b', linewidth=1.7, ls='-.')
ax1.set_xlim(-1, 1)
ax1.set_xlabel(r"$\sin \backslash, \backslash \theta$", fontsize=22) # etiqueta eje x1
ax1.set_ylabel(r"$I_P$", fontsize=22) # etiqueta eje y
ax1.grid(True, which="both")
ax1.minorticks_on()             # subgrilla
ax1.set_ylim(-100, 2600.)      # cambiar con N

ax2 = ax1.twinx()
ax2.set_xlabel(r"$\alpha$", fontsize=22)

plt.draw()
plt.savefig('C9IP2.eps')        # figura en eps
plt.savefig('C9IP2.jpg')        # figura en jpg

Python Tab Width: 8 Ln 5, Col 7 INS

```

Figura 12.8: Comandos Python para obtener figuras 12.6 y 12.7.

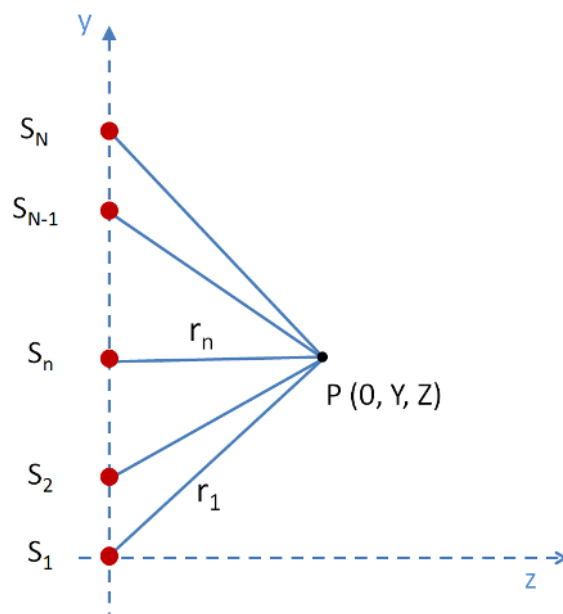


Figura 12.9: Las perturbaciones de N fuentes puntuales colineales equiespaciadas interfieren en un punto P “cercano” en el plano $x = 0$ que contiene a las fuentes.

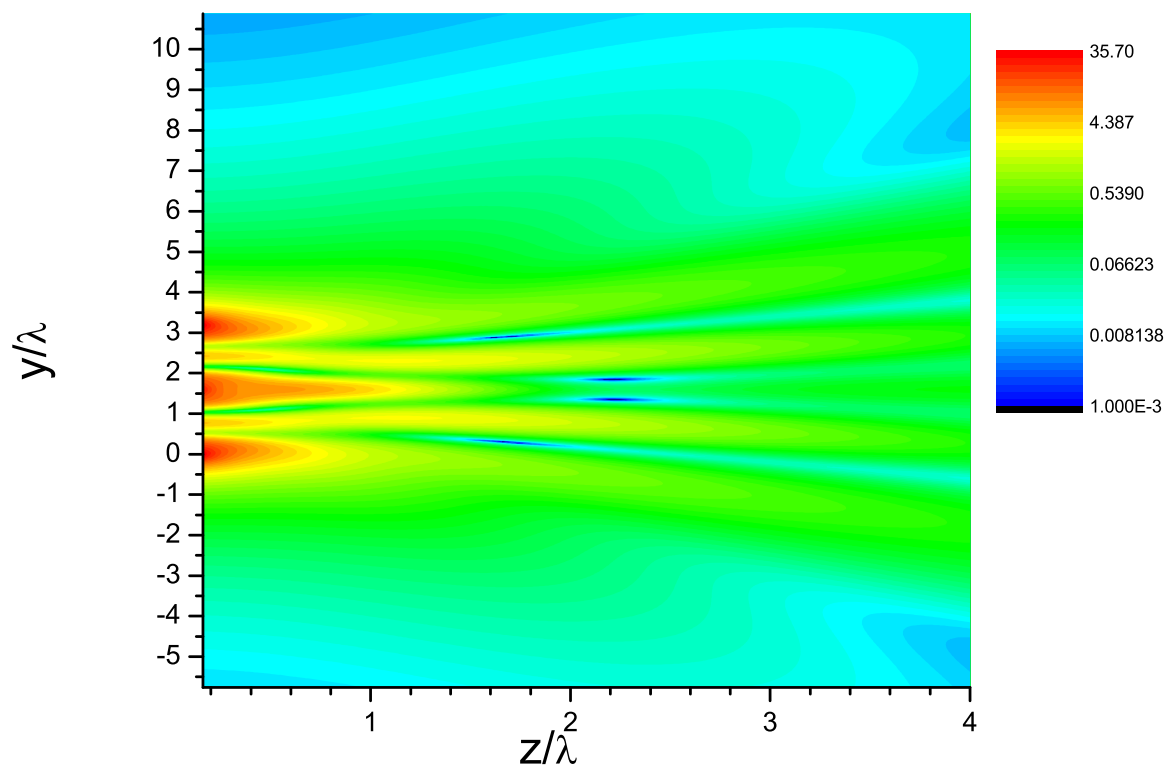


Figura 12.10: mapa de colores en escala logarítmica para la intensidad en la zona $0.12 \leq Z/\lambda \leq 4$ producida por $N = 3$ fuentes idénticas, colocadas en $y_1 = 0$, $y_2 = 1.6\lambda$ e $y_3 = 3.2\lambda$.

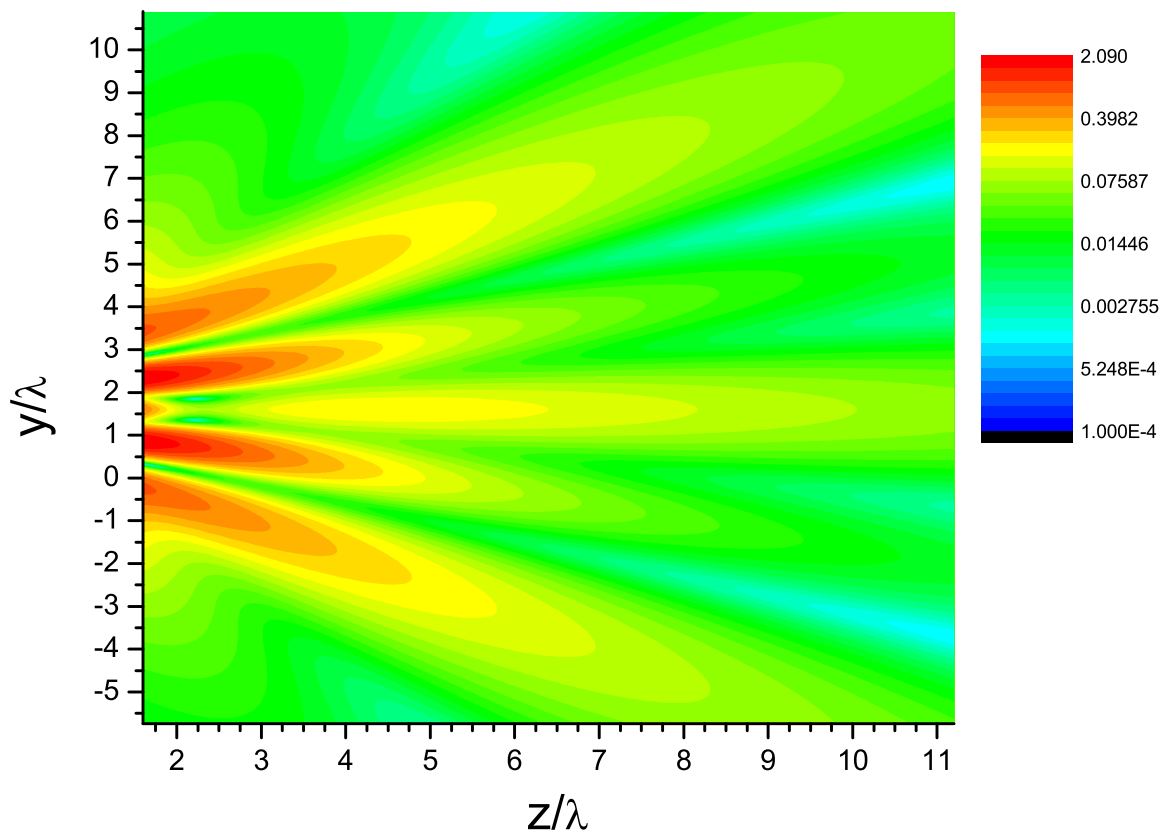


Figura 12.11: como la figura 12.10, pero para la zona $1.6 \leq Z/\lambda \leq 11.2$.

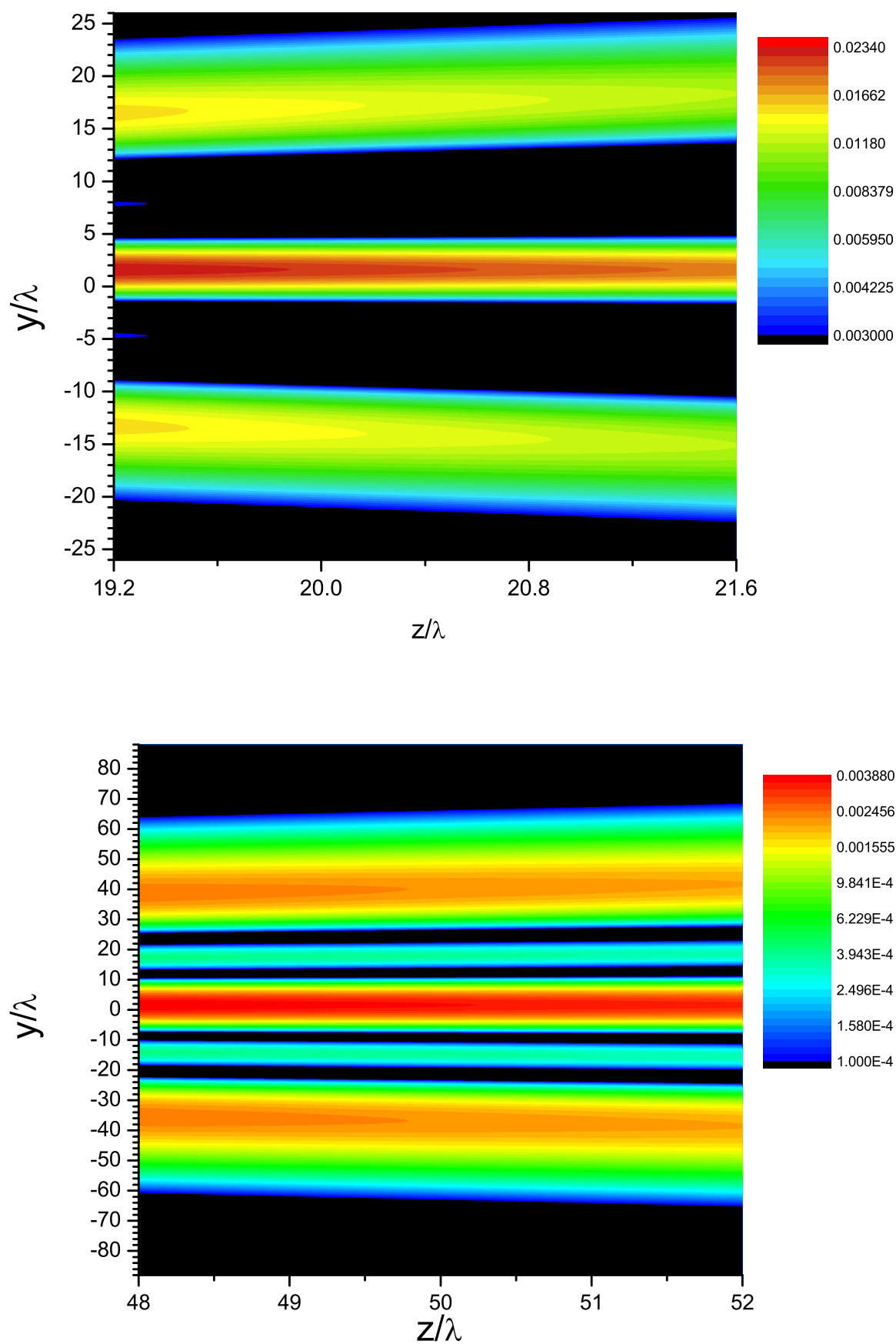


Figura 12.12: como en las figuras 12.10 y 12.11, pero para zonas mas alejadas de las fuentes: a) en la zona $19.2 \leq Z/\lambda \leq 21.6$ (panel superior) y b) $48 \leq Z/\lambda \leq 52$ (panel inferior). Observar la estructura de haces bien direccionales, en las direcciones predichas por la ec. (12.12) y la correlación con los resultados mostrados en el primer panel de la figura 12.7

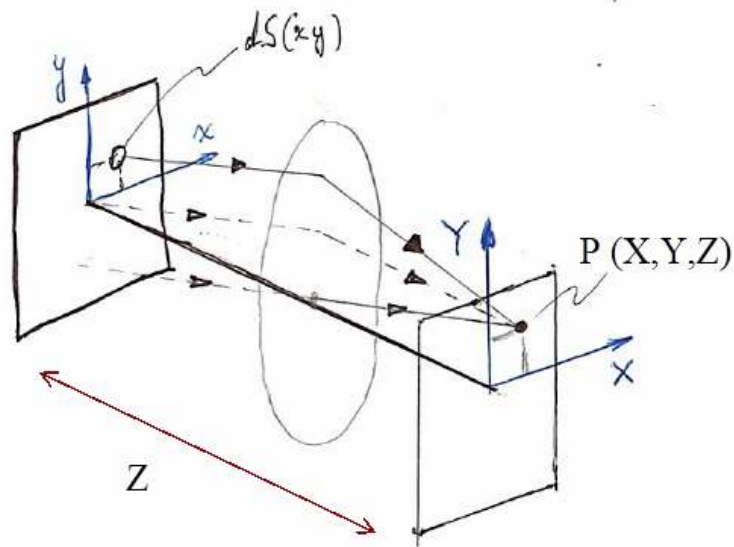


Figura 12.13: Un frente incidente plano ilumina un obstáculo colocado en el plano $x - y$. Cada elemento de superficie dS asociado con parte de un agujero del obstáculo está identificado por coordenadas genéricas (x, y) e irradia hacia adelante para dar lugar a la perturbación resultante en $P(X, Y, Z)$.

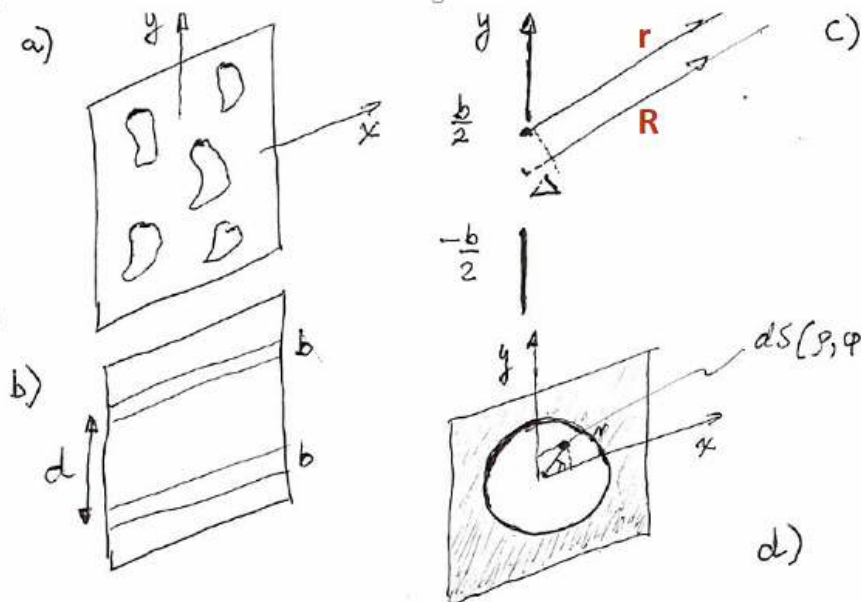


Figura 12.14: Algunos obstáculos planos: a) chapa con perforaciones de la misma o distinta forma; b) interferómetro de Young con dos rendijas lineales del mismo ancho b separadas una distancia d ; c) esquema para rendija lineal de ancho b ; d) disco opaco, o abertura circular en pantalla opaca, según el sombreado represente opacidades o transparencias.)

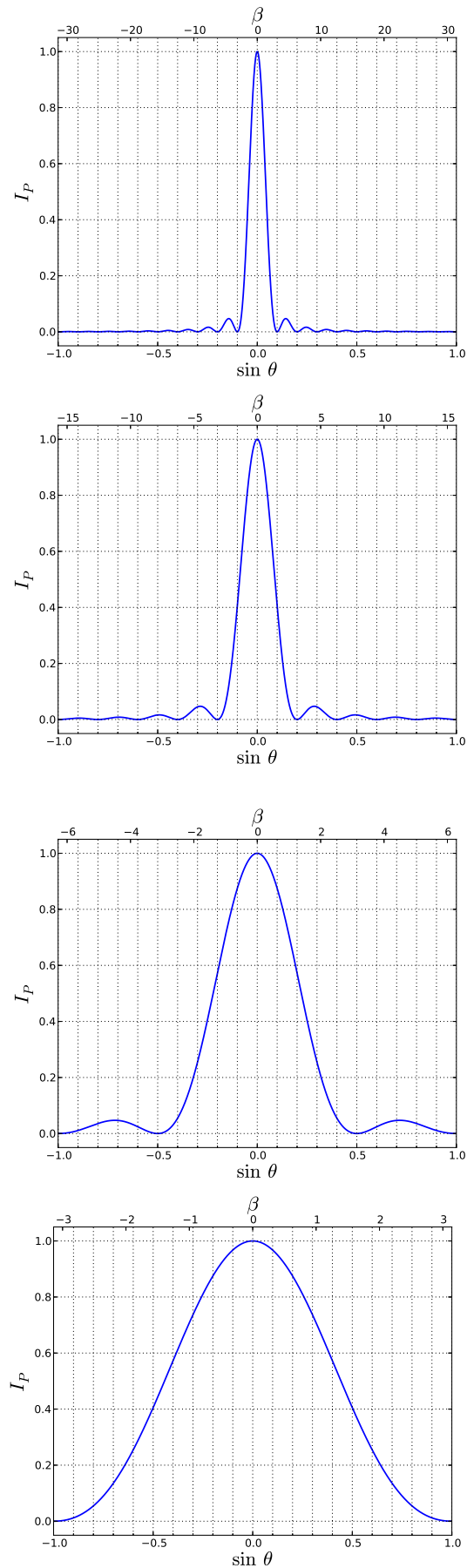


Figura 12.15: Intensidad I_P dada por la ec. (12.22) como función de β (eje superior) y de $\sin \theta$ (eje inferior) para $b/\lambda = 10$ (arriba izquierda), 5 (arriba derecha), 2 (abajo izquierda) y 1 (abajo derecha).

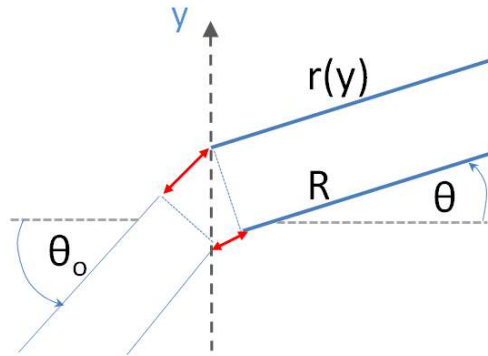


Figura 12.16: Cuando el plano del frente de ondas incidente no es paralelo al plano $x - y$ del obstáculo, las fuentes elementales ubicadas entre y e $y + dy$ **no** se encienden al mismo tiempo. Para un ángulo $\theta_0 > 0$ (como en el dibujo), la fuente con etiqueta $y > 0$ se enciende más tarde que la fuente con etiqueta $y = 0$ (la que tiene un camino R hasta P) y el retardo entre ambas es proporcional a y .

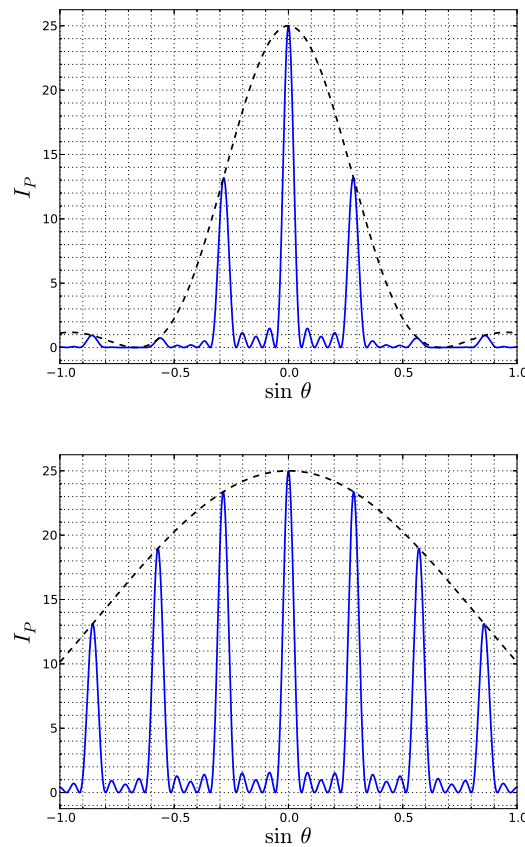


Figura 12.17: Intensidad I_P dada por la ec. (12.34) como función de $\sin \theta$ para $N = 5$ rendijas idénticas equiespaciadas con $d/\lambda = 3.5$ y $b/\lambda = 1.5$ (izquierda) y $b/\lambda = 0.5$ (derecha). La línea punteada corresponde al término de difracción.

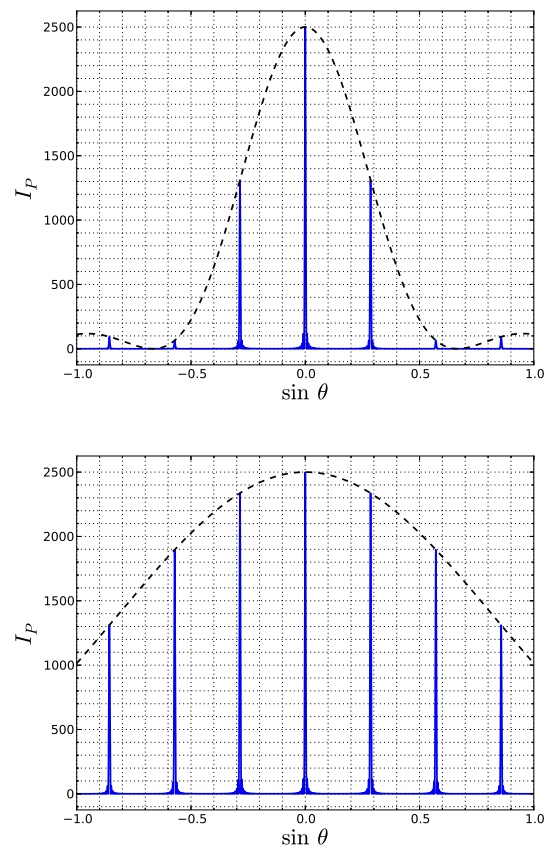


Figura 12.18: Como en la figura 12.17, pero para $N = 50$.

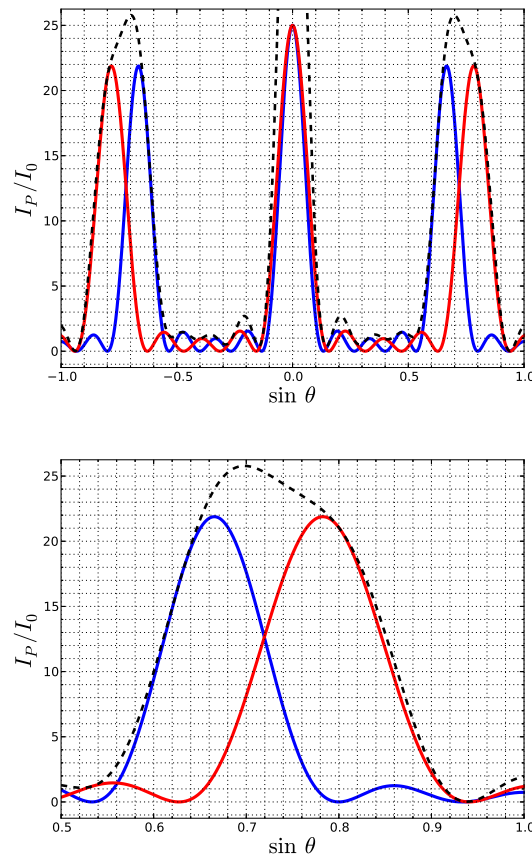


Figura 12.19: Propiedades dispersivas de $N = 5$ rendijas idénticas equiespaciadas con $d = 1\,\mu\text{m}$ y $b = 0.3\,\mu\text{m}$. La línea negra punteada representa el diagrama de difracción de Fraunhofer de una fuente que emite con longitudes de onda $\lambda = 1\,\mu\text{m}$ y $\lambda + \Delta\lambda = 1.176\,\mu\text{m}$, la línea azul (roja) corresponde a una fuente monocromática de λ ($\lambda + \Delta\lambda$). En el panel de la derecha se muestra ampliada la zona del orden de interferencia $+1$.

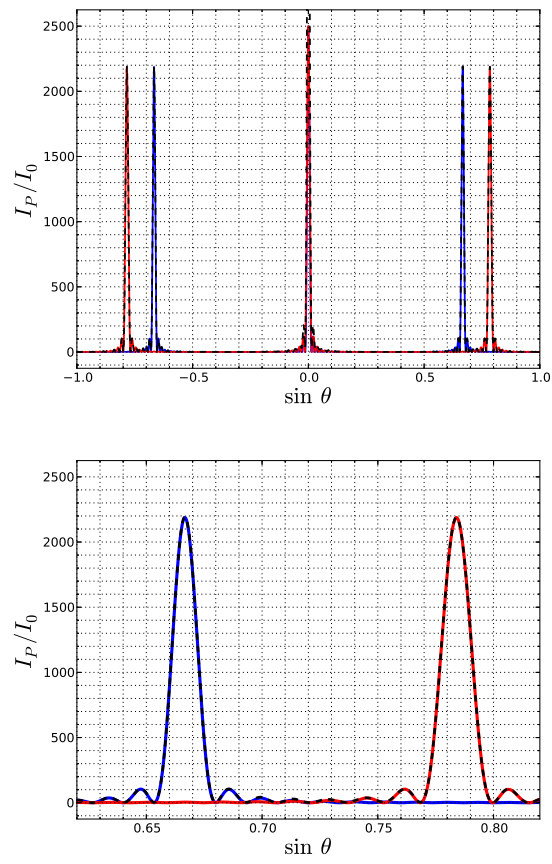


Figura 12.20: Como en la figura 12.19, pero para $N = 50$.

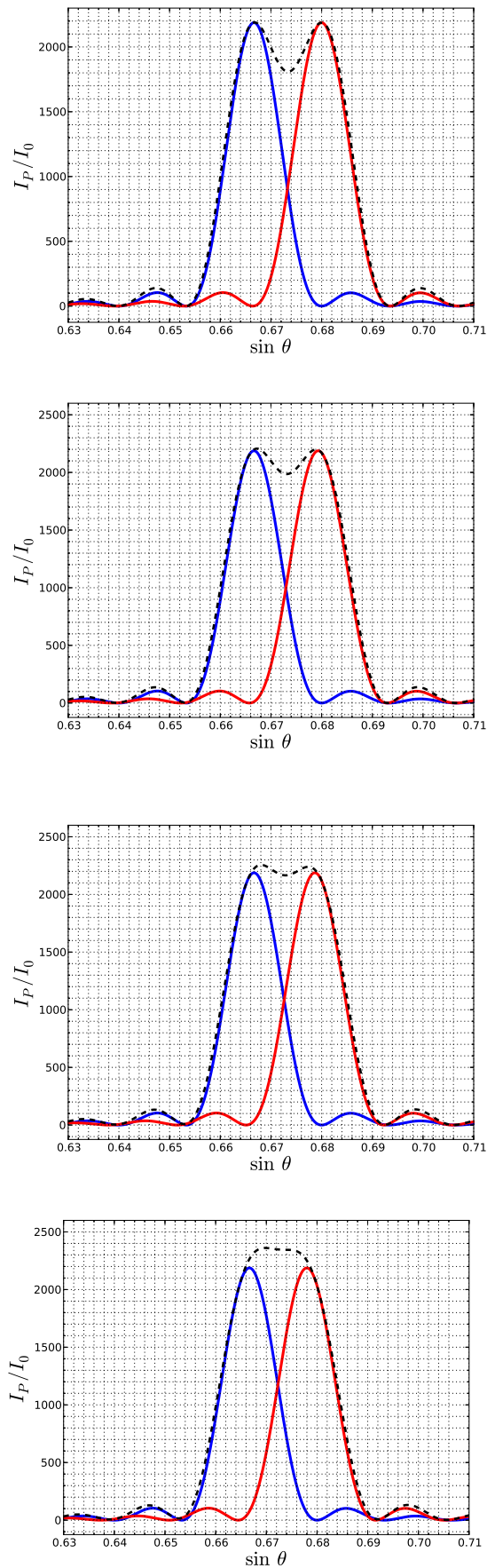


Figura 12.21: Distribución de intensidad cerca del orden $m = +1$ para $N = 50$ rendijas idénticas equiespaciadas con $d = 1\mu\text{m}$ y $b = 0.3\mu\text{m}$. La línea negra punteada representa la intensidad de una fuente que emite con longitudes de onda $\lambda = 1\mu\text{m}$ y $\lambda + \Delta\lambda$ para distintos valores de $\Delta\lambda$. La línea azul (roja) corresponde a una fuente monocromática de λ ($\lambda + \Delta\lambda$). i) arriba izquierda: $\Delta\lambda$ es igual al límite justo calculado con el criterio de Rayleigh, ec. (12.38); arriba derecha: $\Delta\lambda$ es igual al 95% del valor del límite justo; abajo izquierda: $\Delta\lambda$ es igual al 90% del valor del límite justo; abajo derecha:

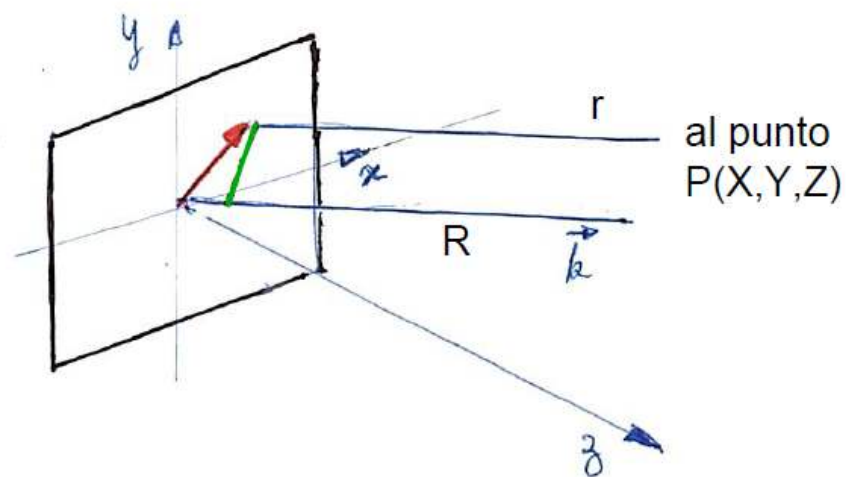


Figura 12.22: el vector $x\hat{x} + y\hat{y}$ apunta desde el origen del obstáculo al diferencial de superficie dS ubicado entre x y $x + dx$ y entre y y $y + dy$. La distancia r entre dS y P difiere de la distancia R entre el origen y P en la proyección del vector $x\hat{x} + y\hat{y}$ a lo largo del versor unitario $\hat{k}$ paralelo al vector de onda representativo de las ondas planas que llegan a P

Bibliografía

- [1] Juan G. Roederer. *Mecánica Elemental*. Manuales. EUDEBA, 2010.
- [2] Frank S. Crawford Jr. *Waves*. Berkeley Physics Course, Vol. 3. McGraw-Hill, New York, 1968.
- [3] Frank S. Crawford Jr. *Ondas*. Curso de Física de Berkeley, Vol. 3. Editorial Reverté, Barcelona, 1971.